ENTERPRISE AI TRANSFORMATION AND CONSULTING PROFESSIONAL

COMPLETE STUDY COLLECTION · COURSES 1–7

The Complete Certificate Study Book

All seven study books of the certificate in one volume — from the advisor's role to a complete advisory package — with case studies, consulting templates, cheat sheets, exercises and mock exams with full answer explanations.

7 COURSES · 641 PAGES OF STUDY MATERIAL · ONE UNIFIED NAVIGATION

Study edition · 2026

Includes full answer explanations

About this collection

This volume brings together the seven study books of the **Enterprise AI Transformation and Consulting Professional** certificate. Each course keeps its own cover, contents, modules, templates, cheat sheets and mock exam, so you can study them in order or jump straight to the one you need.

The learning path

Courses 1–6 each build one capability of the advisor; Course 7 applies all of them to one engagement.

What is inside

#	Course	Focus	Starts at	Pages
1	**The AI Transformation Advisor**	The advisor role, capability landscape, toolkit, engagement framing, due diligence	C1 · 1	118
2	**Identify and Prioritize AI Opportunities**	Finding, qualifying and prioritizing AI opportunities into a portfolio	C2 · 1	94
3	**Assess AI Readiness**	Data, technology and organizational readiness; reporting readiness	C3 · 1	95
4	**Design AI Transformation Roadmaps**	Roadmaps, horizons, workflow and agent opportunities, pressure-testing	C4 · 1	94
5	**AI Business Cases and ROI Models**	Business-case structure, ROI models, value metrics, decision-ready cases	C5 · 1	88
6	**AI Governance, Risk, and Executive Communication**	Governance, ethics, risk management and executive communication	C6 · 1	61
7	**AI Transformation Advisory Package**	Capstone: a full advisory engagement and portfolio piece	C7 · 1	91

How to navigate

Tool	How it works
Bookmarks panel	One tree for the whole volume: Collection → Course → Module → Section. Each course also has a *Reference & Exam* branch (master cheat sheet, templates, glossary, mock exam, answers).
Page references	Pages are labelled per course — **C3 · 27** means Course 3, page 27. The labels match the page number printed in the footer and can be typed into the page box of most PDF readers.
Clickable contents	The master contents on the next pages and each course's own contents page are linked — click any line to jump.

Master contents

COURSE 1
The AI Transformation Advisor — C1 · 1

MODULE 1	What AI Transformation Advisory Is	C1 · 6
MODULE 2	Mapping the Enterprise AI Capability Landscape	C1 · 26
MODULE 3	The Advisory Toolkit	C1 · 46
MODULE 4	Framing and Scoping the Engagement	C1 · 62
MODULE 5	Due Diligence on Claims, Vendors, and Signals	C1 · 80
BACK	Reference & Exam: cheat sheet, templates, glossary, mock exam	C1 · 99

COURSE 2
Identify and Prioritize AI Opportunities — C2 · 1

MODULE 1	Finding AI Opportunities	C2 · 6
MODULE 2	Qualifying Candidate Use Cases	C2 · 24
MODULE 3	Prioritizing with a Value-Effort Lens	C2 · 41
MODULE 4	Building the Prioritized Opportunity Portfolio	C2 · 58
BACK	Reference & Exam: cheat sheet, templates, glossary, mock exam	C2 · 75

COURSE 3
Assess AI Readiness — C3 · 1

MODULE 1	The Dimensions of Readiness	C3 · 6
MODULE 2	Data and Technology Readiness	C3 · 25
MODULE 3	Organizational Readiness	C3 · 43
MODULE 4	Reporting Readiness	C3 · 60
BACK	Reference & Exam: cheat sheet, templates, glossary, mock exam	C3 · 76

COURSE 4
Design AI Transformation Roadmaps — C4 · 1

MODULE 1	Roadmap Fundamentals	C4 · 6
MODULE 2	Sequencing the Investment Portfolio Across Horizons	C4 · 24
MODULE 3	AI-Enabled Workflow and Agent Opportunities	C4 · 40
MODULE 4	Pressure-Testing the Roadmap	C4 · 57
BACK	Reference & Exam: cheat sheet, templates, glossary, mock exam	C4 · 75

COURSE 5
AI Business Cases and ROI Models — C5 · 1

MODULE 1	Structuring the AI Business Case	C5 · 6
MODULE 2	Building the ROI Model	C5 · 23
MODULE 3	Defining Value Metrics	C5 · 39
MODULE 4	Making the Decision-Ready Case	C5 · 55
BACK	Reference & Exam: cheat sheet, templates, glossary, mock exam	C5 · 71

COURSE 6
AI Governance, Risk, and Executive Communication C6 · 1

MODULE 1 AI Governance and Ethics C6 · 6
MODULE 2 AI Risk Management C6 · 25
BACK Reference & Exam: cheat sheet, templates, glossary, mock exam C6 · 45

COURSE 7
AI Transformation Advisory Package C7 · 1

MODULE 1 Opportunity Assessment, Portfolio and Readiness C7 · 6
MODULE 2 Roadmap, Business Case and Governance C7 · 25
MODULE 3 Integrating the Package and the Executive Presentation C7 · 43
MODULE 4 Documenting Reasoning: The Portfolio Piece C7 · 59
BACK Reference & Exam: cheat sheet, templates, glossary, mock exam C7 · 74

CONSULTANT'S TIP · Suggested route

Work through Courses 1–6 in order, completing each mock exam before moving on, then do Course 7 as the capstone. For revision, read the six *Reference & Exam* master cheat sheets back to back, then the Advisor's Playbook in Course 7.

ENTERPRISE AI TRANSFORMATION AND CONSULTING PROFESSIONAL

COURSE 1 · STUDY BOOK

The AI Transformation Advisor

What the role is, how to map an enterprise's AI landscape, the deliverables an engagement produces, how to frame and scope the work, and how to test claims before a plan is built on them.

5 MODULES · CASE STUDIES · TEMPLATES · CHEAT SHEETS · MOCK EXAM

Study edition · 2026

Includes full answer explanations

How to use this book

This study book accompanies **Course 1, The AI Transformation Advisor**, of the Enterprise AI Transformation and Consulting Professional certificate. It is written for consultants, AI strategy and transformation leads, business analysts and change professionals who must give credible advice on enterprise AI without being the ones who build it.

Each module follows the same rhythm so that you always know where you are:

Element	What you get	How to use it
Module opener	Learning objectives, section list, study time	Read first; return to it to self-test at the end
Overview & key takeaways	The module on one screen	Skim before, re-read after
Core sections	Concepts, tables, diagrams, worked examples	Study actively: annotate, restate in your own words
Callout boxes	Key ideas, examples, pitfalls, exam traps, consultant tips	Exam traps are the most frequently tested ideas
Case study	A realistic enterprise scenario with model answers	Answer the questions before reading the model answers
Templates	Fill-in consulting tools (T-codes)	Copy into your own engagements
Do / Avoid	Best practices and pitfalls side by side	Use as a pre-meeting checklist
Cheat sheet	One-page module summary	Revise from these in the final week
Knowledge check	Scenario questions with explanations for **every** option	Understanding why the wrong answers are wrong matters as much as getting the right one

Callout key

KEY IDEA
The idea to remember if you remember nothing else.

EXAM TRAP
A frequently tested distinction or a tempting wrong answer.

EXAMPLE
A worked illustration of the concept.

CONSULTANT'S TIP
Practical wording or technique from the field.

PITFALL
A common mistake made in real engagements.

DEFINITION
A precise definition worth learning word for word.

Suggested study plan (four weeks)

Week	Study	Practise
1	Module 1 and Module 2 (sections 2.1–2.5)	Reframe three real or imagined asks with template T1.1
2	Module 2 (remaining) and Module 3	Build a small capability inventory for an organization you know
3	Module 4 and Module 5	Write a scope statement and a one-page advisory note
4	Revise all cheat sheets and the glossary	Take the mock exam under timed conditions (60 min); review every explanation

NOTE · About the examples

All organizations in case studies and examples are fictional. Figures in them are illustrative and do not describe any real company. Regulatory references are provided for orientation only; always check current law and guidance, and refer legal interpretation to qualified specialists.

Contents

FRONT **How to use this book** 2

MODULE 1 **What AI Transformation Advisory Is** 6
Defining the AI advisory role 7
Advisory vs management consulting vs implementation 8
Advisory judgment vs technical build; demonstration vs deployment 9
Why AI decisions are different: structural uncertainty 10
The three foundations of enterprise AI adoption 11
Pilot purgatory: why AI stalls rather than fails 12
From the stated ask to the underlying business problem 13
Advisory decisions, build decisions, and who decides what 14
Case study: Northbridge Insurance 16
Consulting templates 18
Best practices and pitfalls 20
Module 1 cheat sheet 21
Knowledge check 22
Knowledge check: answers and explanations 24

MODULE 2 **Mapping the Enterprise AI Capability Landscape** 26
Capabilities, not products 27
Define what counts — before counting anything 28
AI vs rules-based automation; the durable classes 29
Beyond the sanctioned: ungoverned AI, unsanctioned tools and sprawl 30
Building the AI capability inventory 31
Scoring on independent axes 31
Inventory vs maturity or readiness assessment 32
Gap conditions and multi-input prioritization 33
Talking to executives: trade-offs, build vs buy, platform fit 34
Case study: Meridian Logistics 36
Consulting templates 38
Best practices and pitfalls 40
Module 2 cheat sheet 41
Knowledge check 42
Knowledge check: answers and explanations 44

MODULE 3 **The Advisory Toolkit** 46
The arc of advisory deliverables 47
The five deliverables in detail 48
Artifact vs report: the three-question test 49
Delivery deliverables vs governance documentation 50
Phase gates, scope and decision rights: what “complete” means 51
Commercial form and the meaning of completion 51
Case study: Halden Bank 53
Consulting templates 55
Best practices and pitfalls 56
Module 3 cheat sheet 57
Knowledge check 58
Knowledge check: answers and explanations 60

MODULE 4 **Framing and Scoping the Engagement** 62
Why AI engagement scoping is different 63
Clarifying business objectives 63
Stakeholder mapping and alignment 64
Requirements gathering 65
The scope boundary 66
Vocabulary work and keeping scope creep out 67
Expectation management with executive sponsors 67
Change control and escalation paths 68
The consultant's scope of authority 68
Case study: Vantis Manufacturing 70
Consulting templates 72
Best practices and pitfalls 74
Module 4 cheat sheet 75
Knowledge check 76
Knowledge check: answers and explanations 78

MODULE 5 **Due Diligence on Claims, Vendors, and Signals** 80
Why due diligence is an advisory duty 81
Reading a capability claim back to its measurement conditions 82
The evidence ladder: demonstration, PoC, pilot, production 83
Questioning benchmarks and headline statistics 83
Structured AI vendor evaluation 84
Third-party and counterparty risk on multi-year commitments 85
Supplier viability and financial durability 85
Separating signal from noise 86
Writing the evidence-based advisory note 86
Case study: Corvel Health 88
Consulting templates 90
Best practices and pitfalls 93
Module 5 cheat sheet 94
Knowledge check 95
Knowledge check: answers and explanations 97

BACK **Course at a glance: the master cheat sheet** 99
Template index 99
Glossary 101
Mock exam 104
Mock exam: answer key and explanations 110
Further reading and next steps 118

MODULE 1

What AI Transformation Advisory Is

Defining the role, separating it from consulting and building, and learning why enterprise AI stalls between pilot and production — not because the model fails, but because the decisions around it were never made.

YOU WILL BE ABLE TO

- Define AI transformation advisory in one sentence and defend it to a sceptical executive
- Distinguish advisory from management consulting and from hands-on AI implementation
- Separate advisory judgment from technical build, and a demonstration from a deployment
- Explain the structural uncertainty that makes AI decisions unlike ordinary IT and analytics decisions
- Assess the three foundations of enterprise AI adoption
- Diagnose why initiatives stall between pilot and production
- Reframe a sponsor's stated ask into the underlying business problem
- Sort work into advisory decisions and build decisions, and name who must decide what

IN THIS MODULE

1. Defining the advisory role
2. Advisory vs consulting vs implementation
3. Judgment vs build; demo vs deployment
4. Why AI decisions are different
5. The three foundations
6. Pilot purgatory
7. From stated ask to real problem
8. Decision rights
9. Case study · templates · cheat sheet · knowledge check

Suggested study time · 3–4 hours

AI transformation advisory is the work of deciding **which AI problems** an enterprise should solve, **for whom**, **in what order**, and **under what rules of use**. Everything in this certificate builds on that sentence. This module teaches you to own it: to explain the role, to show why it creates value that neither strategy consultants nor engineers provide on their own, and to apply its first diagnostic moves in a real engagement.

KEY TAKEAWAYS

1. Advisory is a **decision** discipline, not a build discipline. Its output is a better, faster, defensible decision — not a working model.
2. AI decisions carry **structural uncertainty**: performance is probabilistic, data-dependent, hard to measure in advance, and the technology keeps shifting.
3. Production value rests on **three foundations**: data readiness, rules of use and governance decision rights, and an organization ready to act on an output.
4. Most enterprise AI initiatives do not fail — they **stall** between pilot and production because the non-technical decisions were never taken.
5. Always separate the **stated ask** from the **underlying problem**, sort advisory decisions from build decisions, and name **who decides what** before anything is scoped, funded or built.

1.1 Defining the AI advisory role

Every enterprise today has more AI ideas than it can fund, more vendors than it can evaluate and more enthusiasm than governance. The AI transformation advisor exists to turn that noise into a small number of good, owned, sequenced decisions. The definition has four load-bearing parts:

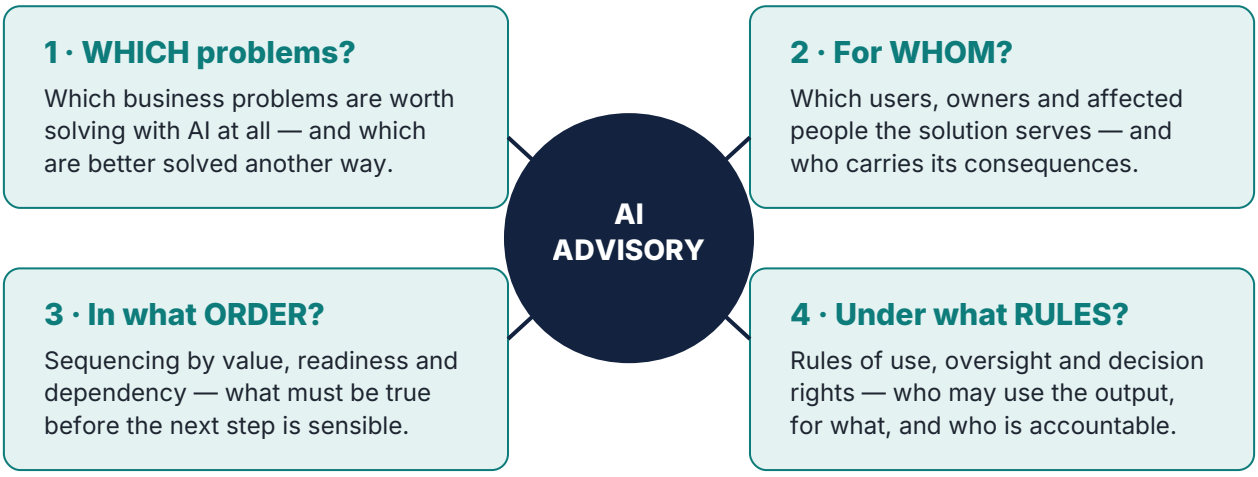


Figure 1.1 — The four questions that define AI transformation advisory

Element	What it means in practice	Typical advisor question
Which problems	Selecting the problems worth solving — including concluding that AI is the wrong tool.	"If this worked perfectly, what would change in the P&L or risk profile?"

Element	What it means in practice	Typical advisor question
For whom	Identifying users, process owners, beneficiaries and the people affected by outputs (customers, employees, citizens).	"Who acts on this output, and who lives with its mistakes?"
In what order	Sequencing by value, readiness and dependencies; deciding what to prove first and what to defer.	"What has to be true before step two makes sense?"
Under what rules of use	Defining permitted uses, human oversight, accountability, escalation and the limits of reliance on outputs.	"Who may rely on this output, for which decisions, and who signs off?"

DEFINITION · AI transformation advisory

The discipline of helping an enterprise decide *which* AI problems to solve, *for whom*, *in what order* and *under what rules of use* — and of making sure the people who must take those decisions are identified and equipped to take them.

The value the role creates

Advisory value is measured by the quality of decisions, not by the volume of output. Concretely, a good advisor creates value in five ways:

- **Avoided waste** — stopping or redirecting initiatives that would never reach production, before the money is spent.
- **Faster time-to-decision** — replacing months of circling debate with a framed choice, named owners and a clear evidence threshold.
- **Higher conversion from pilot to production** — by surfacing the data, governance and organizational prerequisites early.
- **Reduced risk** — legal, reputational, operational and vendor risk identified while it is still cheap to address.
- **Defensibility** — a decision record the board, regulator, auditor or works council can follow and challenge.

CONSULTANT'S TIP · Positioning in one breath

> *"I don't build your models and I don't write your corporate strategy. I help you decide which AI problems are worth solving, in what order, under what rules — and I make sure the right people own those decisions, so what you build actually reaches production."*

1.2 Advisory vs management consulting vs implementation

Clients often lump these three together, and so do many practitioners. Being able to separate them crisply is the first test of credibility. The roles are complementary — a single engagement may need all three — but they answer different questions, produce different outputs and fail in different ways.

Dimension	General management consulting	AI transformation advisory	AI implementation
Core question	Where should the business compete and how should it be organized?	Which AI problems should we solve, for whom, in what order, under what rules?	How do we build, integrate and run this AI system?

Dimension	General management consulting	AI transformation advisory	AI implementation
Primary output	Strategy, operating model, cost programme	Framed decisions, prioritized portfolio, rules of use, decision-rights map, evidence-based recommendations	Working models, pipelines, integrations, deployed software
Key expertise	Industry, finance, organization design	AI capability literacy + business framing + governance + change	Data science, ML/LLM engineering, MLOps, software engineering
Success looks like	Strategy adopted and executed	Right initiatives funded, wrong ones stopped, pilots convert to production	System meets spec, runs reliably in production
Typical failure	AI treated as a generic "digital" lever; no feasibility grounding	Drifting into building, or producing reports that enable no decision	Building the wrong thing well; no owner or rules for the output

EXAM TRAP · "Advisory is just consulting about AI"

Exam questions often offer this as a tempting option. The distinction is not the topic — it is the **structural uncertainty** of AI (Section 1.4) and the need for decisions about **rules of use and decision rights** that ordinary strategy work does not have to make. An advisor must also be literate enough in AI capabilities to judge feasibility, which a generalist strategist is not expected to be.

PITFALL · The drift into building

Advisors who are technically capable are often pulled into prototyping "just to show it's possible". The prototype becomes the deliverable, the decision questions go unasked, and the advisor is now an implementer without the mandate, budget or accountability of one. Building a quick demonstration can be a legitimate advisory tool — but only when it is explicitly designed to answer a decision question and labelled as a demonstration.

1.3 Advisory judgment vs technical build; demonstration vs deployment

Advisory judgment vs technical build

Advisory judgment is the application of evidence, context and experience to a decision under uncertainty: should we, which one, when, under what conditions? **Technical build** is the engineering work of making a chosen solution function: architecture, data pipelines, model selection and tuning, testing, integration, operations. Judgment determines *what is worth building*; build determines *whether it can be made to work*. Good advisors respect build expertise and know where their own judgment ends (see Module 4 on scope of authority).

Demonstration vs deployment

This is the single most important distinction in enterprise AI, and you will meet it again in Module 5. A **demonstration** shows that something *can* happen under chosen conditions. A **deployment** shows that it *does* happen, reliably, in the real process, with real data, real users and real accountability.

Attribute	Demonstration	Deployment
Data	Curated, cleaned, often hand-picked examples	Live, messy, drifting, includes edge cases

Attribute	Demonstration	Deployment
Users	Builders or enthusiastic volunteers	Ordinary users under time pressure, some sceptical
Volume	Tens or hundreds of cases	Thousands to millions; latency and cost matter
Integration	Standalone or mocked	Connected to systems of record, identity, logging
Failure handling	Failures quietly skipped or retried	Failures must be detected, escalated, explained
Governance	Usually none	Owner, rules of use, oversight, monitoring, audit trail
What it proves	Possibility	Value under operating conditions

> **KEY IDEA**
> A demonstration is evidence of **possibility**, not of **value**. The gap between the two is where most enterprise AI money is lost. Every claim you hear — internal or vendor — should be tagged with where on this spectrum its evidence came from.

1.4 Why AI decisions are different: structural uncertainty

Traditional IT decisions largely concern *deterministic* systems: if the specification is met, the system behaves as specified. Analytics decisions concern *descriptive* answers about the past. AI decisions are different in kind, not just degree. The uncertainty is **structural** — it comes from the nature of the technology and does not disappear with better project management.

Source of uncertainty	What it means	Consequence for advisory
Probabilistic outputs	The same system gives correct answers most of the time, wrong ones some of the time — and generative systems can vary between runs.	Decisions must specify acceptable error, who catches errors and what happens when they slip through.
Data dependency	Performance depends on the data the system meets in production, which may differ from training or test data, and drifts over time.	Pilot results do not transfer automatically; data readiness is a gating question, not a detail.
Performance is hard to know in advance	You often cannot know how well it will work on your task until you try it on your data.	Plans need learning milestones and decision gates, not fixed promises.
Non-obvious failure modes	Confident-sounding errors, bias against sub-groups, sensitivity to phrasing, unexpected tool actions.	Rules of use and human oversight must be designed, not assumed.
Moving technology frontier	Capabilities, prices and vendor offerings change within the life of a single engagement.	Avoid locking decisions to a specific product; decide on capabilities and keep options open.
Variable running cost	Usage-based costs (e.g., per-token inference) scale with adoption and prompt design.	Business cases need cost-per-transaction scenarios, not a single licence figure.
Socio-technical effects	Outputs change how people work, decide and are accountable.	Change management and decision rights are part of the solution, not an afterthought.

EXAMPLE · Same budget, different uncertainty

An ERP upgrade and an AI invoice-coding tool both cost €2m. The ERP project's main risk is delivery: will it be built on time to spec? The AI tool's main risk is *performance in context*: on this company's invoices, from these suppliers, in these languages, what fraction will be coded correctly, and what happens to the rest? The second question cannot be answered by a project plan — only by evidence gathered under realistic conditions.

1.5 The three foundations of enterprise AI adoption

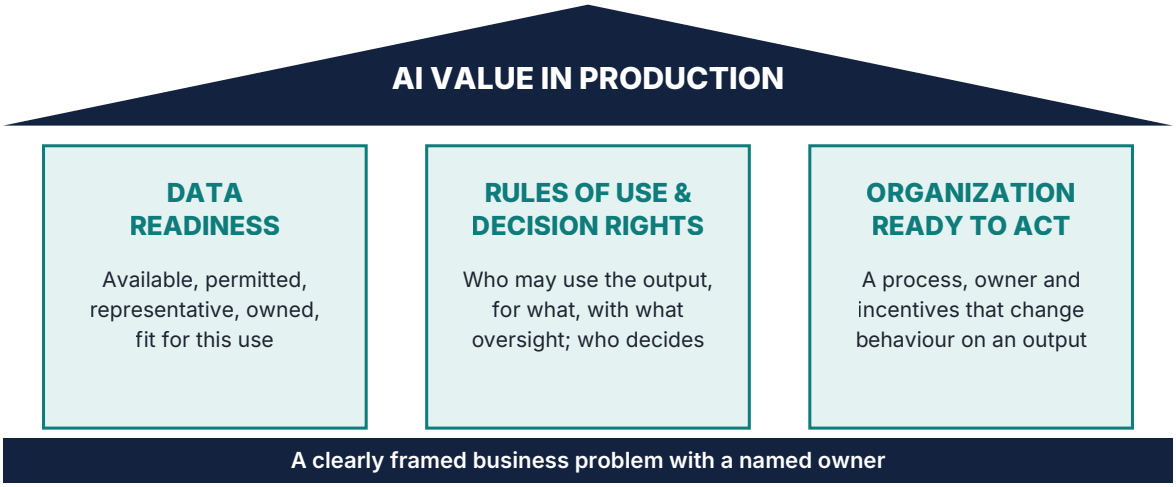


Figure 1.2 — The three foundations of enterprise AI adoption, resting on a framed, owned problem

Foundation 1 — Data readiness

Data readiness is not "do we have data?" but "do we have the *right* data, that we are *allowed* to use, of *sufficient quality*, that represents *production conditions*, with a *named owner*?" Five tests:

- **Availability** — does it exist, can it be accessed at the needed latency?
- **Permission** — legal basis, consent, contractual and confidentiality constraints (e.g., GDPR purpose limitation).
- **Quality** — completeness, accuracy, consistency, labelling where needed.
- **Representativeness** — does it reflect the cases the system will meet, including minorities and edge cases?
- **Ownership** — someone accountable for its upkeep and fitness for this purpose.

Foundation 2 — Rules of use and governance decision rights

Rules of use define what an AI output may be used for, by whom, with what level of human review and under which limits. **Governance decision rights** define who is entitled to approve a use case, accept residual risk, change the rules, and stop the system. Without both, legal and risk functions will — rightly — block go-live, usually late.

Foundation 3 — An organization ready to act on an output

A model that predicts machine failure is worthless if no maintenance planner is empowered, incentivised and scheduled to act on the prediction. Readiness to act means: a process redesigned around the output, an accountable owner, users trained and trusting (but not over-trusting) the system, and incentives that reward using it.

EXAM TRAP

Questions may describe a technically excellent model that nobody uses and ask which foundation is missing. If data and rules are in place but behaviour does not change, the answer is **organizational readiness to act** — not data, and not model accuracy.

1.6 Pilot purgatory: why AI stalls rather than fails

Enterprise AI initiatives rarely end in a clear failure. Instead they stall: the pilot "worked", everyone agrees it was promising, and twelve months later it is still a pilot. Stalling is worse than failing, because it consumes attention, blocks the budget and produces no learning. Understanding the mechanism is a core advisory skill.

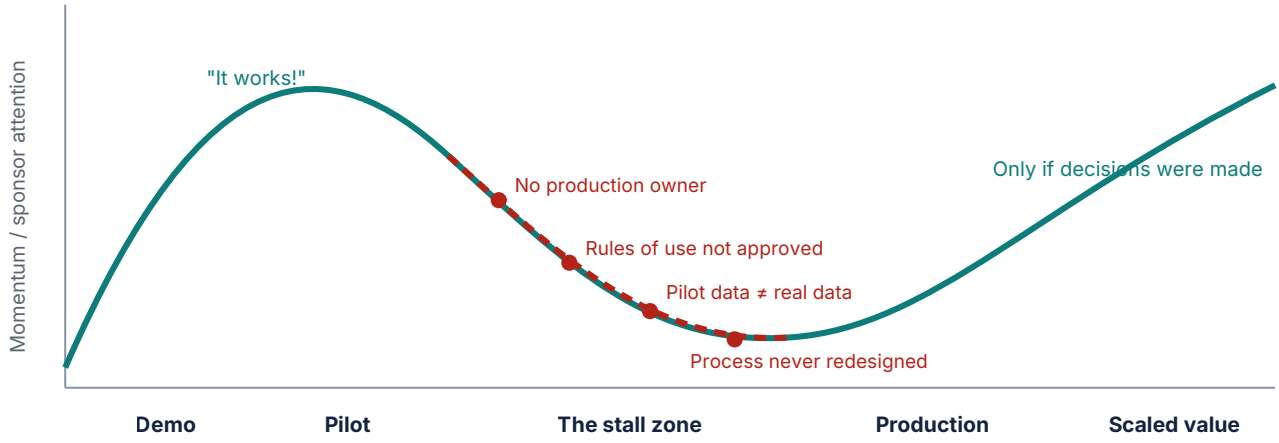


Figure 1.3 — The pilot-to-production stall zone

Stall cause	What it looks like	Advisory prevention
Pilot designed to prove possibility, not value	Success criterion was "it works"; no baseline, no business metric.	Define the production decision and its evidence threshold before the pilot starts.
No production owner	Innovation team owns the pilot; no business unit agrees to run it.	Name the production owner as a condition of funding the pilot.
Rules of use never approved	Legal/risk reviews start at go-live; open questions multiply.	Involve risk, legal, privacy and (where relevant) works council at framing.
Pilot data ≠ production data	Curated sample performed well; real feed is messier.	Test on representative, untidied data as early as possible.
Process not redesigned	Output is "available" but nobody's workflow changes.	Design the to-be process and role changes alongside the model.
Integration and run cost underestimated	Security, identity, logging, monitoring, support add months and cost.	Include run-state requirements and cost in the business case.
Sponsor attention moves on	The next exciting demo gets the budget.	Phase gates with explicit go / no-go / pivot decisions and dates.

KEY IDEA

Stalls are caused by **undecided decisions**, not by bad models. The advisory remedy is to bring those decisions forward — into framing and scoping — where they are cheap.

1.7 From the stated ask to the underlying business problem

Sponsors rarely arrive with a problem. They arrive with a **solution-shaped ask**: “we need a chatbot”, “we want to use agents”, “can you build us a copilot?”. The ask is real information — it tells you about pressure, ambition and politics — but it is not yet a problem. Advisory begins by holding the solution loosely and working back to the problem.

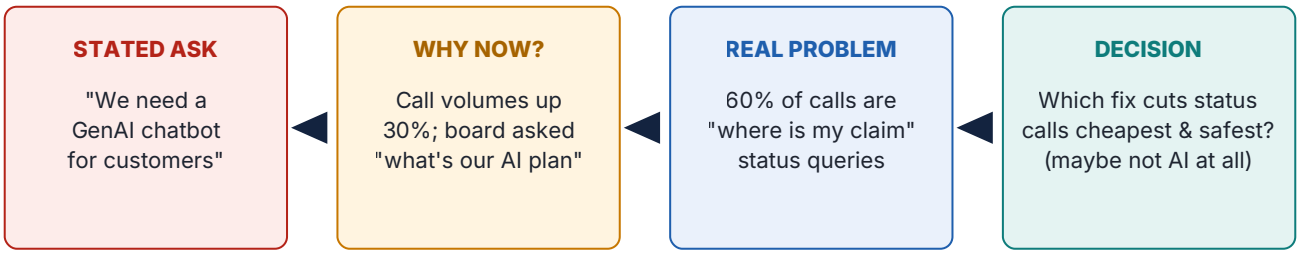


Figure 1.4 — Working back from a stated ask to a decision

The reframing sequence

1. **Capture the ask verbatim.** Do not correct it yet; record who asked and in what words.
2. **Ask “why now?”** What triggered it — a board question, a competitor, a cost spike, a vendor pitch?
3. **Find the business outcome.** Which metric, cost, risk or experience would change if this succeeded? By how much would make it worthwhile?
4. **Locate the problem in a process.** Where exactly, and for whom, does the pain occur? What is the baseline?
5. **Widen the solution space.** List non-AI and simpler options (process fix, rules engine, self-service portal, training). AI must beat them.
6. **State the decision.** Write the question the sponsor actually needs answered, who must answer it and by when.

EXAMPLE · Five reframes

Stated ask	Underlying problem (after reframing)
“We need a GenAI chatbot.”	Customer status queries overwhelm the contact centre and drive cost and dissatisfaction.
“Let’s use AI for HR.”	Time-to-hire for technical roles is 90 days and losing candidates to competitors.
“We want an agentic AI strategy.”	The board wants to know whether competitors’ automation claims threaten margin.
“Build us a forecasting model.”	Inventory write-offs of perishable goods are 4% of revenue.
“Give everyone Copilot.”	Knowledge workers lose hours locating documents; also, staff already use unsanctioned tools.

CONSULTANT'S TIP · Language that works

> *"That could be exactly right. Before we commit to the chatbot, help me understand what success would look like a year from now — what number would have moved?"*

This respects the sponsor's idea while opening the problem. Never open with "that's the wrong question".

1.8 Advisory decisions, build decisions, and who decides what

Once the problem is clear, sort the remaining work. **Advisory decisions** are choices about whether, what, for whom, in what order and under what rules. **Build decisions** are choices about how to make the chosen solution work. Mixing them causes two classic failures: engineers making business-risk decisions by default, and executives making technical choices they are not equipped to make.

Advisory decisions (whether / what / rules)	Build decisions (how)
Is this problem worth solving, and is AI the right approach?	Which model family, architecture or vendor component fits the requirements?
Which use cases go first, and why?	How should data pipelines, retrieval and integrations be designed?
What error rate is acceptable, and what happens to errors?	How do we measure and monitor that error rate technically?
Who may rely on the output, for which decisions, with what oversight?	How is human review implemented in the user interface and workflow?
Build, buy or partner — strategically?	Which specific product, configuration and deployment pattern?
What evidence is required to move from pilot to production?	How is the test harness, evaluation set and logging designed?

Naming who must decide what — before scoping, funding or building

For each material decision, name a single accountable decision-maker and the parties who must be consulted. Typical roles in a European enterprise:

Role	Typically decides / approves	Typically consulted on
Executive sponsor	Strategic fit, funding envelope, go/no-go at major gates	Priorities, risk appetite
Business process owner	Whether and how the process changes; acceptance of output into operations	Error tolerance, user impact
Data owner / steward	Whether data may be used for this purpose and on what terms	Quality, retention, access
Risk, compliance & legal	Risk classification, regulatory obligations, acceptable residual risk	Rules of use, vendor contracts
Data Protection Officer	Advises on GDPR; DPIA where required (advises, does not own the business decision)	Personal data processing
CISO / security	Security controls, vendor security approval	Data flows, third-party access
IT / enterprise architecture	Platform and integration standards	Build vs buy, run cost

Role	Typically decides / approves	Typically consulted on
Employee representatives (e.g., works council)	Co-determination rights where AI monitors or affects employees (jurisdiction-specific)	Rollout, monitoring, training
Finance	Business case acceptance, cost model	Run-cost scenarios

NOTE · Jurisdiction matters

In Germany, for example, the works council has co-determination rights over technical systems capable of monitoring employee behaviour or performance (Works Constitution Act, §87(1) no. 6). The EU AI Act assigns obligations to *providers* and *deployers* of AI systems according to risk class. The advisor's job is to make sure these parties are on the decision map early — not to interpret the law (see Module 4, scope of authority).

PITFALL · The phantom decision-maker

"The steering committee will decide" usually means nobody will. Committees advise; a named individual decides. If you cannot write a person's role next to a decision, that decision is your engagement's biggest risk.

1.9 Case study: Northbridge Insurance

CASE STUDY **"We need a claims chatbot by Q3"**

Northbridge Insurance (fictional) · mid-size European P&C insurer · 4,000 staff

Situation

The COO of Northbridge returns from an industry conference and asks the innovation team for "a GenAI chatbot for claims customers, live by Q3". A vendor has already demonstrated a slick prototype answering questions about policy coverage. The innovation lead engages you as advisor to "help get this done quickly".

What you find in the first two weeks

- Contact-centre call volume rose 28% in a year. Analysis of call reasons shows 57% are "where is my claim?" status queries; only 9% are coverage questions — the scenario the vendor demonstrated.
- Claim status lives in a legacy claims system with a nightly batch export; there is no real-time API.
- Legal has not been consulted. Coverage answers could be construed as binding statements about policy terms.
- No business unit has agreed to own the chatbot after launch. The contact-centre director was not invited to the vendor demo.
- The works council has asked whether conversation logs will be used to evaluate agents.

The advisory moves

1. **Reframe:** the real problem is status-query volume, not coverage questions. A proactive status notification (SMS/e-mail on claim milestones) — no AI required — might remove a large share of calls.
2. **Sort decisions:** advisory decisions (Which problem first? May an AI answer coverage questions at all? Who owns it?) are separated from build decisions (real-time API vs batch, vendor selection).
3. **Map decision rights:** COO (sponsor, funding), contact-centre director (process owner — now invited), Head of Legal (rules of use for coverage answers), DPO, CISO, works council (log usage).
4. **Recommend a sequence:** (1) proactive status notifications; (2) an AI assistant for *agents* (not customers) on coverage questions, with mandatory human review; (3) revisit a customer-facing assistant once rules of use are approved and status traffic has fallen.

Outcome

Status notifications launched in 10 weeks and cut status calls substantially. The agent-assist pilot had a named owner and legal-approved rules of use from day one, and moved to production at its first gate. The COO got a visible AI success and a lower call volume — the outcome the original ask was really about.

Discussion questions

1. Which of the three foundations was weakest at the start, and what evidence shows it?
2. Why would building the vendor's coverage chatbot quickly have been likely to stall rather than fail?
3. Which decisions in this case were advisory decisions and which were build decisions?
4. How would you explain to the COO, in two sentences, why the first step is not AI?

Model answers (brief)

1. Organizational readiness and rules of use were weakest: no production owner, contact-centre director absent, legal not consulted. Data readiness was also limited (batch-only status data).
2. The demo would likely have “worked”, generating enthusiasm, but go-live would have been blocked by legal review of coverage answers, lack of an owner and works-council questions — a classic stall.
3. Advisory: which problem first, whether AI may answer coverage questions, who owns it, sequencing. Build: real-time API design, vendor product, retrieval approach.
4. “Most of the calls you want to remove are status questions we can answer automatically without AI in ten weeks. That frees capacity and buys time to put an AI assistant in front of your agents safely, with legal sign-off, as step two.”

1.10 Consulting templates

Use these templates in the first fortnight of any AI engagement. They are deliberately short: each should fit on one page and be completed in a working session with the sponsor.

T1.1 Ask-to-Problem Reframing Canvas

Purpose: Turn a solution-shaped ask into a decision-ready problem statement. **Use when:** first sponsor meeting; any new AI request

1 · THE ASK

Stated ask (verbatim): .....

Requested by / date: .....

Trigger — why now? .....

2 · THE OUTCOME

Business metric that would move: .....

Current baseline: .....

Change that would make this worthwhile: .....

By when: .....

3 · THE PROBLEM

Where in the process does the pain occur? .....

Who experiences it? Who acts on it? .....

Evidence (data, interviews): .....

4 · SOLUTION SPACE

Option (incl. non-AI)	Expected impact	Cost / effort	Key risk
Proactive status SMS	*High on status calls*	*Low*	*Data latency*

5 · THE DECISION

Decision the sponsor must take: .....

Decision owner: .....

Evidence needed to decide: .....

Decision date: .....

T1.2 Three Foundations Quick Check

Purpose: Rapid red/amber/green assessment of whether a use case can reach production. **Use when:** before funding a pilot; at each phase gate

Foundation	Test question	R / A / G	Evidence	Action / owner
Data — availability	*Can we access the data at required latency?*			
Data — permission	*Do we have a legal/contractual basis for this use?*			
Data — quality & representativeness	*Does it reflect production cases, incl. edge cases?*			
Data — ownership	*Is a named owner accountable for fitness?*			
Rules of use	*Are permitted uses, oversight and limits written down?*			
Decision rights	*Is it clear who approves, accepts risk, can stop it?*			
Org — owner	*Has a business owner agreed to run it in production?*			
Org — process	*Is the to-be process designed around the output?*			
Org — people	*Are users trained, consulted, incentivised?*			

T1.3 Decision Rights Register

Purpose: Name who must decide what before an AI initiative is scoped, funded or built. **Use when:** engagement start; refresh at every gate

Decision	Type (Advisory / Build)	Accountable (one name)	Must consult	Inform	Needed by
May an AI answer coverage questions to customers?	*Advisory*	*Head of Legal*	*COO, DPO, CC Director*	*Works council*	*Gate 1*

T1.4 Advisory vs Build Decision Sort

Purpose: Separate decisions the advisor facilitates from those owned by technical teams.

Open question	Advisory decision?	Build decision?	Who answers	Advisor's role (lead / facilitate / refer)

1.11 Best practices and pitfalls

✓ DO THIS

- Open every engagement by reframing the ask into a business problem with a baseline.
- Tag every piece of evidence as demo, pilot or production.
- Make a named production owner a condition of pilot funding.
- Bring legal, risk, privacy and employee representatives in at framing.
- Write the go/no-go criteria for a pilot before it starts.
- Decide on capabilities and problems, not on products.
- Include non-AI options in every solution space.

× AVOID THIS

- Accepting "build us X" as the problem statement.
- Building a prototype that answers no decision question.
- Treating a successful demo as evidence of production value.
- Letting "the steering committee" be the decision-maker.
- Interpreting law or designing model architecture yourself.
- Measuring pilot success as "it works".
- Promising fixed performance figures before testing on real data.

1.12 Module 1 cheat sheet

MODULE 1 · ONE-PAGE SUMMARY

What AI Transformation Advisory Is

DEFINITION

Deciding **which** AI problems to solve, **for whom**, **in what order**, **under what rules of use** — and making sure the right people own those decisions.

THREE ROLES

- **Mgmt consulting**: where to compete, how to organize
- **AI advisory**: which AI decisions, in what order, under what rules
- **Implementation**: how to build and run it

DEMO VS DEPLOYMENT

- Demo = possibility, curated data, no governance
- Deployment = value, live data, real users, owner, oversight, monitoring

STRUCTURAL UNCERTAINTY

- Probabilistic outputs
- Data-dependent, drifting performance
- Unknowable in advance
- Non-obvious failure modes
- Moving frontier · variable run cost

THREE FOUNDATIONS

- **Data readiness**: available, permitted, quality, representative, owned
- **Rules of use & decision rights**
- **Organization ready to act**: owner, process, people, incentives

WHY PILOTS STALL

- Proved possibility, not value
- No production owner
- Rules of use not approved
- Pilot data ≠ real data
- Process not redesigned · run cost ignored

ASK → PROBLEM

- Capture ask verbatim → why now?
- Business outcome + baseline
- Locate in process & people
- Widen options (incl. non-AI)
- State the decision, owner, date

ADVISORY VS BUILD

Advisory: whether, what, for whom, order, rules, evidence threshold. **Build**: how — model, pipelines, integration, test harness.

DECISION RIGHTS RULE

One accountable **named role** per decision. Committees advise; people decide. Map before scope, funding or build.

1.13 Knowledge check

Answer all questions before turning the page. Explanations for every option — right and wrong — follow.

K1 **A CIO asks what distinguishes your AI advisory role from the strategy firm already on site. Which answer is most accurate?**

(A) "We provide independent vendor selection."
(B) "We build working prototypes so the strategy can be validated."
(C) "We cover AI topics, while they cover general business topics."
(D) "We decide which AI problems to solve, for whom, in what order and under what rules of use — decisions shaped by AI's structural uncertainty."

K2 **A vendor shows a model that summarises 50 hand-picked contracts flawlessly. What has been demonstrated?**

(A) That data readiness is confirmed
(B) Production value
(C) Possibility under chosen conditions
(D) Readiness for deployment pending a security review

K3 **Which statement best explains why AI decisions carry *structural* uncertainty?**

(A) Outputs are probabilistic and depend on data the system meets in production, so performance often cannot be known until tested on real data.
(B) AI vendors are newer companies with less track record.
(C) AI projects are usually managed with agile methods, which are less predictable.
(D) AI is more expensive than traditional software.

K4 **A churn-prediction model is accurate, the data is well governed and its use has been approved, yet churn has not changed after six months. Which foundation is most likely missing?**

(A) Model accuracy
(B) Organizational readiness to act on the output
(C) Rules of use
(D) Data readiness

K5 **Why do enterprise AI initiatives typically *stall* rather than fail?**

(A) Because cloud capacity is insufficient for production volumes.
(B) Because the models usually stop working after the pilot.
(C) Because pilots prove possibility while decisions on ownership, rules of use, data and process are left untaken, so nothing blocks the project outright but nothing lets it proceed.
(D) Because sponsors deliberately avoid production to limit risk.

K6 **A sponsor says, "We want an agentic AI strategy." What is the best first advisory move?**

(A) Ask what triggered the request and what business outcome would change if it succeeded.
(B) Shortlist three agent platform vendors for evaluation.
(C) Explain that agentic AI is overhyped and suggest traditional ML instead.
(D) Present a reference architecture for agentic systems.

K7 **Which of the following is an advisory decision rather than a build decision?**

(A) Choosing the embedding model for a retrieval pipeline
(B) Selecting the GPU instance type for inference
(C) Designing the logging schema for model outputs
(D) Deciding what error rate is acceptable and who handles errors

K8 **In a decision-rights register, "Steering committee" is entered as accountable for approving go-live. What is the best advice?**

(A) Replace it with the advisor, who is the most neutral party.
(B) Replace it with a single named role who decides, with the committee as consulted.
(C) Leave it; committees provide balanced governance.
(D) Remove the decision from the register until later.

1.14 Knowledge check: answers and explanations

K1 **Answer: D** Review: 1.1–1.2

A CIO asks what distinguishes your AI advisory role from the strategy firm already on site. Which answer is most accurate?

✗ (A) "We provide independent vendor selection." — **Wrong.** Vendor due diligence is one advisory activity (Module 5), not the definition of the role.

✗ (B) "We build working prototypes so the strategy can be validated." — **Wrong.** Building is implementation. A prototype may be a tool of advisory, but it does not define the role.

✗ (C) "We cover AI topics, while they cover general business topics." — **Wrong.** Topic alone is not the distinction. Strategy firms discuss AI too; what differs is the type of decision and the structural uncertainty handled.

✓ (D) "We decide which AI problems to solve, for whom, in what order and under what rules of use — decisions shaped by AI's structural uncertainty." — **Correct.** This is the module's definition, and it names what generalist strategy work does not have to address: rules of use and AI-specific uncertainty.

K2 **Answer: C** Review: 1.3

A vendor shows a model that summarises 50 hand-picked contracts flawlessly. What has been demonstrated?

✗ (A) That data readiness is confirmed — **Wrong.** The data was curated by the vendor; it says nothing about the client's production data.

✗ (B) Production value — **Wrong.** Production value requires live data, real users, integration and governance — none present here.

✓ (C) Possibility under chosen conditions — **Correct.** Hand-picked examples and a vendor-run session show only that the task *can* be done under favourable conditions.

✗ (D) Readiness for deployment pending a security review — **Wrong.** Security is only one gap; representativeness, volume, ownership and rules of use are also untested.

K3 **Answer: A** Review: 1.4

Which statement best explains why AI decisions carry structural uncertainty?

✓ (A) Outputs are probabilistic and depend on data the system meets in production, so performance often cannot be known until tested on real data. — **Correct.** This is the core of structural uncertainty: it comes from the technology itself and does not disappear with better project management.

✗ (B) AI vendors are newer companies with less track record. — **Wrong.** Supplier risk is real (Module 5) but it is not what makes AI decisions structurally different.

✗ (C) AI projects are usually managed with agile methods, which are less predictable. — **Wrong.** Delivery method is not the source; even perfectly managed AI projects face performance uncertainty.

✗ (D) AI is more expensive than traditional software. — **Wrong.** Cost may be variable, but price level is not what makes the uncertainty structural.

K4 **Answer: B** Review: 1.5

A churn-prediction model is accurate, the data is well governed and its use has been approved, yet churn has not changed after six months. Which foundation is most likely missing?

✗ (A) Model accuracy — **Wrong.** The model is accurate; accuracy without action produces no value.

✓ (B) Organizational readiness to act on the output — **Correct.** Nothing changes because nobody acts on predictions — no owner, process or incentive turns an output into a retention action.

✗ (C) Rules of use — **Wrong.** Use has been approved, so rules of use exist.

✗ (D) Data readiness — **Wrong.** The scenario states the data is well governed; nothing suggests data is the issue.

K5 **Answer: C** Review: 1.6

Why do enterprise AI initiatives typically stall rather than fail?

✗ (A) Because cloud capacity is insufficient for production volumes. — **Wrong.** Infrastructure can be a factor but is seldom the main cause, and would usually show as a defined blocker.

✗ (B) Because the models usually stop working after the pilot. — **Wrong.** Model failure would produce a clear failure, not a stall.

✓ (C) Because pilots prove possibility while decisions on ownership, rules of use, data and process are left untaken, so nothing blocks the project outright but nothing lets it proceed. — **Correct.** Undecided non-technical decisions create the limbo characteristic of pilot purgatory.

✗ (D) Because sponsors deliberately avoid production to limit risk. — **Wrong.** Deliberate avoidance is rare; the stall is usually unintended.

K6 **Answer: A** Review: 1.7

A sponsor says, "We want an agentic AI strategy." What is the best first advisory move?

✓ (A) Ask what triggered the request and what business outcome would change if it succeeded. — **Correct.** Capturing the ask and asking "why now?" and "what outcome?" starts the reframing sequence.

✗ (B) Shortlist three agent platform vendors for evaluation. — **Wrong.** Vendor selection is premature until the problem and decision are framed.

✗ (C) Explain that agentic AI is overhyped and suggest traditional ML instead. — **Wrong.** This dismisses the sponsor and substitutes your solution for theirs — still solution-first.

✗ (D) Present a reference architecture for agentic systems. — **Wrong.** That jumps to a build answer before the problem is known.

K7 **Answer: D** Review: 1.8

*Which of the following is an **advisory** decision rather than a build decision?*

✗ (A) Choosing the embedding model for a retrieval pipeline — **Wrong.** This concerns how to make a solution work — a build decision.

✗ (B) Selecting the GPU instance type for inference — **Wrong.** This is infrastructure engineering — a build decision.

✗ (C) Designing the logging schema for model outputs — **Wrong.** This is technical design — a build decision, even though logs serve governance.

✓ (D) Deciding what error rate is acceptable and who handles errors — **Correct.** Acceptable error and error handling are about whether and under what rules to use the system, which makes this an advisory decision.

K8 **Answer: B** Review: 1.8

In a decision-rights register, "Steering committee" is entered as accountable for approving go-live. What is the best advice?

✗ (A) Replace it with the advisor, who is the most neutral party. — **Wrong.** The advisor facilitates decisions but must not own the client's business decisions.

✓ (B) Replace it with a single named role who decides, with the committee as consulted. — **Correct.** One accountable named role per decision is the rule; the committee keeps its advisory voice.

✗ (C) Leave it; committees provide balanced governance. — **Wrong.** Committees advise and consult; diffuse accountability tends to leave decisions untaken.

✗ (D) Remove the decision from the register until later. — **Wrong.** Postponing the naming of a decision-maker is how pilots end up stalling.

MODULE 2

Mapping the Enterprise AI Capability Landscape

Building a capability inventory that survives the next product cycle: defining what counts, finding the AI nobody sanctioned, scoring on independent axes, classifying gaps and talking about it with executives without overpromising.

YOU WILL BE ABLE TO

- Map AI by durable capability category rather than by product or vendor
- Set counting rules before building an AI capability inventory
- Extend the inventory to ungoverned AI, unsanctioned tools and tool sprawl
- Score capabilities on independent axes instead of one blended number
- Distinguish a capability inventory from a maturity or readiness assessment
- Classify every shortfall by its structural gap condition and rank gaps with multi-input prioritization
- Separate AI from rules-based automation and place systems in ML, DL, GenAI or agentic classes
- Discuss trade-offs, build vs buy and platform fit with executives without overpromising

IN THIS MODULE

1. Capabilities, not products
2. Define before you count
3. AI vs automation; the durable classes
4. Beyond the sanctioned: shadow AI and sprawl
5. Building the inventory
6. Scoring on independent axes
7. Inventory vs maturity assessment
8. Gap conditions and prioritization
9. Talking to executives
10. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4–5 hours

Before you can advise an enterprise on what AI to do next, you need an honest picture of what AI it already has. Most organizations cannot produce one: their lists are organized by vendor, miss the AI that employees adopted on their own, and summarise everything into a single "maturity score" that hides the real problems. This module teaches a capability-based mapping method that stays useful when products change, and the vocabulary to discuss the results with executives.

KEY TAKEAWAYS

1. Map by **durable capability** (what the organization can do) — not by product or vendor (which will change).
2. **Define what counts** as a capability, and the counting rules, *before* you count anything.
3. The real landscape includes **shadow AI**, unsanctioned tools, AI embedded in existing SaaS and tool sprawl.
4. Score on **independent axes**; a blended number hides which remedy is needed.
5. An **inventory** describes what exists; a **maturity/readiness assessment** judges organizational ability. Do not confuse them.
6. Classify each shortfall by its **structural gap condition**, then rank with **multiple visible inputs**.
7. Use precise class vocabulary — **rules-based automation, ML, DL, GenAI, agentic AI** — and never promise beyond the evidence.

2.1 Capabilities, not products

A vendor-organized list ("we have Product X, Product Y and three licences of Z") goes out of date with every product release, acquisition and renaming. It also hides overlap — three products may deliver the same capability — and gaps, because a capability nobody has bought does not appear on a product list at all.

DEFINITION · AI capability
A business-relevant ability to perform a **class of task** using AI, described independently of the tool that currently delivers it. Example: *"extract structured fields from unstructured documents"* is a capability; *"Vendor X Intelligent Document Processing v4"* is one current implementation of it.

Product-based entry (fragile)	Capability-based entry (durable)
Vendor A chatbot licence	Conversational self-service for customer queries
Copilot seats in sales team	Drafting and summarising business text
Vendor B IDP platform	Document information extraction
Data science team churn model	Customer behaviour prediction (churn propensity)
Vendor C "AI agent" for IT tickets	Autonomous multi-step task execution (IT service requests)

A practical capability taxonomy

Most enterprise AI capabilities fall into a manageable set of durable categories. Adapt names to the client, but keep categories stable over time so the map can be compared year on year.

Capability category	Examples of tasks	Typical class
Prediction & forecasting	Demand, churn, failure, credit risk	ML
Classification & routing	Ticket triage, document type, fraud flag	ML / DL
Perception (vision, speech)	Defect detection, transcription, ID verification	DL
Information extraction	Fields from invoices, contracts, forms	DL / GenAI
Search & knowledge retrieval	Enterprise search, Q&A over documents	DL / GenAI
Content generation & transformation	Drafting, summarising, translating, coding	GenAI
Optimization & recommendation	Pricing, scheduling, next-best-action	ML / optimization
Conversational interaction	Assistants for customers or staff	GenAI
Autonomous task execution	Multi-step workflows across systems via tools	Agentic AI

2.2 Define what counts — before counting anything

Two analysts inventorying the same company will produce wildly different numbers unless they agree on rules first. “We have 140 AI use cases” is meaningless if one team counted every chatbot prompt template and another counted only production models. **Set the counting rules first, publish them with the result, and keep them stable.**

Rule	Question it settles	Example decision
Boundary rule	What is AI for this inventory, and what is not?	Rules-based RPA and static dashboards are excluded; ML, DL, GenAI, agentic included
Unit of analysis	Is the unit a capability, a use case, a deployment or a tool?	Unit = capability; deployments and tools are attributes of it
Granularity	How specific is a capability?	“Document extraction” — not “invoice extraction for Supplier Group 3”
Lifecycle scope	Which stages count?	Production and pilots counted separately; ideas listed in a backlog, not the inventory
Sanction scope	Do unsanctioned tools count?	Yes — tagged as unsanctioned, never omitted
Embedded AI	Do AI features inside existing SaaS count?	Yes, when enabled and used on company data
De-duplication	When are two entries the same capability?	Same task class + same business purpose = one capability, many deployments

EXAM TRAP

If a question asks what to do *first* when asked to “count our AI use cases”, the answer is to **define the unit of analysis and the counting rules** — not to send a survey, buy a discovery tool or start interviewing.

2.3 AI vs rules-based automation; the durable classes

Executives and vendors use "AI" for everything from a macro to an autonomous agent. An advisor needs vocabulary precise enough to place any system in a durable class, because the class determines data needs, failure modes and governance requirements.

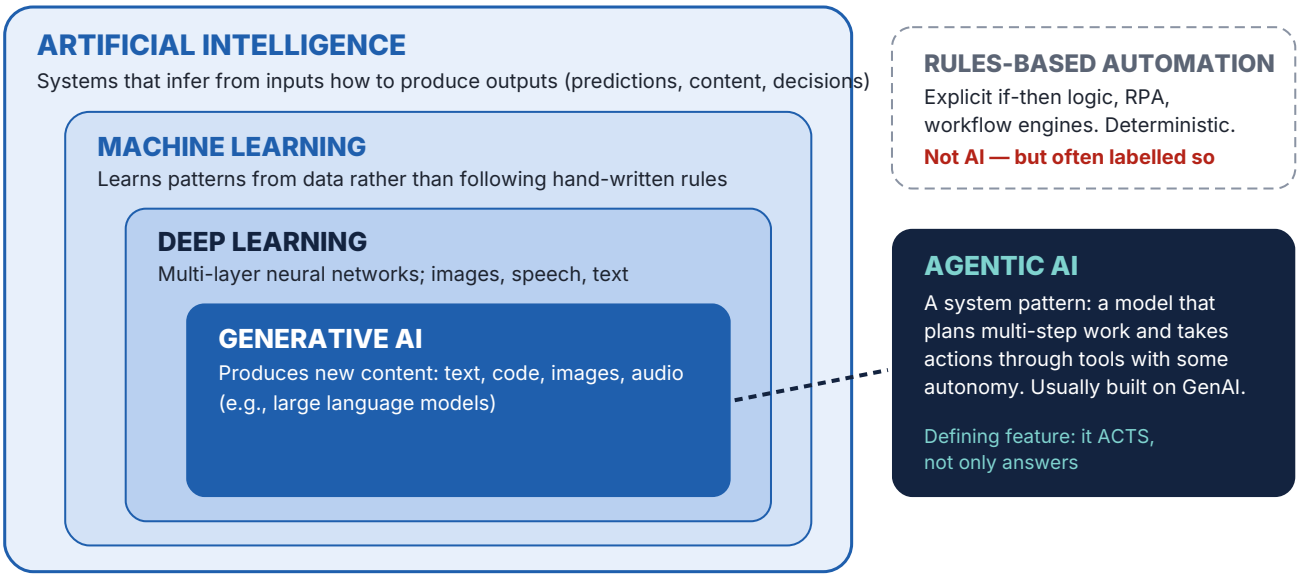


Figure 2.1 — The durable classes. Agentic AI is a system pattern layered on (usually generative) models; rules-based automation sits outside AI.

Class	How it works	Typical data need	Characteristic failure modes	Governance emphasis
Rules-based automation (not AI)	Executes explicit, human-written rules	None to learn from; rules must be maintained	Breaks on unforeseen cases; brittle but predictable	Change control of rules
Machine learning	Learns statistical patterns from historical, often labelled, data	Structured historical data with outcomes	Drift, bias from historical data, overfitting	Validation, monitoring, fairness
Deep learning	Multi-layer neural networks learn representations from large, often unstructured data	Large volumes of images, audio, text	Opaque reasoning, sensitivity to unusual inputs	Explainability limits, robustness testing
Generative AI	Large pre-trained models generate new content from prompts	Often little task-specific data; quality context (documents) matters	Plausible but false output ("hallucination"), prompt sensitivity, data leakage, IP issues	Rules of use, human review, content and data controls
Agentic AI	A model plans, calls tools and takes actions over multiple steps with some autonomy	Tool access, permissions, process knowledge	Compounding errors, unintended actions, security exposure through tools	Permission scoping, action logging, human checkpoints, kill switch

EXAMPLE · Placing systems in a class

System	Class	Why
Invoice approval routed by amount thresholds	Rules-based automation	Explicit if-then rules; nothing learned
Credit-default model trained on 10 years of loans	Machine learning	Learns patterns from labelled historical outcomes
Camera-based weld-defect detection	Deep learning	Neural network on image data
Assistant drafting replies to customer emails	Generative AI	Generates new text with an LLM
Assistant that reads a ticket, queries systems, resets access and closes the ticket	Agentic AI	Plans and takes actions through tools

PITFALL · "AI-washing"

Rebadging workflow automation or RPA as "AI" inflates the inventory and misdirects governance effort. Equally, under-labelling (treating an LLM feature as "just a plugin") lets high-risk use slip through. Classify by mechanism, not by marketing.

2.4 Beyond the sanctioned: ungoverned AI, unsanctioned tools and sprawl

The centrally sanctioned list is never the whole landscape. Generative AI in particular is available to anyone with a browser, and many enterprise SaaS products have switched on AI features through routine updates. An inventory that only lists what IT approved is a map of intentions, not of reality.

Category	Definition	Why it matters
Ungoverned AI adoption	AI in use with no owner, rules of use or risk review — sanctioned or not	Risk is accepted by nobody, by default
Unsanctioned tool use ("shadow AI")	Staff using AI tools that were not approved, often personal accounts	Data leakage, IP exposure, compliance breaches — and evidence of real demand
Embedded AI in existing software	AI features activated in CRM, office, HR or ITSM tools through updates	Enters without procurement; new data flows to vendors
AI tool sprawl	Many overlapping tools for the same capability across units	Duplicate cost, inconsistent controls, fragmented data and skills

Discovery techniques

Source	What it reveals	Watch-out
Expense and card data	Individually purchased AI subscriptions	Misses free tiers
Network, proxy, CASB, SSO logs	Traffic to AI services; sanctioned app sign-ins	Requires security partnership; privacy/ works-council sensitivity
Procurement and contract records	Licensed tools and AI clauses in renewals	Embedded features rarely itemised

Source	What it reveals	Watch-out
SaaS admin consoles	Which AI features are enabled	Needs each platform owner
Anonymous staff survey	Tools used, tasks, perceived value	Self-reporting bias; keep it non-punitive
Interviews and workshops	Workarounds, real workflows, unmet needs	Time-intensive; sample deliberately

CONSULTANT'S TIP · Make discovery safe

Frame discovery as learning, not policing — an "AI amnesty". Unsanctioned use is your best evidence of where demand and value lie. If people fear punishment, they will hide usage and your inventory will be wrong. Check with HR, legal and employee representatives before analysing logs that could identify individuals.

2.5 Building the AI capability inventory

The inventory is a register with one row per capability and attributes describing each of its deployments. Keep it lightweight enough to maintain — an inventory that is accurate once and never updated quickly becomes misleading.

Field group	Fields
Identity	Capability ID · capability name · category · class (ML/DL/GenAI/agentic)
Deployments	Business units · tools/vendors · lifecycle stage (pilot/production) · sanction status · users (approx.)
Ownership	Business owner · technical owner · data owner(s)
Use & risk	Purpose · decisions it informs or makes · people affected · personal data? · regulatory risk class (if assessed)
Scores (independent axes)	Value · performance evidence · data readiness · governance coverage · adoption · risk exposure
Gap	Gap condition · notes · proposed action
Housekeeping	Source of information · date verified · next review

2.6 Scoring on independent axes

The temptation is to give each capability a single score ("maturity 3.2 out of 5"). Resist it. A blended number averages strengths against weaknesses and conceals which problem needs solving. Score each axis separately, define each level in observable terms, and only combine for a specific purpose with visible weights.

**The blended score rates the first two capabilities identically (3.2) — yet one has an adoption problem and the other a governance prob
They need completely different remedies. Independent axes keep that visible.**

Figure 2.2 — Why blending hides the remedy (scores 1 = weak, 4 = strong)

Axis	Question	1 (weak)	4 (strong)
Business value	How much value does it create or could it create?	Marginal or unmeasured	Material, measured against a baseline
Performance evidence	How good is it, on what evidence?	Vendor claim or demo only	Measured in production on representative data
Data readiness	Is data available, permitted, fit, owned?	Major gaps or no legal basis	Fit, governed, owned
Governance coverage	Owner, rules of use, oversight, monitoring?	None	Complete and operating
Adoption	Is it actually used as intended?	Little or no use	Embedded in the process
Risk exposure	What harm could errors cause? (scored inversely)	Severe, poorly controlled	Low or well controlled

KEY IDEA

Axes are **independent** when a change in one does not mechanically change another. Value and governance can move separately — that is exactly why each must be visible. Define every level with observable evidence so two assessors would give the same score.

2.7 Inventory vs maturity or readiness assessment

	AI capability inventory	AI maturity / readiness assessment
Question	What AI capabilities exist, where, in what state?	How able is the organization to adopt and scale AI?
Unit	Individual capabilities and deployments	The organization (or a unit) as a whole
Nature	Descriptive, evidence-based register	Evaluative, against a staged model or criteria
Typical content	Rows per capability; owners; scores; gaps	Dimensions such as strategy, data, talent, governance, culture, technology
Output used for	Gap analysis, rationalisation, governance, prioritization	Setting organizational ambition, capability-building programme

	AI capability inventory	AI maturity / readiness assessment
Risk if confused	A maturity "level" is mistaken for knowledge of what is deployed	An inventory count is mistaken for organizational ability

EXAM TRAP

"We are at maturity level 3" tells you nothing about which tools are in use, who owns them or which ones expose the company to risk. Conversely, an inventory of 200 deployments does not show that the organization can *scale* AI. They are complementary instruments. If a question asks which one reveals unsanctioned tools, the answer is the **inventory**.

2.8 Gap conditions and multi-input prioritization

A "gap" is not one thing. Classifying each shortfall by its **structural gap condition** tells you what kind of remedy is required, and prevents the common error of answering every gap with "buy a tool".

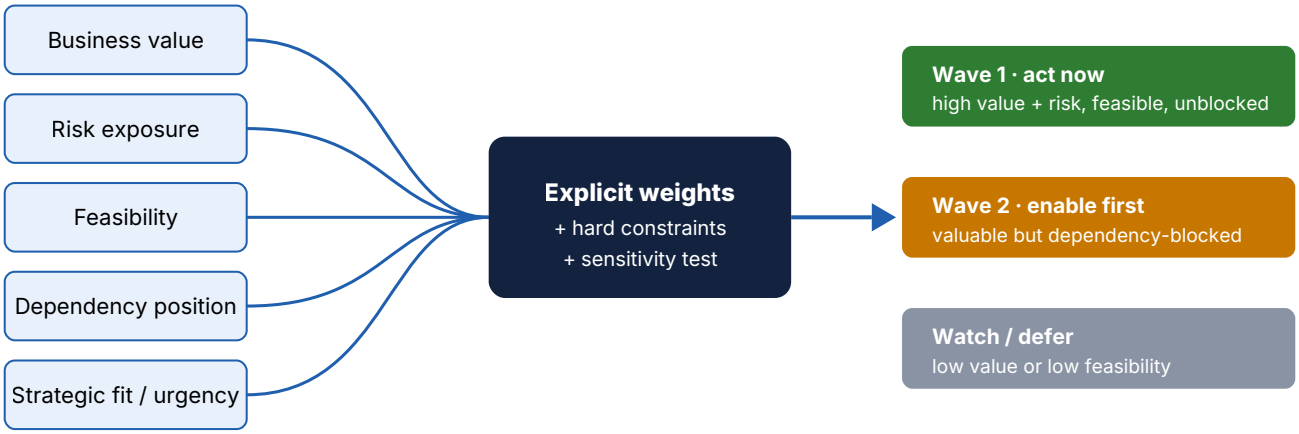
Gap condition	Description	Typical remedy
Absent	A needed capability does not exist anywhere	Build, buy or partner decision
Ungoverned	Exists and is used, but lacks owner, rules of use or oversight	Assign owner, define rules, risk review — not new technology
Underperforming	Exists but does not meet the required standard on evidence	Improve data/model, change vendor, or narrow use
Unadopted	Exists and works, but is not used as intended	Process redesign, training, incentives, UX
Fragmented (sprawl)	Duplicated across tools and units with inconsistent controls	Rationalise; choose a standard; migrate
Dependency-blocked	Cannot progress until a prerequisite (data, platform, policy, skill) is in place	Fix the dependency first; sequence accordingly

EXAMPLE

A bank finds four teams using three GenAI drafting tools, one with no data-processing agreement. That single finding contains two gap conditions: **fragmented** (three tools for one capability) and **ungoverned** (no agreement, no owner). The remedy is rationalisation plus governance — not another tool.

Multi-input prioritization

Rank gaps using several inputs kept visible side by side rather than a single opaque score. Use explicit weights agreed with the sponsor, apply hard constraints (e.g., regulatory deadlines override ranking), and test whether the ranking changes if weights move moderately.



Inputs are kept visible after ranking — the ranking is a conversation aid, not an oracle.

Figure 2.3 — Multi-input prioritization produces waves, not a single winner

Input	Why it matters
Business value	What is gained by closing the gap
Risk exposure	Ungoverned high-risk use may need action regardless of value
Feasibility	Effort, cost, data and skills required
Dependency position	Gaps that unblock others go earlier, even if their own value is modest
Strategic fit and urgency	Alignment with strategy; external deadlines (regulation, contracts)

PITFALL · False precision

A ranked list with scores to two decimal places invites the executive to trust the arithmetic instead of debating the trade-offs. Present ranks with the input profile next to each, and say plainly which judgements drove the order.

2.9 Talking to executives: trade-offs, build vs buy, platform fit

Capability trade-offs to name explicitly

Trade-off	Executive framing
Accuracy vs explainability	"The more complex model is a little more accurate but we cannot easily explain individual decisions — does this use need explanations?"
Generality vs specialization	"A general model handles many tasks adequately; a specialized one handles this task better but only this task."
Speed vs cost vs quality	"Larger models answer better but cost more per transaction and respond slower."
Autonomy vs control	"Letting the agent act directly saves time; requiring approval keeps a human accountable."
Time-to-value vs lock-in	"Buying gets us live faster; it also ties us to this vendor's roadmap and pricing."

Build, buy or partner

Criterion	Favours BUY	Favours BUILD	Favours PARTNER
Differentiation	Commodity capability	Core source of advantage	Advantage needs outside expertise
Data	Generic data is enough	Proprietary data is the edge	Shared or ecosystem data
Speed needed	Fast	Can afford longer horizon	Medium
Internal skills	Limited	Strong and retainable	Partial, building up
Control & compliance	Vendor controls acceptable	Need full control	Contractually shared
Total cost of ownership	Lower up front; recurring fees	Higher up front; maintenance burden	Shared; complex contracts

Platform fit asks whether a capability should run on the organization's existing data and AI platform, on a vendor's platform, or standalone — judged by integration with systems of record, identity and access, security and logging standards, skills, and the cost of operating yet another platform.

CONSULTANT'S TIP · Speaking without overpromising

Tie every performance statement to its evidence and conditions:

> *"In the pilot on 2,000 of our own invoices, it extracted the key fields correctly about 9 times in 10. We have not yet tested handwritten or non-German invoices, so I would plan for lower performance there until we do."*

Avoid: "It's 95% accurate", "It will replace the team", "Agents will run the process end-to-end".

2.10 Case study: Meridian Logistics

CASE STUDY **"How much AI do we actually have?"**

Meridian Logistics (fictional) · European freight forwarder · 9,000 staff in 14 countries

Situation

Meridian's CFO wants a single slide for the board: "our AI footprint and maturity". IT's list shows 12 sanctioned AI tools. You are asked to validate it within six weeks.

What you do

1. **Agree counting rules** with the CIO and CFO: unit = capability; production and pilot counted separately; unsanctioned and embedded AI included and tagged; RPA excluded.
2. **Discover** using expense data, proxy logs (aggregated, approved by the works council), SaaS admin consoles, a non-punitive survey and 20 interviews.
3. **Catalogue**: 47 tools resolved into 11 capabilities. Five tools on IT's "AI" list were rules-based workflow automation — excluded. Document extraction was delivered by six different tools across countries. GenAI drafting was used by ~2,300 staff, 70% through personal accounts.
4. **Score** each capability on six independent axes. Route optimization scored high on value and performance but low on adoption: dispatchers overrode it routinely because it ignored driver rest constraints they cared about.
5. **Classify gaps**: GenAI drafting — ungoverned and fragmented; document extraction — fragmented; route optimization — unadopted; customs classification — absent; demand forecasting — dependency-blocked (no harmonised shipment master data).
6. **Prioritise** with five visible inputs. Wave 1: govern GenAI drafting (risk) and fix route-optimization adoption (value already paid for). Wave 2: harmonise master data (unblocks forecasting). Watch: customs classification pending a build-vs-buy review.

The board slide

Instead of a maturity number, the CFO presented: *"11 AI capabilities, 4 in production; our largest exposure is ungoverned GenAI use by a quarter of staff; our quickest value is making the routing tool we already own usable by dispatchers."*

Discussion questions

1. Why did the count fall from 47 tools to 11 capabilities — and why does that matter to the board?
2. Why were five tools excluded, and what risk would including them have created?
3. Route optimization had a high blended score. What would a single score have hidden?
4. Why is master-data harmonisation in Wave 2 even though it creates no AI value directly?

Model answers (brief)

1. Counting rules made capability the unit; many tools delivered the same capability. The board needs to understand capabilities and exposure, not licence counts.
2. They were rules-based automation. Including them would inflate the AI footprint (AI-washing) and misdirect governance effort.

3. That the problem is adoption, not technology — which calls for process and UX change with dispatchers rather than a new tool.
4. It is a dependency: forecasting is blocked until it is fixed, so its dependency position earns it priority.

2.11 Consulting templates

T2.1 Inventory Counting Rules Sheet

Purpose: Agree and publish what counts before counting anything. **Use when:** day 1 of any landscape assessment

Boundary — included classes: .....

Boundary — excluded (e.g., RPA, BI): .....

Unit of analysis: .....

Granularity rule + example: .....

Lifecycle stages counted: .....

Unsanctioned tools — how tagged: .....

Embedded SaaS AI — rule: .....

De-duplication rule: .....

Approved by / date: .....

T2.2 AI Capability Inventory Register

Purpose: One row per capability; deployments as attributes.

ID	Capability	Category / class	Deployments (unit · tool · stage)	Sanctioned?	Business owner	Personal data?	Gap condition
C-04	*Document info extraction*	*Extraction / DL+GenAI*	*DE·Tool A·prod; PL·Tool B·pilot; …*	*Partly*	*Head of Ops*	*Yes*	*Fragmented*

T2.3 Independent-Axis Scorecard

Purpose: Score each capability on separate axes with evidence; do not blend.

Capability	Value 1–4	Performance evidence 1–4	Data readiness 1–4	Governance 1–4	Adoption 1–4	Risk exposure 1–4	Evidence source

SCORING DISCIPLINE

- ☐ Each level defined in observable terms
- ☐ Evidence source recorded for every score
- ☐ Two assessors calibrated on three sample capabilities
- ☐ No column labelled “overall”

T2.4 Gap Classification & Prioritization Matrix

Purpose: Classify each gap structurally and rank using visible inputs.

Gap	Condition (Absent / Ungoverned / Underperf. / Unadopted / Fragmented / Dep.-blocked)	Value	Risk	Feasibility	Unblocks…	Hard constraint?	Wave

Weights agreed with sponsor: .....

Sensitivity test result (does ranking change at ±20% weights?): .....

T2.5 Build / Buy / Partner Grid

Purpose: Structure the strategic sourcing conversation for one capability.

Criterion	Buy	Build	Partner	Notes
Differentiation				
Proprietary data advantage				
Time-to-value				
Internal skills				
Control & compliance				
5-year TCO				
Lock-in / exit				
Platform fit				

2.12 Best practices and pitfalls

✓ DO THIS

- Name capabilities in task language that will outlive the current product.
- Publish counting rules alongside every number.
- Include shadow and embedded AI, tagged — never silently omitted.
- Partner with security, HR and employee representatives on discovery.
- Score on independent axes with evidence sources.
- Classify gaps by condition before proposing remedies.
- Show the inputs behind every priority ranking.

× AVOID THIS

- Organizing the landscape by vendor.
- Counting before defining the unit of analysis.
- Treating unsanctioned use purely as a compliance breach.
- Reporting a single blended maturity score.
- Calling an inventory a maturity assessment (or vice versa).
- Answering every gap with “buy a tool”.
- Labelling RPA “AI” — or GenAI features “just plugins”.

2.13 Module 2 cheat sheet

MODULE 2 · ONE-PAGE SUMMARY

Mapping the Enterprise AI Capability Landscape

CAPABILITY, NOT PRODUCT

Capability = ability to perform a class of task with AI, described independently of tool. Products change; capabilities persist.

DEFINE BEFORE COUNTING

- Boundary · unit · granularity
- Lifecycle · sanction · embedded AI
- De-duplication rule

DURABLE CLASSES

- **Rules-based**: explicit rules, NOT AI
- **ML**: learns patterns from data
- **DL**: neural nets, unstructured data
- **GenAI**: generates content
- **Agentic**: plans + acts via tools

BEYOND SANCTIONED

- Ungoverned adoption
- Shadow AI (unsanctioned)
- Embedded SaaS AI
- Tool sprawl

Discover safely: amnesty framing.

INDEPENDENT AXES

Value · performance evidence · data readiness · governance · adoption · risk. **Never blend** — blends hide the remedy.

INVENTORY ≠ MATURITY

Inventory: what exists (descriptive, per capability). **Maturity/readiness**: how able the org is (evaluative, organization-wide).

GAP CONDITIONS → REMEDY

- Absent → build/buy/partner
- Ungoverned → owner + rules
- Underperforming → improve/replace
- Unadopted → process, training, UX
- Fragmented → rationalise
- Dependency-blocked → fix prerequisite

MULTI-INPUT PRIORITIZATION

Value · risk · feasibility · dependency · strategic fit/urgency. Explicit weights, hard constraints, sensitivity test → waves.

EXECUTIVE LANGUAGE

Name trade-offs (accuracy/explainability, general/specialised, cost/speed/quality, autonomy/control, speed/lock-in). Tie every number to its evidence and conditions. Build vs buy vs partner by differentiation, data, speed, skills, control, TCO.

2.14 Knowledge check

Answer all questions before turning the page.

K1 **Why should an AI landscape be mapped by capability rather than by product?**

(A) Because products cannot be scored.
(B) Because vendors refuse to disclose product details.
(C) Because capabilities are durable across product changes and reveal overlap and absence that product lists hide.
(D) Because capability maps are required by the EU AI Act.

K2 **A COO asks you to "count our AI use cases by Friday". What should you do first?**

(A) Send an all-staff survey immediately.
(B) Agree the boundary, unit of analysis and counting rules.
(C) Purchase a SaaS discovery tool.
(D) Ask IT for its list of approved tools.

K3 **An invoice workflow routes approvals using fixed amount thresholds and a supplier whitelist. How should it be classified?**

(A) Machine learning
(B) Generative AI
(C) Agentic AI, because it acts without a human
(D) Rules-based automation — outside the AI inventory boundary

K4 **What most distinguishes agentic AI from generative AI?**

(A) It plans multi-step work and takes actions through tools with some autonomy.
(B) It uses larger models.
(C) It is always fully autonomous with no human involvement.
(D) It does not need any data.

K5 **Two capabilities both have a blended score of 3.2. One has very low adoption; the other has very low governance coverage. What does this show?**

(A) They are equally mature and should be treated the same.
(B) The scoring scale is too narrow.
(C) Adoption should be weighted more heavily.
(D) Blended scores hide which remedy is needed; independent axes should be reported.

K6 **Which instrument would reveal that 2,000 employees use an unapproved GenAI tool?**

(A) An AI capability inventory that includes unsanctioned use
(B) An AI maturity assessment
(C) An AI readiness survey of executives
(D) A vendor benchmark report

K7 **A forecasting capability cannot progress until shipment master data is harmonised. Which gap condition applies?**

(A) Absent
(B) Dependency-blocked
(C) Underperforming
(D) Unadopted

K8 **An executive asks, "Is the extraction tool accurate?" Which response best avoids overpromising?**

(A) "It is state of the art, so it will be fine."
(B) "AI is never accurate, so we must keep humans."
(C) "On 2,000 of our invoices it got key fields right about 9 in 10 times; we haven't tested handwritten ones yet."
(D) "Yes — the vendor says 98%."

2.15 Knowledge check: answers and explanations

K1 **Answer: C** Review: 2.1

Why should an AI landscape be mapped by capability rather than by product?

- ✗ (A) Because products cannot be scored. — **Wrong.** Products can be scored; the problem is that the result does not stay meaningful.
- ✗ (B) Because vendors refuse to disclose product details. — **Wrong.** Disclosure is not the issue; even with full details a product list goes stale and hides overlap.
- ✓ (C) Because capabilities are durable across product changes and reveal overlap and absence that product lists hide. — **Correct.** Capabilities outlive products, and grouping by capability exposes both duplication and missing abilities.
- ✗ (D) Because capability maps are required by the EU AI Act. — **Wrong.** The Act requires various obligations by risk class, but does not prescribe a capability-based map.

K2 **Answer: B** Review: 2.2

A COO asks you to "count our AI use cases by Friday". What should you do first?

- ✗ (A) Send an all-staff survey immediately. — **Wrong.** Without agreed rules, survey answers cannot be counted consistently.
- ✓ (B) Agree the boundary, unit of analysis and counting rules. — **Correct.** Defining what counts before counting is the prerequisite for any meaningful number.
- ✗ (C) Purchase a SaaS discovery tool. — **Wrong.** A tool finds traffic; it does not decide what counts as an AI capability.
- ✗ (D) Ask IT for its list of approved tools. — **Wrong.** IT's list omits unsanctioned and embedded AI and is product-organized.

K3 **Answer: D** Review: 2.3

An invoice workflow routes approvals using fixed amount thresholds and a supplier whitelist. How should it be classified?

- ✗ (A) Machine learning — **Wrong.** Nothing is learned from data; the thresholds are hand-written.
- ✗ (B) Generative AI — **Wrong.** It generates no content.
- ✗ (C) Agentic AI, because it acts without a human — **Wrong.** Acting automatically is not enough; agentic AI involves a model planning and choosing actions with some autonomy.
- ✓ (D) Rules-based automation — outside the AI inventory boundary — **Correct.** Explicit, deterministic if-then rules are rules-based automation, not AI.

K4 **Answer: A** Review: 2.3

What most distinguishes agentic AI from generative AI?

- ✓ (A) It plans multi-step work and takes actions through tools with some autonomy. — **Correct.** The defining feature is acting — calling tools and changing things — not just answering.
- ✗ (B) It uses larger models. — **Wrong.** Model size is not the defining feature; small models can be used agentically.
- ✗ (C) It is always fully autonomous with no human involvement. — **Wrong.** Agentic systems may include human checkpoints; autonomy is a matter of degree.
- ✗ (D) It does not need any data. — **Wrong.** Agents need tool access, permissions and context; they are not data-free.

K5 **Answer: D** Review: 2.6

Two capabilities both have a blended score of 3.2. One has very low adoption; the other has very low governance coverage. What does this show?

✗ (A) They are equally mature and should be treated the same. — **Wrong.** Their identical blend conceals completely different problems.

✗ (B) The scoring scale is too narrow. — **Wrong.** A wider scale would not fix the problem caused by averaging.

✗ (C) Adoption should be weighted more heavily. — **Wrong.** Reweighting still produces a blend that hides the difference.

✓ (D) Blended scores hide which remedy is needed; independent axes should be reported. — **Correct.** Low adoption calls for process and people change; low governance calls for owners and rules. Only separate axes reveal this.

K6 **Answer: A** Review: 2.4, 2.7

Which instrument would reveal that 2,000 employees use an unapproved GenAI tool?

✓ (A) An AI capability inventory that includes unsanctioned use — **Correct.** A descriptive inventory with discovery beyond sanctioned tools is designed to find exactly this.

✗ (B) An AI maturity assessment — **Wrong.** Maturity assessments evaluate organizational ability by dimension; they do not catalogue specific tools in use.

✗ (C) An AI readiness survey of executives — **Wrong.** Executives are unlikely to know about, or report, unsanctioned tool use.

✗ (D) A vendor benchmark report — **Wrong.** Benchmarks evaluate products, not the organization's actual usage.

K7 **Answer: B** Review: 2.8

A forecasting capability cannot progress until shipment master data is harmonised. Which gap condition applies?

✗ (A) Absent — **Wrong.** Absent means the capability does not exist and needs to be created; here the blocker is a prerequisite.

✓ (B) Dependency-blocked — **Correct.** Progress depends on fixing a prerequisite (data), which also makes that fix a priority.

✗ (C) Underperforming — **Wrong.** Underperforming describes a working capability that falls short; here it cannot progress at all.

✗ (D) Unadopted — **Wrong.** Unadopted means it works but is not used; the issue here is upstream.

K8 **Answer: C** Review: 2.9

An executive asks, "Is the extraction tool accurate?" Which response best avoids overpromising?

✗ (A) "It is state of the art, so it will be fine." — **Wrong.** This is vague and makes an unsupported promise.

✗ (B) "AI is never accurate, so we must keep humans." — **Wrong.** This is an overgeneralisation that is unhelpful in the other direction.

✓ (C) "On 2,000 of our invoices it got key fields right about 9 in 10 times; we haven't tested handwritten ones yet." — **Correct.** This ties performance to evidence and conditions, and states what is still unknown.

✗ (D) "Yes — the vendor says 98%." — **Wrong.** This repeats an unverified claim with no conditions.

MODULE 3

The Advisory Toolkit

A tool-agnostic mental model of what an AI advisory engagement produces — from opportunity assessment to executive recommendation — and of what turns a document into an artifact that enables a decision.

YOU WILL BE ABLE TO

- Describe the arc of advisory deliverables and the transformation stage each supports
- State the purpose, decision and decision-holder for each deliverable
- Separate an advisory artifact from a report using the three-question test
- Distinguish delivery deliverables from the governance documentation an organization must maintain
- Explain how phase gates, scoping and stakeholder decision rights define completion
- Explain how the commercial form of a contract changes what “done” means

IN THIS MODULE

1. The arc of deliverables
2. The five deliverables in detail
3. Artifact vs report
4. Delivery deliverables vs governance documentation
5. Phase gates and what “complete” means
6. Commercial form and completion
7. Case study · templates · cheat sheet · knowledge check

Suggested study time · 3 hours

Advisory work is judged through what it leaves behind. This module gives you a mental model of the deliverables an AI advisory engagement produces — what each one is for, which decision it enables and who holds that decision — independent of any template, tool or consultancy method. It deliberately teaches the *arc and purpose* of deliverables, not how to produce them: production methods come later in the certificate.

KEY TAKEAWAYS

1. Deliverables follow an arc: **opportunity assessment → roadmap → business case → trial report → executive recommendation**, each tied to a transformation stage.
2. An **advisory artifact** names the decision it enables, who holds it and what happens if the answer is no. Otherwise it is a **report**.
3. **Delivery deliverables** serve the engagement; **governance documentation** (system inventory, ownership records, oversight specifications, audit logs) is maintained by the organization for the life of the system.
4. What "complete" means depends on **phase gates**, **engagement scope**, **stakeholder decision rights** and the **commercial form** of the contract.

3.1 The arc of advisory deliverables

AI transformation moves through recognisable stages — exploring possibilities, prioritising and sequencing, committing resources, proving value, and scaling or stopping. Each stage requires a different decision, and therefore a different deliverable. Thinking in an arc prevents two common mistakes: producing a business case before anyone has agreed which problem matters, and producing a final recommendation without evidence from a trial.

Figure 3.1 — The arc of AI advisory deliverables mapped to transformation stages

NOTE · Not every engagement covers the whole arc

A short engagement may deliver only an opportunity assessment; another may begin with an existing roadmap and deliver only a trial report and recommendation. Knowing where your deliverable sits on the arc tells you what it must assume from earlier stages and what it must hand to later ones.

3.2 The five deliverables in detail

EXPLORE

Opportunity assessment

Purpose	Decision it enables	Decision holder	If the answer is no	Typical failure
Identify and screen AI opportunities against business problems, value, feasibility and readiness.	Which opportunities deserve further investment of time and attention?	Executive sponsor / leadership team	Opportunities are dropped or parked; attention moves to non-AI levers or other problems.	A long list of AI ideas with no screening criteria; vendor-driven ideas presented as opportunities.

PRIORITIZE & SEQUENCE

Roadmap

Purpose	Decision it enables	Decision holder	If the answer is no	Typical failure
Sequence chosen initiatives and enablers (data, platform, governance, skills) over time, with dependencies and gates.	In what order do we pursue these initiatives, and which enablers must come first?	Sponsor with portfolio / transformation board	Initiatives are re-sequenced or deferred; enabler investment is reconsidered.	A Gantt chart of projects with no dependencies, enablers or decision gates.

COMMIT

Business case

Purpose	Decision it enables	Decision holder	If the answer is no	Typical failure
Justify investment in a specific initiative: value, costs (including run cost), risks, assumptions, alternatives.	Should we fund this initiative, at this level, on these assumptions?	Budget holder / investment committee	Not funded, funded at a smaller scope, or sent back for evidence.	Benefits based on vendor claims; run costs omitted; no “do nothing” or non-AI alternative.

PROVE

Trial (pilot) report

Purpose	Decision it enables	Decision holder	If the answer is no	Typical failure
Report what a time-boxed trial showed under stated conditions against pre-agreed criteria.	Did the trial meet the criteria to proceed, under realistic conditions?	Sponsor and production owner at the phase gate	Stop, extend with a changed design, or pivot to an alternative.	Moving goalposts; criteria defined after results; curated data presented as production evidence.

SCALE / STOP **Executive recommendation**

Purpose	Decision it enables	Decision holder	If the answer is no	Typical failure
A concise, evidence-graded recommendation with conditions, risks and the decision required.	Do we scale, pivot or stop — and under what conditions?	Executive sponsor or board	Initiative stopped or redirected; resources released.	A long report with the recommendation buried on page 40; no conditions or confidence level.

EXAM TRAP

A **trial report** answers "did it work under the agreed conditions?"; an **executive recommendation** answers "what should we do now?". The recommendation uses the trial's evidence but also weighs strategy, risk, cost and alternatives. Questions often test whether you would ask an executive to decide on the trial report alone.

3.3 Artifact vs report: the three-question test

A report informs. An advisory artifact enables a decision. The difference is not length, polish or format — it is whether the document is built around a decision. Apply the test to anything you produce:

1 · WHAT decision?
The specific choice this document must enable, phrased as a question

2 · WHO holds it?
The named role with the authority to say yes or no — and by when

3 · IF NO, then what?
What happens instead — stop, pivot, defer, fund the alternative

If any of the three cannot be answered, the document is a report — informative, but not an advisory artifact.

Figure 3.2 — The three-question test

EXAMPLE · Turning a report into an artifact

Report version	Artifact version
"Current state of AI in Customer Service" — 35 pages describing tools, vendors and market trends.	"Decision: should Customer Service fund an agent-assist pilot in Q2? Holder: VP Customer Service, by 15 March. If no: we invest the budget in self-service status notifications instead." — followed by 6 pages of evidence.
"Findings from the GenAI pilot."	"Decision: does the pilot meet the three agreed criteria to move to production? Holder: COO at Gate 2. If no: we stop and release the team, or extend 6 weeks to test multilingual cases."

KEY IDEA

Writing the "if no" answer is the hardest and most valuable part. It forces you to show the decision-holder that a genuine choice exists — and protects you from producing documents that can only be approved.

3.4 Delivery deliverables vs governance documentation

Advisory deliverables are produced *for the engagement*: they enable decisions and then age. AI **governance documentation** is different: it is maintained *by the organization*, for the life of each AI system, often because regulation, audit or internal policy requires it. Advisors frequently help design it, but must not confuse it with their own deliverables or leave it without an owner.

	Delivery deliverables	Governance documentation
Examples	Opportunity assessment, roadmap, business case, trial report, executive recommendation	AI system inventory, ownership records, human-oversight specifications, audit logs
Purpose	Enable a decision at a point in time	Demonstrate ongoing control and accountability
Owner	Engagement (advisor produces; sponsor accepts)	The organization — named internal owners
Lifespan	Engagement or decision cycle	Life of the AI system, plus retention periods
Audience	Decision-holders	Risk, audit, regulators, internal oversight bodies
Updated	Rarely after acceptance	Continuously, as systems and use change

Governance document	What it records	Why it matters
AI system inventory	Every AI system in use: purpose, owner, risk class, status, vendor	Basis for all oversight; regulators and auditors start here
Ownership records	Accountable business, technical and data owners per system; changes over time	Ensures someone answers for each system after the project team leaves
Oversight specifications	How humans supervise: who reviews which outputs, when they can intervene or stop the system, required competence	Turns “human in the loop” into an operable, testable design
Audit logs	Automatically recorded events: inputs, outputs, decisions, overrides, changes	Enables investigation, accountability and monitoring

NOTE · Regulatory orientation

The EU AI Act (Regulation (EU) 2024/1689) places documentation, logging and human-oversight obligations on providers and deployers of high-risk AI systems; for example, deployers must assign human oversight to people with the necessary competence and keep automatically generated logs under their control for an appropriate period. Application dates and details have been subject to change, so always confirm current requirements with legal specialists. ISO/IEC 42001 (AI management systems) also expects documented roles, records and controls.

PITFALL · The orphaned inventory

An advisor builds a beautiful AI system inventory as an engagement deliverable, the engagement ends, nobody owns it, and within six months it is wrong. Every governance document you help create needs a named internal owner and an update process *before* you hand it over.

3.5 Phase gates, scope and decision rights: what "complete" means

A **phase gate** is a pre-agreed decision point at which a named decision-holder chooses to proceed, pivot or stop based on criteria set in advance. Gates turn deliverables into decisions: a deliverable is complete when it allows the gate decision to be taken — not when it has been written.

Gate element	Description
Entry criteria	What must be true to hold the gate (e.g., trial run for 8 weeks on production data)
Decision criteria	Measurable thresholds agreed before the phase (e.g., ≥ 30% handling-time reduction, no critical errors)
Decision-holder	One named role; others consulted
Possible outcomes	Go · Go with conditions · Pivot · Extend (time-boxed) · Stop
Evidence deliverable	The artifact presented at the gate (business case, trial report, recommendation)

Three things that redefine "done"

- **Engagement scoping** — what was in scope determines which decisions your deliverables must enable (Module 4).
- **Stakeholder decision rights** — a deliverable accepted by the sponsor but not by the person with authority to decide is not complete in any useful sense.
- **Commercial form** — the contract defines what you are paid to deliver, which may be effort, a document, or an outcome (next section).

> **CONSULTANT'S TIP**
> Agree acceptance criteria for each deliverable at the start: "The trial report is complete when it reports results against criteria C1–C3 and is presented at Gate 2." This prevents endless revision cycles and makes clear that *acceptance of the document* is different from *a yes decision*.

3.6 Commercial form and the meaning of completion

Commercial form	You are paid for...	"Complete" means...	Watch-out in AI work
Time & materials	Effort (days/hours) at agreed rates	Effort delivered as directed; deliverables are best-effort	Scope can expand indefinitely; the client carries uncertainty
Capped T&M	Effort up to a ceiling	Cap reached or work finished	Need early warning before cap
Fixed fee / fixed scope	Specified deliverables	Deliverables accepted against criteria	AI uncertainty makes fixed outcomes risky; define deliverables, not results
Outcome- or value-based	Achieved business results	Measured outcome achieved	Outcome depends on factors outside the advisor's control (data, adoption); attribution disputes
Retainer / advisory subscription	Availability and ongoing advice	Period elapsed with service levels met	Value can become invisible; agree decision cadence

EXAM TRAP

Under a **fixed-fee** contract, "complete" is normally the accepted deliverable — *not* a positive decision and *not* a successful AI system. A trial report recommending "stop" can be a fully complete, high-value deliverable. Under an **outcome-based** contract, by contrast, completion depends on the measured result.

3.7 Case study: Halden Bank

CASE STUDY **The 60-page report that nobody could approve**

Halden Bank (fictional) · retail bank · 2.1m customers

Situation

Halden Bank engaged a consultancy on a fixed fee for an "AI opportunity assessment for operations". Eight weeks later, the consultancy delivered a 60-page report describing 34 possible AI use cases, market trends and vendor landscapes. The COO thanked the team, shared the report, and nothing happened. Six months later, a new advisor is brought in to "get AI moving".

Diagnosis

- The report named no decision. It listed opportunities but did not recommend which to pursue.
- No decision-holder was identified; the COO assumed the transformation board would decide, the board assumed the COO would.
- No "if no" — so there was no alternative to compare against, and no urgency.
- Three of the use cases were already running as unsanctioned pilots in branches; the report did not mention governance documentation for them.

The new advisor's approach

1. Reframes the deliverable: *"Decision: which three operational AI opportunities should receive €400k discovery funding in H1? Holder: COO, at the March transformation board. If no: budget is reallocated to the core-banking migration."*
2. Screens the 34 ideas on value, feasibility, readiness and risk; shortlists five, recommends three.
3. Flags the three unsanctioned pilots, and recommends they be entered in the AI system inventory with named owners and oversight specifications — as governance documentation owned by Operational Risk, not as part of the advisory deliverable.
4. Defines a roadmap with Gate 1 (business cases) in June.

Outcome

The board approved two of three opportunities at the March meeting and stopped one unsanctioned pilot that processed customer data without a legal basis.

Discussion questions

1. Was the original consultancy's deliverable "complete" under its fixed-fee contract? Was it valuable?
2. Apply the three-question test to the original report and to the reframed one.
3. Why should the unsanctioned pilots' documentation be owned by Operational Risk rather than the advisor?

Model answers (brief)

1. Probably complete contractually, since the report was accepted. But it enabled no decision, so its advisory value was low — completion and value are different things.
2. Original: no decision, no holder, no "if no" — a report. Reframed: all three answered — an artifact.

3. Governance documentation must persist for the life of the systems and be maintained by the organization; the advisor leaves when the engagement ends.

3.8 Consulting templates

T3.1 Deliverable Decision Card

Purpose: Attach to the front of every advisory deliverable; if it cannot be completed, the deliverable is a report. **Use when:** before drafting any deliverable

Deliverable: .....

Stage on the arc (Explore / Prioritize / Commit / Prove / Scale-Stop): .....

Decision it enables (as a question): .....

Decision-holder (named role): .....

Decision date / gate: .....

If the answer is NO, what happens instead: .....

Evidence threshold the holder needs: .....

Acceptance criteria for this document: .....

T3.2 Deliverable Arc Map

Purpose: Place the engagement's deliverables on the arc and show hand-offs.

Stage	Deliverable	Decision	Holder	Assumes from earlier stage	Hands to next stage	Gate date
Explore				—		
Prioritize						
Commit						
Prove						
Scale / stop					—	

T3.3 Phase Gate Checklist

Purpose: Prepare a gate so that a decision — not a discussion — results.

BEFORE THE PHASE

- [ ] Decision criteria written and agreed with decision-holder
- [ ] Single decision-holder named
- [ ] Possible outcomes listed (go / conditions / pivot / extend / stop)
- [ ] Evidence deliverable and format agreed

AT THE GATE

- [ ] Results reported against the original criteria (no moved goalposts)
- [ ] Conditions and limitations of evidence stated
- [ ] Run cost and ownership for next phase confirmed
- [ ] Decision recorded with date, conditions and owner

T3.4 Governance Documentation Handover Register

Purpose: Ensure every governance document has an internal owner before the engagement ends.

AI system	Inventory entry?	Ownership record	Oversight specification	Audit logging in place	Internal owner	Review cadence

3.9 Best practices and pitfalls

✓ DO THIS

- Start every deliverable by writing its decision card.
- Place each deliverable on the arc and state what it assumes.
- Agree decision criteria before trials begin.
- Separate document acceptance from the business decision.
- Hand governance documents to named internal owners.
- Match the contract form to the uncertainty of the work.

× AVOID THIS

- Delivering long-lists without a recommendation.
- Writing a business case before the problem is prioritized.
- Defining trial success after seeing results.
- Treating a governance inventory as a one-off deliverable.
- Promising AI outcomes under a fixed-fee contract.
- Burying the recommendation at the end.

3.10 Module 3 cheat sheet

MODULE 3 · ONE-PAGE SUMMARY

The Advisory Toolkit

THE ARC

Opportunity assessment (Explore: which problems?) → **Roadmap** (Prioritize: what order, which enablers?) → **Business case** (Commit: fund it?) → **Trial report** (Prove: did it meet criteria?) → **Executive recommendation** (Scale/stop: what now, under what conditions?)

ARTIFACT TEST

- What decision does it enable?
- Who holds that decision?
- What happens if the answer is no?

Can't answer all three → it's a report.

GOVERNANCE DOCUMENTATION

- AI system inventory
- Ownership records
- Oversight specifications
- Audit logs

Owned by the organization, for the system's life.

PHASE GATE

Pre-agreed criteria · one decision-holder · outcomes: go / go with conditions / pivot / extend / stop · evidence deliverable.

"COMPLETE" DEPENDS ON

- Phase gates
- Engagement scope
- Stakeholder decision rights
- Commercial form

COMMERCIAL FORMS

T&M: effort delivered · **Capped T&M**: effort to ceiling · **Fixed fee**: accepted deliverables (not a yes, not a working system) · **Outcome-based**: measured results (attribution risk) · **Retainer**: availability over a period.

REMEMBER

Document accepted ≠ decision taken ≠ value created. A "stop" recommendation can be a complete, high-value deliverable.

3.11 Knowledge check

Answer all questions before turning the page.

K1 **Which deliverable primarily enables the decision "in what order should we pursue these initiatives, and which enablers come first?"**

- (A) Business case
- (B) Trial report
- (C) Opportunity assessment
- (D) Roadmap

K2 **A 40-page document describes market trends and vendors but names no decision, holder or alternative. What is it?**

- (A) An executive recommendation
- (B) A report rather than an advisory artifact
- (C) Governance documentation
- (D) A business case

K3 **Which of these is governance documentation the organization must maintain, rather than a delivery deliverable?**

- (A) The roadmap
- (B) The trial report
- (C) The oversight specification for a deployed credit-scoring system
- (D) The opportunity assessment

K4 **Under a fixed-fee contract for a trial report, the trial misses its criteria and the report recommends stopping. Is the engagement complete?**

- (A) Yes, provided the report meets its agreed acceptance criteria.
- (B) No — the advisor must extend the trial at no cost until it succeeds.
- (C) Only if the sponsor agrees with the recommendation.
- (D) No — the contract is only complete if the AI goes to production.

K5 **What is the main risk of defining a trial's success criteria after results are available?**

- (A) It delays the trial.
- (B) It allows goalposts to move, so the gate decision is no longer based on a fair test.
- (C) It increases the vendor's costs.
- (D) It makes the report longer.

K6 **Which commercial form makes completion depend on factors such as client data quality and user adoption that the advisor does not control?**

- (A) Outcome- or value-based
- (B) Fixed fee
- (C) Retainer
- (D) Time and materials

K7 **An advisor builds an AI system inventory during an engagement. What must happen before handover?**

(A) It should be archived with the final report.
(B) It should be published externally.
(C) The advisor should agree to update it annually.
(D) A named internal owner and an update process must be in place.

3.12 Knowledge check: answers and explanations

K1 **Answer: D** Review: 3.1–3.2

Which deliverable primarily enables the decision "in what order should we pursue these initiatives, and which enablers come first?"

- ✗ A — Business case — **Wrong.** A business case justifies funding a specific initiative, not the order of a portfolio.
- ✗ B — Trial report — **Wrong.** A trial report reports evidence from a test against criteria.
- ✗ C — Opportunity assessment — **Wrong.** This screens which opportunities merit attention; it precedes sequencing.
- ✓ D — Roadmap — **Correct.** The roadmap sequences initiatives and enablers with dependencies and gates.

K2 **Answer: B** Review: 3.3

A 40-page document describes market trends and vendors but names no decision, holder or alternative. What is it?

- ✗ A — An executive recommendation — **Wrong.** A recommendation must state the decision required and conditions.
- ✓ B — A report rather than an advisory artifact — **Correct.** It fails the three-question test; it may inform, but it enables no decision.
- ✗ C — Governance documentation — **Wrong.** Governance documents record systems, owners, oversight and logs — not market trends.
- ✗ D — A business case — **Wrong.** A business case justifies a specific funding decision.

K3 **Answer: C** Review: 3.4

Which of these is governance documentation the organization must maintain, rather than a delivery deliverable?

- ✗ A — The roadmap — **Wrong.** A roadmap is a delivery deliverable that enables a sequencing decision.
- ✗ B — The trial report — **Wrong.** A trial report is a delivery deliverable for a gate decision.
- ✓ C — The oversight specification for a deployed credit-scoring system — **Correct.** It records how humans supervise a live system and must be maintained for the system's life.
- ✗ D — The opportunity assessment — **Wrong.** This is an early-stage delivery deliverable.

K4 **Answer: A** Review: 3.5–3.6

Under a fixed-fee contract for a trial report, the trial misses its criteria and the report recommends stopping. Is the engagement complete?

- ✓ A — Yes, provided the report meets its agreed acceptance criteria. — **Correct.** Completion is acceptance of the deliverable; a well-evidenced "stop" is a valid and valuable result.
- ✗ B — No — the advisor must extend the trial at no cost until it succeeds. — **Wrong.** Nothing in a fixed-fee form obliges this; it confuses deliverable with outcome.
- ✗ C — Only if the sponsor agrees with the recommendation. — **Wrong.** Acceptance of the document is separate from agreement with the decision it supports.
- ✗ D — No — the contract is only complete if the AI goes to production. — **Wrong.** Fixed-fee contracts normally pay for specified, accepted deliverables, not for AI success.

K5 **Answer: B** Review: 3.5

What is the main risk of defining a trial's success criteria after results are available?

- ✗ A — It delays the trial. — **Wrong.** Timing of definition does not itself delay the trial.
- ✓ B — It allows goalposts to move, so the gate decision is no longer based on a fair test. — **Correct.** Criteria set afterwards can be bent to fit results, which undermines the gate.
- ✗ C — It increases the vendor's costs. — **Wrong.** Costs are not the core problem.
- ✗ D — It makes the report longer. — **Wrong.** Length is irrelevant to the integrity of the decision.

K6 **Answer: A** Review: 3.6

Which commercial form makes completion depend on factors such as client data quality and user adoption that the advisor does not control?

✓ A Outcome- or value-based — **Correct.** Payment depends on measured results, which are affected by factors outside the advisor's control.

✗ B Fixed fee — **Wrong.** Fixed fee pays for accepted deliverables.

✗ C Retainer — **Wrong.** A retainer pays for availability over a period.

✗ D Time and materials — **Wrong.** T&M pays for effort; completion does not depend on business results.

K7 **Answer: D** Review: 3.4

An advisor builds an AI system inventory during an engagement. What must happen before handover?

✗ A It should be archived with the final report. — **Wrong.** Archiving makes it stale; it must be a living record.

✗ B It should be published externally. — **Wrong.** Publication is not the requirement; ownership and maintenance are.

✗ C The advisor should agree to update it annually. — **Wrong.** This keeps ownership outside the organization, the reverse of what is needed.

✓ D A named internal owner and an update process must be in place. — **Correct.** Governance documentation must be maintained by the organization for the life of the systems.

MODULE 4

Framing and Scoping the Engagement

Clarifying objectives, mapping stakeholders, gathering requirements and drawing a scope boundary that a later change request can be tested against — designed for the uncertain outcomes, data dependencies and moving technology of AI work.

YOU WILL BE ABLE TO

- Explain why AI engagements need different scoping from ordinary IT or consulting projects
- Turn a sponsor's aspiration into clear, measurable business objectives
- Map stakeholders, analyse their positions and build alignment
- Gather business, user, data, non-functional and governance requirements
- Write a scope boundary of inclusions, exclusions, assumptions and dependencies that can test a change request
- Use shared vocabulary, change control and escalation paths to keep scope creep out
- Know which questions an advisor answers directly and which belong to a specialist

IN THIS MODULE

1. Why AI scoping is different
2. Clarifying business objectives
3. Stakeholder mapping and alignment
4. Requirements gathering
5. The scope boundary
6. Vocabulary work and scope creep
7. Expectation management
8. Change control and escalation
9. The consultant's scope of authority
10. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4 hours

AI engagement scoping decides whether an advisory project delivers value or drifts. A well-framed engagement has clear business objectives, the right stakeholders around the table, requirements that include data and governance, and a scope boundary sharp enough that any later request can be tested against it. This module teaches that framing — and the discipline to protect it while outcomes stay uncertain and the technology keeps moving.

KEY TAKEAWAYS

1. AI scoping must accommodate **uncertain outcomes**, **data readiness dependencies** and **shifting technical capability**.
2. Objectives are business outcomes with baselines — not technology outputs.
3. Map stakeholders by power and interest; **gatekeepers** (legal, security, employee representatives) need early engagement even if their interest is low.
4. Requirements include **data** and **governance** requirements, not just functional ones.
5. A scope boundary lists **inclusions, exclusions, assumptions and dependencies** precisely enough to **test a change request** against.
6. Shared **vocabulary**, **change control** and **escalation paths** keep scope creep out.
7. Know your **scope of authority**: advise on what is yours; refer law, security testing, deep technical design and employment matters to specialists.

4.1 Why AI engagement scoping is different

AI characteristic	Scoping consequence	Scoping response
Uncertain outcomes	You cannot promise performance before testing on real data	Scope learning milestones and decisions, not results; time-box experiments
Data readiness dependencies	Work stalls if data is late, unusable or not permitted	Make data access an explicit dependency with an owner and date; add a data gate
Shifting technical capability	New models or features appear mid-engagement and invite re-scoping	Scope by capability and decision; agree how new options will be assessed (change control)
Socio-technical impact	Affected people and gatekeepers can block late	Include them in stakeholder mapping from day one
Ambiguous vocabulary	“Accuracy”, “pilot”, “agent”, “automate” mean different things to different people	Agree an engagement glossary in the charter

4.2 Clarifying business objectives

A business objective states what will be different for the business, measurably, if the engagement succeeds. It is not a description of a technology. Use an **objective tree**: start from the strategic goal, break it into business outcomes, then into measurable indicators with baselines.

Weak objective (output-focused)	Strong objective (outcome-focused)
"Implement a GenAI assistant for claims handlers."	"Reduce average claims handling time for standard motor claims from 42 to 30 minutes by Q4, without increasing leakage."
"Use AI in recruitment."	"Cut time-to-offer for engineering roles from 90 to 60 days while keeping candidate-quality ratings stable."
"Develop an AI strategy."	"Agree by June a funded portfolio of up to three AI initiatives that address the top-five cost drivers in operations."

CONSULTANT'S TIP · Objective checklist

- Is it a business outcome, not a technology?
- Is there a baseline and a target?
- Is there a guardrail metric (what must *not* get worse)?
- Is there a date?
- Is there a named owner of the outcome?

NOTE · Advisory objectives vs business objectives

The engagement often has its own objective — usually *a decision* ("the COO decides at Gate 2 whether to scale agent-assist") — which serves the business objective. Write both. Advisors control the first far more than the second.

4.3 Stakeholder mapping and alignment

Stakeholder analysis identifies everyone who can affect or is affected by the engagement, assesses their power and interest, and records their current and needed position. In AI work, pay special attention to **gatekeepers** — functions with low day-to-day interest but the power to stop go-live — and to **affected people** who may not be in the room.

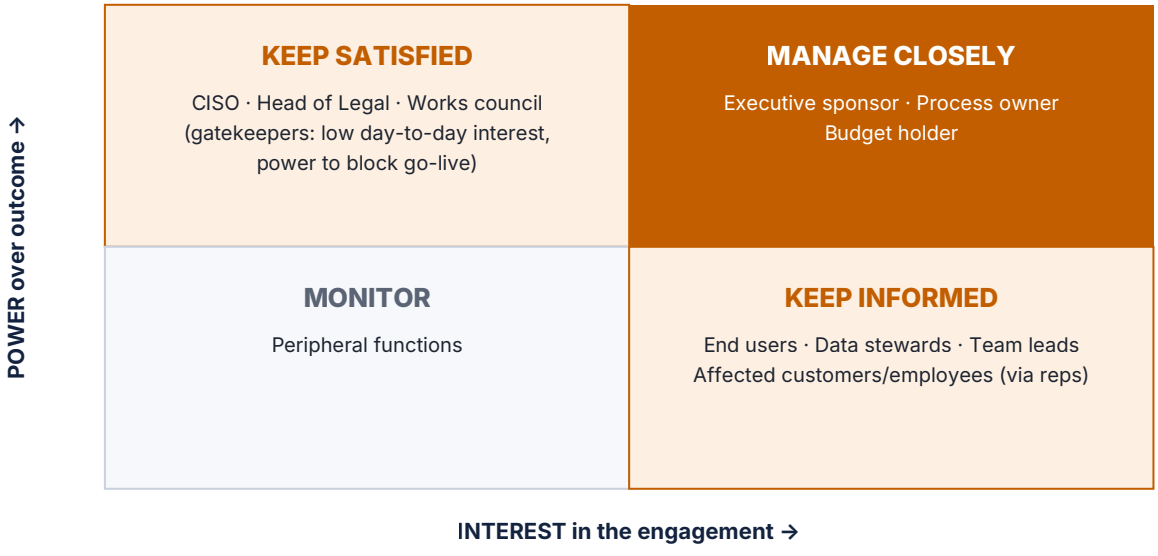


Figure 4.1 — Power/interest grid with typical AI engagement stakeholders

Stakeholder	Typical concerns in AI engagements	Alignment move
Executive sponsor	Visible results, speed, board narrative	Agree decision-focused objectives and a gate calendar
Process owner	Disruption, accountability for errors, headcount implications	Co-design the to-be process; make them the production owner
End users	Usability, trust, job impact, workload	Involve in requirements and trial; collect feedback visibly
Legal / compliance	Liability, regulatory classification, contracts	Early briefing; clear questions with deadlines
DPO	Personal data, purpose limitation, DPIA	Data-flow description early; DPIA trigger check
CISO	Data leakage, vendor access, new attack surface	Security requirements in scope; vendor review slot
Works council / employee reps	Monitoring, performance evaluation, job changes	Early information; transparency on logs and use
Finance	Credible costs and benefits, run cost	Joint assumptions in the business case

BEST PRACTICE · Stakeholder alignment techniques

- **Pre-wiring**: brief key stakeholders individually before group decisions.
- **Explicit positions**: record current vs needed stance (resist / neutral / support / champion) and plan to close the gap.
- **Shared facts**: agree baselines and definitions before debating options.
- **Decision records**: write down what was decided, by whom, and why.

4.4 Requirements gathering

Requirement type	Example questions	Example requirement
Business	What outcome? What is good enough? What must not get worse?	Handling time −25% with leakage unchanged
User	Who uses it, in what workflow, with what skills?	Handlers review suggested text before sending; one-click accept/edit
Data	What data, from where, with what legal basis, quality, latency?	Claims history 2021–2025, DE/AT, pseudonymised, nightly refresh
Non-functional	Latency, availability, volume, cost per transaction, languages?	Response < 5 s; German and English; cost < €0.10 per claim
Governance & risk	Risk class, oversight, logging, explainability, approvals?	All outputs reviewed by a human; logs retained per policy; Legal approves rules of use
Integration & security	Systems of record, identity, data residency, vendor access?	EU data residency; SSO; no training on client data by vendor

PITFALL · Requirements as wish-lists

"It should be accurate, fast and explainable" is not a requirement. Ask for the measure, the threshold and the test: "≥ 85% of suggested answers accepted without substantive edit, measured on a 500-case sample from production." Where a threshold cannot yet be set, write the decision about how it will be set — that is still a requirement.

4.5 The scope boundary

The scope boundary is the single most important artifact for preventing drift. It must be precise enough that, when someone asks "can you also…?", anyone can test the request against it and reach the same conclusion.

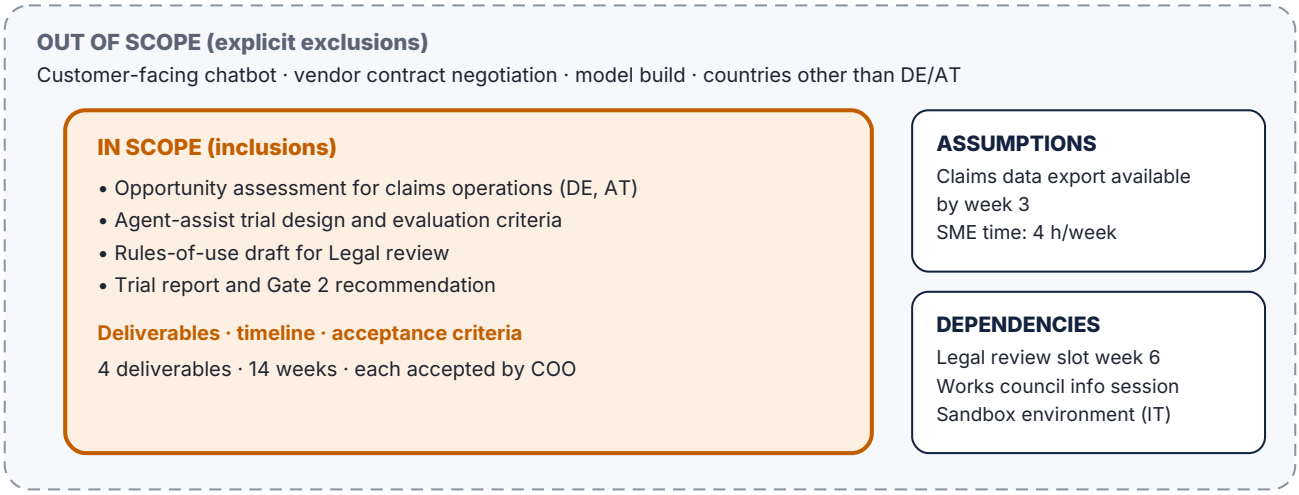


Figure 4.2 — A scope boundary: inclusions, explicit exclusions, assumptions and dependencies

Element	Purpose	Writing rule
Inclusions	What will be done and delivered	Name deliverables, populations, geographies, processes and time
Exclusions	What will explicitly not be done	List the most likely "can you also" requests — exclusions are the test's teeth
Assumptions	Conditions taken as true for the plan	Make each one checkable; give it an owner
Dependencies	Things others must deliver	Name who, what and by when; state the consequence of delay
Constraints	Fixed limits (budget, dates, policies)	Quantify
Acceptance criteria	How deliverables are accepted, by whom	Tie to the decision each deliverable enables (Module 3)

KEY IDEA · The change-request test

For any new request ask: **(1)** Is it within the stated inclusions? **(2)** Does it touch a stated exclusion? **(3)** Does it invalidate an assumption or dependency? **(4)** Does it change a deliverable, acceptance criterion or the decision being enabled? If the answer to (1) is no, or to any of (2)–(4) is yes, it is a change and goes to change control.

4.6 Vocabulary work and keeping scope creep out

Many scope disputes in AI engagements are really vocabulary disputes. The sponsor heard "pilot" and imagined a live service for 500 users; the team meant a four-week test with 20 volunteers. Agree an **engagement glossary** in the charter.

Term	What people may mean	Agree instead
Pilot	Demo · lab test · limited live service	"A time-boxed trial with N real users on production data, with criteria C1–C3"
Accuracy	Overall % correct · precision · recall · user acceptance	The specific metric, on which dataset, against which baseline
Automate	Fully replace a human · assist a human	Level of automation: suggest · draft for review · act with approval · act autonomously
Agent	Chatbot · workflow · autonomous tool-using system	Whether the system takes actions, which ones, with what permissions
Done / live	Built · tested · in production · adopted	The exact state and acceptance criteria

Common source of scope creep in AI work	Defence
"While you're at it, could it also handle…" (adjacent use case)	Test against exclusions; log as candidate for next phase
A new model or vendor release mid-engagement	Agreed rule: assess via change control; no silent switch
A new data source appears "helpful"	Treat as a change — new legal, quality and integration questions
Sponsor wants a production system from a trial	Point to the trial's definition and gate; production is a separate decision
Stakeholders add success metrics late	Criteria fixed at gate entry; new metrics go to the next phase

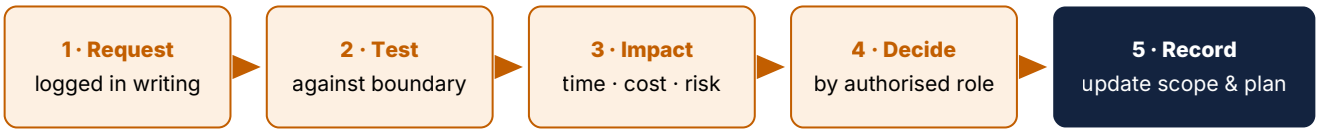
4.7 Expectation management with executive sponsors

- **Promise decisions, not performance.** "By week 10 you will be able to decide whether to scale, on evidence" rather than "by week 10 it will be 90% accurate".
- **Use ranges and conditions.** "Our best estimate is 20–35% time saving, if data access is granted by week 3."
- **Name the learning milestones.** Tell the sponsor when uncertainty will reduce and what could change the plan.
- **No surprises.** Short, regular updates with traffic-light status on objectives, risks and dependencies; bad news early.
- **Make "stop" a legitimate success.** Agree up front that a well-evidenced stop saves money and is a good outcome.

CONSULTANT'S TIP · Script for the over-eager sponsor

> *"I share your ambition. What I can commit to is that by the end of March you'll have the evidence to decide whether this deserves a full rollout. What I can't honestly promise yet is the performance figure — that depends on how it behaves on your data, which is exactly what the trial tells us."*

4.8 Change control and escalation paths



Outcomes at step 4: approve (with re-plan) · trade off (swap for something in scope) · defer to next phase · reject

Figure 4.3 — A lightweight change-control process

Escalation paths define who resolves what, at which level, within what time — so that issues are neither escalated too early (undermining the team) nor too late (missing deadlines).

Level	Who	Typical issues	Target resolution
1	Engagement lead + client project lead	Day-to-day scheduling, minor clarifications, SME availability	2 working days
2	Engagement partner + process owner	Changes within budget tolerance, dependency slips, conflicting requirements	1 week
3	Executive sponsor (steering)	Changes beyond tolerance, scope or objective changes, unresolved gatekeeper blocks	Next steering meeting or ad hoc within 2 weeks

4.9 The consultant's scope of authority

Credibility depends as much on what you decline to answer as on what you answer. An advisor answers questions within their competence and engagement mandate; questions that require licensed, specialist or deep technical judgment are framed by the advisor but answered by the right specialist.

Advisor answers directly	Advisor frames, specialist answers
Which business problem to prioritise and why	Whether a use case is "high-risk" under the EU AI Act → legal counsel
How to structure the decision and the gate criteria	Whether a DPIA is legally required and its outcome → DPO / privacy counsel
Who should hold which decision	Whether the system is secure against prompt injection → security testing specialists
What evidence is needed to decide	Which model architecture will meet latency targets → ML / solution architects
Whether vendor evidence supports the claim made (Module 5)	Co-determination rights and agreements → employment law / HR
How to sequence initiatives and enablers	Tax or accounting treatment of AI investments → finance / tax experts

EXAM TRAP

When a question asks how an advisor should respond to "Is this legal under the AI Act?", the correct pattern is: **frame the question precisely, identify the facts the specialist will need, and refer it to legal counsel** — not give a legal opinion, and not ignore the question.

4.10 Case study: Vantis Manufacturing

CASE STUDY **From predictive maintenance to "everything AI"**

Vantis Manufacturing (fictional) · industrial pumps · 6 plants in DE, PL, CZ

Situation

Vantis engaged an advisor for a 12-week, fixed-fee engagement to "assess and trial predictive maintenance on Line 3 in the Dortmund plant". By week 5 the scope was drifting: the plant director asked whether the same sensors could detect quality defects; the CIO wanted the trial switched to a newly released vendor platform; the sponsor mentioned to the board that "AI maintenance will be live across all plants by year-end".

What had gone wrong in framing

- The charter listed inclusions but no exclusions.
- "Trial" was never defined; the sponsor understood it as the start of rollout.
- Sensor data access depended on the OT security team, who were not on the stakeholder map; data arrived in week 4 instead of week 2.
- No change-control process existed; requests were discussed informally in corridors.

The reset

1. Re-issue the scope statement with explicit exclusions: quality-defect detection, other lines and plants, platform migration, production rollout.
2. Run the four-part change test on each request: defect detection (exclusion → logged for phase 2), platform switch (changes assumptions and dependencies → impact assessed: +5 weeks, new security review → sponsor declined for this phase), board statement (expectation issue → escalated to Level 3).
3. Add OT Security as a dependency owner; re-baseline the timeline by two weeks via change control.
4. Agree a glossary: "trial = 6 weeks on Line 3, criteria: ≥ 60% of failures flagged ≥ 48 h ahead, false alarms ≤ 2 per week".
5. Sponsor corrects the board message: "decision on scaling at Gate 2 in Q1".

Discussion questions

1. Which AI-specific characteristics (uncertain outcomes, data dependencies, shifting capability) showed up in this case?
2. Apply the change-request test to the CIO's platform request.
3. Which stakeholder was missing from the map, and in which quadrant of the power/interest grid should they have been?
4. Draft one sentence the advisor could use with the sponsor about the board statement.

Model answers (brief)

1. All three: uncertain outcome (board promise), data dependency (late sensor data), shifting capability (new vendor platform).
2. Not an inclusion; it invalidates assumptions (existing platform) and dependencies (security review), and changes the timeline → a change; impact assessed and decided by the sponsor.
3. OT Security — high power (can block data access), low interest → "keep satisfied".

4. “To protect your credibility with the board, let’s describe Q1 as the point where we’ll decide on scaling, based on evidence from Line 3, rather than a go-live date.”

4.11 Consulting templates

T4.1 Engagement Charter & Scope Statement

Purpose: Single source of truth for objectives, boundary and authority. **Use when:** signed off before work starts

OBJECTIVES

Business objective (outcome, baseline, target, guardrail, date):

Engagement objective (decision to be enabled, holder, gate):

BOUNDARY

Inclusions (deliverables, processes, populations, geographies, time):

Explicit exclusions:

Assumptions (each with owner):

Dependencies (who · what · by when · impact of delay):

Constraints:

GOVERNANCE

Acceptance criteria per deliverable:

Change-control owner and tolerance:

Escalation levels:

Engagement glossary attached? ☐ yes

T4.2 Stakeholder Map & Alignment Tracker

Purpose: Track power, interest and positions; plan alignment moves.

Stakeholder	Role / decision right	Power H/L	Interest H/L	Current position	Needed position	Key concern	Alignment action · owner · date
Head of OT Security	*Approves sensor data access*	*H*	*L*	*Neutral*	*Support*	*Network exposure*	*Briefing wk 1 · Lead · 12/01*

T4.3 Requirements Capture Sheet

Purpose: Capture requirements of every type with a testable measure.

ID	Type (Bus./User/Data/NFR/Gov./Sec.)	Requirement	Measure & threshold	Test method	Source	Priority (M/S/C)

T4.4 Change Request Form & Impact Test

Purpose: Test every new request against the scope boundary.

REQUEST

Requested by / date: .....

Description: .....

FOUR-PART TEST

- ☐ (1) Within stated inclusions? ☐ yes ☐ no
- ☐ (2) Touches an exclusion? ☐ yes ☐ no
- ☐ (3) Invalidates an assumption or dependency? ☐ yes ☐ no
- ☐ (4) Changes a deliverable, acceptance criterion or the decision enabled? ☐ yes ☐ no

IMPACT & DECISION

Impact on time / cost / risk / quality: .....

Options (approve · trade off · defer · reject): .....

Decision · by whom · date: .....

T4.5 Authority Boundary Card

Purpose: Decide quickly whether to answer, frame or refer a question.

Question received	Within my competence & mandate?	Answer / frame / refer	Specialist	Facts specialist needs	By when

T4.6 Engagement Glossary

Purpose: Fix the meaning of contested terms for this engagement.

Term	Agreed meaning in this engagement	Explicitly does NOT mean	Agreed by / date
Pilot / trial			
Accuracy			
Automate			
Agent			
Live / done			

4.12 Best practices and pitfalls

✓ DO THIS

- Write outcome objectives with baseline, target, guardrail and date.
- Put gatekeepers on the stakeholder map from day one.
- List exclusions as carefully as inclusions.
- Make data access a named dependency with a date.
- Agree an engagement glossary.
- Promise decision points, not performance.
- Route every change through the four-part test.

× AVOID THIS

- Technology-shaped objectives (“implement X”).
- Discovering legal, security or works-council concerns at go-live.
- Corridor agreements on scope changes.
- Silently switching models or vendors mid-trial.
- Letting “pilot” mean different things to different people.
- Giving legal or security opinions outside your competence.
- Escalating everything — or nothing.

4.13 Module 4 cheat sheet

MODULE 4 · ONE-PAGE SUMMARY

Framing and Scoping the Engagement

WHY AI SCOPING DIFFERS

- Uncertain outcomes → scope decisions & learning milestones
- Data dependencies → named, dated, gated
- Shifting capability → change control

GOOD OBJECTIVE

Business outcome · baseline · target · guardrail · date · owner. Plus an **engagement objective**: the decision enabled.

STAKEHOLDERS

Power/interest grid. **Gatekeepers** (legal, DPO, CISO, works council) = high power, low interest → keep satisfied, engage early. Pre-wire; record positions.

REQUIREMENT TYPES

Business · user · data · non-functional · governance & risk · integration & security. Each with measure + threshold + test.

SCOPE BOUNDARY

Inclusions · **exclusions** · assumptions · dependencies · constraints · acceptance criteria.

CHANGE-REQUEST TEST

1. Within inclusions?
2. Touches exclusion?
3. Breaks assumption/dependency?
4. Changes deliverable/criterion/decision?

CHANGE CONTROL

Request → test → impact → decide (approve · trade off · defer · reject) → record.

VOCABULARY

Define pilot, accuracy, automate, agent, done/live in an engagement glossary.

EXPECTATIONS

Promise decisions not performance · ranges + conditions · learning milestones · no surprises · “stop” is a success.

SCOPE OF AUTHORITY

Answer: problem, decision structure, decision rights, evidence, sequencing. **Frame & refer**: law, DPIA, security testing, architecture, employment law, tax.

4.14 Knowledge check

Answer all questions before turning the page.

K1 **Which is the best-formed business objective for an AI engagement?**

- (A) "Become an AI-first insurer."
- (B) "Deploy a large language model for the claims team."
- (C) "Reduce standard-claim handling time from 42 to 30 minutes by Q4 without increasing leakage."
- (D) "Achieve 95% model accuracy."

K2 **The Head of Legal shows little interest in a GenAI engagement but must approve the rules of use before go-live. Where are they on the power/interest grid, and what should you do?**

- (A) Low power, low interest — monitor.
- (B) High power, high interest — manage closely in weekly meetings.
- (C) Low power, high interest — keep informed.
- (D) High power, low interest — keep satisfied, and engage early with clear questions and deadlines.

K3 **What makes a scope boundary usable for testing change requests?**

- (A) Precise inclusions plus explicit exclusions, assumptions and dependencies.
- (B) A long list of inclusions in general terms.
- (C) A detailed project plan with all tasks.
- (D) Sponsor approval of the budget.

K4 **Mid-trial, a new model version is released and the CIO asks to switch to it. What is the right response?**

- (A) Refuse, since the scope is fixed.
- (B) Treat it as a change request: test against the boundary, assess impact, and let the authorised role decide.
- (C) Switch immediately — newer models are better.
- (D) Run both models without telling the sponsor.

K5 **The sponsor says "pilot", meaning a live service for 500 users. The team means a four-week test with 20 volunteers. What does this show is missing?**

- (A) More data scientists
- (B) A vendor contract
- (C) An agreed engagement glossary
- (D) A larger budget

K6 **A sponsor asks, "Is this use case high-risk under the EU AI Act?" What should the advisor do?**

- (A) Say it is not your area and move on.
- (B) Ask the vendor to confirm the classification.
- (C) Give a definitive yes or no to show expertise.
- (D) Frame the question, assemble the facts the specialist needs, and refer it to legal counsel with a deadline.

K7 **Which statement best manages a sponsor's expectations at the start of an AI trial?**

- (A) "By week 10 you'll be able to decide whether to scale, based on evidence from your own data."
- (B) "It will be 90% accurate by week 10."
- (C) "The vendor guarantees the result."
- (D) "AI results are unpredictable, so we cannot commit to anything."

K8 **Which requirement is written most testably?**

(A) “The assistant should use the latest model.”

(B) “≥ 85% of suggested replies accepted without substantive edit on a 500-case production sample; response < 5 s.”

(C) “Users must like the assistant.”

(D) “The assistant must be accurate and fast.”

4.15 Knowledge check: answers and explanations

K1 **Answer: C** Review: 4.2

Which is the best-formed business objective for an AI engagement?

- ✗ (A) "Become an AI-first insurer." — **Wrong.** This is aspirational and unmeasurable, with no baseline, target or date.
- ✗ (B) "Deploy a large language model for the claims team." — **Wrong.** This describes a technology output, not a business outcome.
- ✓ (C) "Reduce standard-claim handling time from 42 to 30 minutes by Q4 without increasing leakage." — **Correct.** It is an outcome with a baseline, target, guardrail and date.
- ✗ (D) "Achieve 95% model accuracy." — **Wrong.** This is a technical metric with no business outcome, dataset or baseline.

K2 **Answer: D** Review: 4.3

The Head of Legal shows little interest in a GenAI engagement but must approve the rules of use before go-live. Where are they on the power/interest grid, and what should you do?

- ✗ (A) Low power, low interest — monitor. — **Wrong.** Approval authority means high power; monitoring alone risks a late block.
- ✗ (B) High power, high interest — manage closely in weekly meetings. — **Wrong.** Their interest is low; weekly meetings would waste their time and goodwill.
- ✗ (C) Low power, high interest — keep informed. — **Wrong.** This understates their power to block.
- ✓ (D) High power, low interest — keep satisfied, and engage early with clear questions and deadlines. — **Correct.** Gatekeepers have power to stop go-live and need early, targeted engagement.

K3 **Answer: A** Review: 4.5

What makes a scope boundary usable for testing change requests?

- ✓ (A) Precise inclusions plus explicit exclusions, assumptions and dependencies. — **Correct.** Exclusions, assumptions and dependencies give the test something to check against.
- ✗ (B) A long list of inclusions in general terms. — **Wrong.** General inclusions without exclusions cannot settle borderline requests.
- ✗ (C) A detailed project plan with all tasks. — **Wrong.** A task plan is not a boundary; it does not say what is out of scope.
- ✗ (D) Sponsor approval of the budget. — **Wrong.** Budget approval does not define scope.

K4 **Answer: B** Review: 4.6, 4.8

Mid-trial, a new model version is released and the CIO asks to switch to it. What is the right response?

- ✗ (A) Refuse, since the scope is fixed. — **Wrong.** Refusing without assessment ignores potentially valuable options; change control exists to assess them.
- ✓ (B) Treat it as a change request: test against the boundary, assess impact, and let the authorised role decide. — **Correct.** This is the disciplined response to shifting capability.
- ✗ (C) Switch immediately — newer models are better. — **Wrong.** This changes assumptions and invalidates the trial's comparability without assessment.
- ✗ (D) Run both models without telling the sponsor. — **Wrong.** This is an unauthorised scope change and erodes trust.

K5 **Answer: C** Review: 4.6

The sponsor says "pilot", meaning a live service for 500 users. The team means a four-week test with 20 volunteers. What does this show is missing?

- ✗ (A) More data scientists — **Wrong.** Staffing is irrelevant to the misunderstanding.
- ✗ (B) A vendor contract — **Wrong.** The vendor contract does not define the client's internal terms.
- ✓ (C) An agreed engagement glossary — **Correct.** Defining terms like "pilot" prevents disputes that are really about vocabulary.
- ✗ (D) A larger budget — **Wrong.** Money does not resolve a definitional mismatch.

K6 **Answer: D** Review: 4.9

A sponsor asks, "Is this use case high-risk under the EU AI Act?" What should the advisor do?

✗ (A) Say it is not your area and move on. — **Wrong.** Ignoring it leaves a critical dependency unmanaged.

✗ (B) Ask the vendor to confirm the classification. — **Wrong.** The vendor has a conflict of interest and is not the client's legal adviser.

✗ (C) Give a definitive yes or no to show expertise. — **Wrong.** Legal classification is a legal judgment outside the advisor's authority.

✓ (D) Frame the question, assemble the facts the specialist needs, and refer it to legal counsel with a deadline. — **Correct.** The advisor frames and refers, keeping the issue on the critical path without overstepping.

K7 **Answer: A** Review: 4.7

Which statement best manages a sponsor's expectations at the start of an AI trial?

✓ (A) "By week 10 you'll be able to decide whether to scale, based on evidence from your own data." — **Correct.** This promises a decision point rather than a performance figure.

✗ (B) "It will be 90% accurate by week 10." — **Wrong.** This promises performance before evidence exists.

✗ (C) "The vendor guarantees the result." — **Wrong.** This outsources credibility to a party with an interest in the sale.

✗ (D) "AI results are unpredictable, so we cannot commit to anything." — **Wrong.** This is unhelpfully vague; advisors can commit to decisions and milestones.

K8 **Answer: B** Review: 4.4

Which requirement is written most testably?

✗ (A) "The assistant should use the latest model." — **Wrong.** This is a design preference, not a testable requirement.

✓ (B) "≥ 85% of suggested replies accepted without substantive edit on a 500-case production sample; response < 5 s." — **Correct.** It has measures, thresholds and a test basis.

✗ (C) "Users must like the assistant." — **Wrong.** This is not measurable as written.

✗ (D) "The assistant must be accurate and fast." — **Wrong.** There is no measure, threshold or test.

MODULE 5

Due Diligence on Claims, Vendors, and Signals

Protecting a client's plan from the market's noise: reading every capability claim back to the conditions that produced it, evaluating vendors and their staying power, and turning an honest reading of the evidence into a short note an executive can act on.

YOU WILL BE ABLE TO

- Read a capability claim back to the measurement conditions that produced it
- Tell a demonstration or proof of concept apart from a production deployment
- Question benchmark methodology and headline statistics before a plan is built on them
- Run a structured AI vendor evaluation
- Assess third-party and counterparty risk on multi-year platform commitments
- Assess supplier viability and financial durability
- Write an evidence-based advisory note an executive can act on

IN THIS MODULE

1. Why due diligence is an advisory duty
2. Reading a claim back to its conditions
3. The evidence ladder
4. Benchmarks and headline statistics
5. Vendor evaluation
6. Counterparty risk on multi-year commitments
7. Supplier viability and financial durability
8. Separating signal from noise
9. Writing the advisory note
10. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4–5 hours

AI vendor due diligence is what separates an advisor who protects a client's plan from one who repeats what the market is saying. Every AI market is full of confident numbers, impressive demos and urgent signals. This module teaches you to trace each claim to the conditions that produced it, to judge evidence strength, to evaluate vendors and their durability over a multi-year commitment, and to write the short, honest note an executive needs.

KEY TAKEAWAYS

1. A capability claim is only as good as its **measurement conditions**: task, data, metric, baseline, who measured, when, with what help, what was excluded.
2. Place every piece of evidence on the **evidence ladder**; a demo or PoC is not a production deployment.
3. Question **benchmark methodology** and **headline statistics** — fit to task, contamination, metric choice, averages, "up to", survey sampling.
4. Evaluate vendors on capability fit, evidence, security and privacy, integration, total cost, contract terms, and exit.
5. Multi-year AI platform commitments carry **counterparty risk**: lock-in, upstream model dependency, pricing and deprecation changes.
6. Assess **supplier viability**: funding runway, revenue quality, concentration, ownership — and contract for failure.
7. Turn evidence into a **short advisory note**: recommendation first, decision required, graded evidence, confidence, conditions, and what would change your view.

5.1 Why due diligence is an advisory duty

When an executive bases a plan on a number, they are trusting that someone checked it. If the advisor merely passes on vendor claims, analyst forecasts or a competitor's press release, the advisor adds nothing — and may add harm by lending them credibility. Due diligence is the advisor's protective function: it is how the client avoids building a multi-year plan on evidence that would not survive five minutes of questioning.

KEY IDEA

The advisor's question is never "is this claim true?" in the abstract. It is: **"Under what conditions was this measured, and do those conditions match ours?"**

5.2 Reading a capability claim back to its measurement conditions

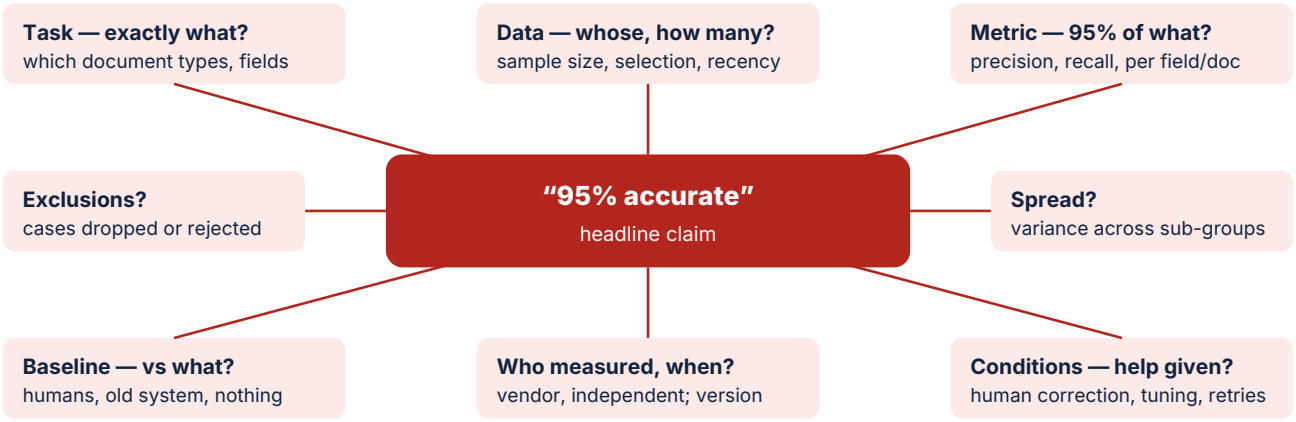


Figure 5.1 — Every headline number is the tip of a set of measurement conditions

Condition	Questions to ask	Why it matters
Task definition	Exactly what task? Which inputs, outputs, languages, document types?	A claim on a simpler or different task does not transfer
Data	Whose data? How many cases? How were they selected? How recent?	Hand-picked or small samples flatter results
Metric	Which metric? Per item or per field? Precision, recall, F1, acceptance rate?	"95%" can mean very different things
Baseline	Compared with what — humans, previous system, random, nothing?	Improvement claims are meaningless without a baseline
Who measured, when, which version	Vendor or independent? Date? Product/model version?	Self-measurement and old versions weaken evidence
Conditions and assistance	Was there tuning, human correction, retries, prompt engineering by experts?	Results may depend on effort the client will not replicate
Exclusions	Were some cases dropped, rejected or routed to humans?	Accuracy "on the cases it handled" can hide low coverage
Spread	How does performance vary across sub-groups, languages, edge cases?	Averages hide failures where it matters most

EXAMPLE · Reading back a claim

Claim: "Our assistant resolves 70% of customer queries automatically."

Read-back finds: measured at one retailer over two weeks; "resolved" meant the customer did not contact again within 24 hours; queries the bot could not classify (30% of traffic) were routed to agents before measurement and excluded; version two releases ago.

Honest restatement: "At one retailer, of the ~70% of queries the bot accepted, about 70% did not lead to a repeat contact within 24 hours — roughly half of all traffic — on an older version."

5.3 The evidence ladder: demonstration, PoC, pilot, production

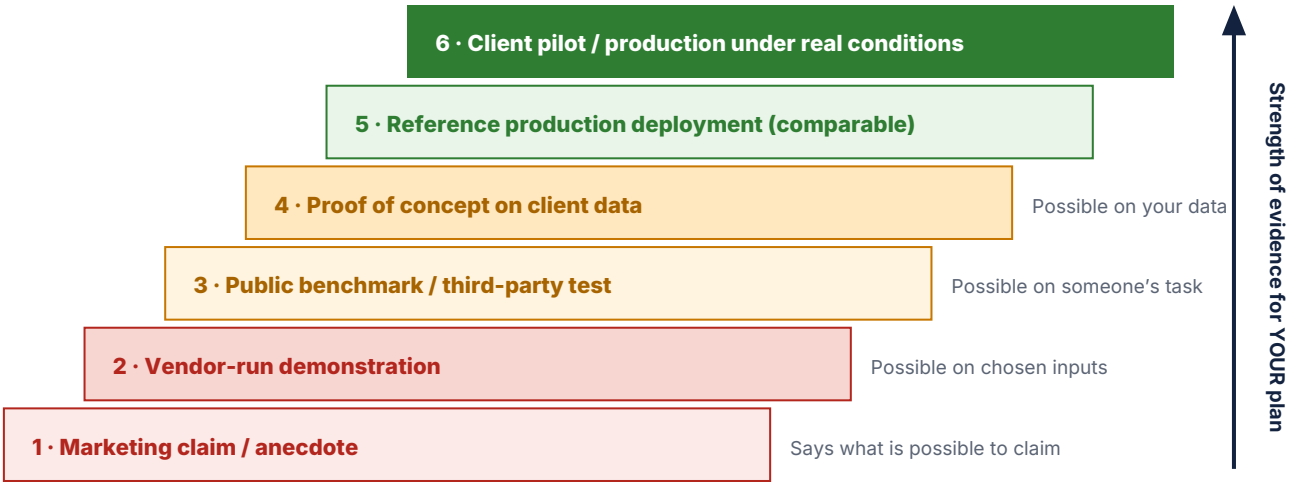


Figure 5.2 — The evidence ladder

Evidence type	What it proves	What it does not prove
Demonstration	The capability can be shown on chosen inputs	Performance on your data, at your volume, with your users
Proof of concept (PoC)	Technical feasibility on a sample, often of client data	Operational value, integration, adoption, run cost
Pilot / trial	Performance and usability under defined real conditions for a limited scope	Behaviour at full scale and over time
Production deployment	Sustained value with real users, integration, governance and monitoring	Transfer to a different context without checking similarity

> **EXAM TRAP · Reference customers**
> "Used by 300 enterprises" is an **adoption** claim, not a **performance** claim. Ask for references that are comparable in task, data and scale, that are **in production** (not pilots), and speak to them without the vendor present.

5.4 Questioning benchmarks and headline statistics

Issue	What to look for	Question to ask
Task fit	Benchmark measures a generic task unlike yours	"How does this benchmark resemble our task and data?"
Contamination	Test data may have been seen in training	"How do you know the model did not train on the test set?"
Metric choice	Accuracy on imbalanced data (e.g., 99% accuracy when 99% are negatives)	"What are precision and recall on the minority class?"
Averages	Mean hides variance across groups or cases	"What is the worst-performing segment?"

Issue	What to look for	Question to ask
Cherry-picked comparisons	Compared with weak or outdated alternatives	"Why these comparators? Which were left out?"
"Up to X%"	The maximum observed, not the typical	"What is the median, and the range?"
Relative vs absolute	"Errors halved" from 2% to 1%	"What are the absolute numbers?"
Self-reported savings	Time savings from user surveys	"Was time measured, or estimated by users?"
Survey statistics	"80% of CEOs plan to…" from small or sponsored samples	"Who was surveyed, how many, who paid, what exact question?"
Leaderboard volatility	Rankings change monthly	"Does a small ranking difference matter for our decision?"

EXAMPLE · The imbalanced-data trap

A fraud model is "99.2% accurate". Only 0.8% of transactions are fraudulent, so a model that flags *nothing* is also 99.2% accurate. What matters is how much fraud it catches (recall) and how many legitimate transactions it flags (precision / false positives).

5.5 Structured AI vendor evaluation

Dimension	Key questions
Capability fit	Does it address our capability and requirements? On what evidence (ladder level)?
Evidence quality	Can we test on our own representative data before committing? Comparable production references?
Security	Certifications (e.g., ISO/IEC 27001, SOC 2), penetration testing, vulnerability handling, access controls, AI-specific threats (prompt injection, data exfiltration via tools)
Privacy & data use	Data residency, sub-processors, retention, whether client data is used for training, data processing agreement
Governance support	Logging, audit trails, human-oversight features, documentation, model change notifications, support for regulatory obligations
Integration & platform fit	APIs, identity, systems of record, existing platform standards
Total cost of ownership	Licences, usage-based costs at realistic volumes, integration, change management, run and support, exit costs
Contract terms	IP in outputs, liability and indemnities, SLAs, price change protection, termination rights, change-of-control
Exit & portability	Data and configuration export, format, transition assistance, notice periods
Supplier viability	Financial durability, ownership, concentration (Section 5.7)

BEST PRACTICE · Evaluation discipline

- Fix criteria and weights **before** seeing vendor pitches.
- Use the same test data and protocol for every shortlisted vendor (a "bake-off" on client data).
- Separate must-have gates (e.g., EU data residency) from scored criteria.
- Record evidence level for every score.

5.6 Third-party and counterparty risk on multi-year commitments

A three- or five-year AI platform commitment is a bet on a counterparty in a fast-moving market. Risks arise not only from the vendor itself but from its own suppliers — notably the foundation-model provider that many AI products are built on (a **fourth-party** dependency).

Risk	Description	Mitigation
Lock-in	Proprietary formats, workflows and skills make switching costly	Portability clauses, open standards, abstraction layers, exit plan
Upstream model dependency	Vendor depends on a third-party model whose terms, price or availability may change	Disclosure of upstream providers; notice and fallback obligations
Model change / deprecation	Behaviour changes when underlying models are updated or retired	Change-notification duty; regression testing rights; version pinning where possible
Pricing change	Usage-based or renewal prices rise	Price caps, rate cards, volume protections
Concentration	Too many critical processes on one vendor	Portfolio view; multi-vendor strategy for critical capabilities
Regulatory change	New obligations alter vendor's ability or willingness to serve	Compliance cooperation clauses; regulatory change provisions
Change of control	Vendor acquired; roadmap or terms change	Change-of-control termination rights

CONSULTANT'S TIP

Match commitment length to evidence and market volatility. A short initial term with options to extend often beats a large multi-year discount when the capability is still moving quickly.

5.7 Supplier viability and financial durability

Signal	What to examine	Warning signs
Funding & runway	Recent rounds, cash vs burn, profitability path	Runway under ~18 months against a multi-year commitment; repeated down-rounds
Revenue quality	Recurring revenue, net retention, customer count	Revenue mostly from pilots; high churn; reliance on a few customers
Unit economics	Margin after model/inference costs	Negative gross margins sustained by subsidy
Ownership & control	Investors, acquirer interest, strategic owners	Likely acquisition by a competitor of the client
Team & delivery	Key-person dependency, support capacity	Small team supporting many enterprise customers

Signal	What to examine	Warning signs
Dependency position	Thin layer over another provider's model	Product easily replicated by the upstream provider

Where viability is uncertain but the product is compelling, **contract for failure**: escrow or data-export guarantees, step-in or transition assistance, termination rights on insolvency or change of control, shorter terms, and an internal exit plan tested in advance.

> **NOTE**
> Financial due diligence on private companies relies on what the vendor discloses; request it formally, compare it with independent sources, and involve procurement and finance specialists. The advisor's role is to make viability a decision criterion — not to audit accounts.

5.8 Separating signal from noise

"Signals" — analyst forecasts, competitor announcements, conference keynotes, social media — create pressure to act. Treat them as hypotheses to test, not facts to plan on.

Signal	Ask
"Competitor X has deployed AI agents across operations."	Deployed where, at what scale, with what measured result? Press release or production evidence?
"Analysts forecast the market will grow 40% a year."	Forecasting what exactly, with what definition, and does market growth matter for our decision?
"Everyone is moving to platform Y."	Who, how many, in production? What did they give up?
"This changes everything — act now or fall behind."	What is the cost of waiting six months for better evidence, versus the cost of a wrong commitment?

5.9 Writing the evidence-based advisory note

Executives need a short document that states the recommendation up front, shows the strength of evidence honestly and makes the decision easy to take. One to two pages.

Section	Content
1 · Recommendation (bottom line up front)	One or two sentences: what to do
2 · Decision required	What, by whom, by when; what happens if no
3 · Evidence	Key facts, each graded by evidence level (ladder) and source
4 · Confidence	High / medium / low, and why
5 · Risks and conditions	Conditions attached to the recommendation; key risks and mitigations
6 · What would change our view	Specific evidence that would reverse or alter the recommendation
7 · Next steps	Actions, owners, dates

KEY IDEA · Evidence grading scale (example)

Grade	Meaning
A	Measured on client data under realistic conditions (pilot/production)
B	Comparable production references verified independently
C	PoC on client data, or credible independent benchmark with good task fit
D	Vendor-run demo or vendor-reported figures
E	Marketing claim, anecdote, unsourced statistic

PITFALL · False balance and false certainty

Do not present a grade-E claim next to a grade-A measurement as if they were equivalent. And do not convert uncertain evidence into confident language to make the note “more executive”. Confidence must match evidence.

5.10 Case study: Corvel Health

CASE STUDY **The 95% clinical-coding claim**

Corvel Health (fictional) · hospital group · 11 hospitals

Situation

A vendor offers an AI clinical-coding product claiming "95% coding accuracy — cut coder workload by 60%". The CFO wants to sign a five-year enterprise agreement this quarter for a 25% discount. You have three weeks.

Read-back findings

- **Task:** the 95% applied to primary diagnosis codes only; secondary codes and procedures (most of the coders' effort) were not included.
- **Data:** 1,200 discharge summaries from one US hospital, English only; Corvel's records are in German with a different coding system.
- **Metric & exclusions:** accuracy was measured on cases where the model's confidence exceeded a threshold — 62% of cases. The rest went to human coders and were excluded.
- **Workload claim:** from a survey of coders at two customers, self-reported.
- **Viability:** Series B company, ~20 months of runway at current burn; product built on a third-party foundation model; no change-of-control clause in the draft contract.
- **References:** four named customers, three of them pilots; one production customer in a different country.

The advisory note (excerpt)

Recommendation: Do not sign the five-year agreement now. Fund a 12-week paid pilot in two hospitals on German records, with an option to a 2-year term if criteria are met.

Decision: CFO and Chief Medical Officer, by 30 June. If no: continue current coding process; revisit in Q1.

Evidence: 95% applies to primary diagnoses on 62% of English US cases (grade D); workload saving self-reported (E); no German-language evidence (none).

Confidence: Low that the headline results transfer to Corvel. Medium that the product could assist coders on primary diagnoses.

Conditions: data processing agreement with EU residency; no training on Corvel data; change-of-control and insolvency exit rights; upstream model disclosure.

What would change our view: pilot achieving ≥ 85% agreement with senior coders on full code sets for ≥ 80% of cases, and measured (not reported) time savings ≥ 25%.

Discussion questions

1. Restate the vendor's 95% claim honestly in one sentence.
2. What is the strongest piece of evidence the vendor has, and where does it sit on the ladder?
3. Why does the 25% discount not justify the five-year commitment?

4. Which counterparty risks are present?

Model answers (brief)

1. “On English US records, for the 62% of cases where it was confident, the system matched the primary diagnosis code 95% of the time.”
2. The single production customer — a comparable-ish reference (level 5) but in a different country and language, so transfer is uncertain.
3. The discount is certain but small relative to the risk of committing for five years to a product unproven on Corvel’s task, language and coding system, from a supplier with limited runway.
4. Supplier viability (runway), upstream model dependency, change of control, lock-in over five years, model change behaviour.

5.11 Consulting templates

T5.1 Claim Read-Back Worksheet

Purpose: Trace a headline claim to its measurement conditions and restate it honestly.

Claim (verbatim) and source: .....

Condition	What we found	Matches our context? (Y / N / ?)
Task definition		
Data (whose, how many, selection, recency)		
Metric		
Baseline		
Who measured, when, version		
Assistance / tuning		
Exclusions / coverage		
Spread across segments		

Evidence ladder level: .....

Evidence grade (A–E): .....

Honest restatement: .....

T5.2 AI Vendor Due Diligence Questionnaire (core)

Purpose: Minimum questions for any shortlisted AI vendor.

CAPABILITY & EVIDENCE

- [ ] Can we test on our own representative data before contract?
- [ ] Production references comparable in task, data and scale — contactable without vendor present?
- [ ] Which models (including third-party foundation models) underpin the product?

SECURITY, PRIVACY & GOVERNANCE

- [ ] Certifications and latest penetration test summary
- [ ] Data residency, sub-processors, retention; is our data used for training?
- [ ] Logging, audit trail and human-oversight features
- [ ] Notification of model changes; regression-testing rights

COMMERCIAL & EXIT

- [ ] Pricing at our expected volumes; price-change protection
- [ ] IP in outputs; liability and indemnities
- [ ] Data and configuration export; transition assistance
- [ ] Termination on change of control or insolvency

VIABILITY

- [ ] Funding, runway and path to profitability (as disclosed)
- [ ] Revenue concentration and recurring share
- [ ] Ownership and likely acquirers

T5.3 Vendor Evaluation Scorecard

Purpose: Compare vendors on pre-agreed criteria and weights, with evidence levels.

MUST-HAVE GATES (PASS / FAIL)

Gate	Vendor A	Vendor B	Vendor C
EU data residency			
No training on client data			
Test on our data allowed			

SCORED CRITERIA (1–5, WITH EVIDENCE GRADE)

Criterion	Weight	Vendor A	Vendor B	Vendor C
Capability fit (bake-off result)				
Security & privacy				
Governance support				
Integration / platform fit				
5-year TCO				
Contract & exit terms				
Supplier viability				

T5.4 Advisory Note

Purpose: One- to two-page evidence-based recommendation for an executive.

To / from / date: .....

1 · Recommendation (BLUF): .....

2 · Decision required — what, who, by when; if no: .....

3 · Evidence (each item with grade A–E and source): .....

.....

4 · Confidence (H/M/L) and why: .....

5 · Conditions and key risks: .....

6 · What would change our view: .....

7 · Next steps (action · owner · date): .....

5.12 Best practices and pitfalls

✓ DO THIS

- Read every claim back to its conditions before repeating it.
- Tag evidence with its ladder level and grade.
- Test on your own representative data before committing.
- Fix evaluation criteria before vendor pitches.
- Speak to production references without the vendor.
- Contract for failure: exit, change of control, data export.
- Lead the note with the recommendation; say what would change your view.

× AVOID THIS

- Quoting “up to” or headline percentages in plans.
- Treating a PoC as proof of production value.
- Accepting accuracy on imbalanced data at face value.
- Signing long terms for a discount on unproven capability.
- Ignoring the vendor’s upstream model dependency.
- Converting weak evidence into confident language.
- Treating competitor press releases as evidence.

5.13 Module 5 cheat sheet

MODULE 5 · ONE-PAGE SUMMARY

Due Diligence on Claims, Vendors, and Signals

READ-BACK CONDITIONS

Task · data · metric · baseline · who/when/version · assistance · exclusions · spread. Ask: **do these conditions match ours?**

EVIDENCE LADDER

1. Marketing claim
2. Vendor demo
3. Benchmark / 3rd-party test
4. PoC on client data
5. Comparable production reference
6. Client pilot/production

STATISTIC RED FLAGS

- Benchmark ≠ our task · contamination
- Accuracy on imbalanced data
- Averages hide segments
- "Up to" · relative vs absolute
- Self-reported savings · sponsored surveys

VENDOR EVALUATION

Fit · evidence · security · privacy/data use · governance support · integration · TCO · contract · exit · viability. Gates before scores; same test data for all.

COUNTERPARTY RISK

Lock-in · upstream (4th-party) model dependency · model change/deprecation · pricing · concentration · regulation · change of control.

VIABILITY

Runway · revenue quality · unit economics · ownership · key people · thin-wrapper risk. **Contract for failure.**

ADVISORY NOTE

Recommendation (BLUF) → decision required (who, when, if no) → evidence graded A–E → confidence → conditions & risks → what would change our view → next steps. 1–2 pages. Confidence must match evidence.

5.14 Knowledge check

Answer all questions before turning the page.

K1 **A vendor claims "97% accuracy". What is the best first question?**

- (A) "Under what conditions — task, data, metric, baseline — was this measured?"
- (B) "Is that better than your competitors?"
- (C) "Can you guarantee 97% in our contract?"
- (D) "How many customers do you have?"

K2 **A fraud model is 99% accurate on data where 1% of transactions are fraud. What can you conclude?**

- (A) It catches 99% of fraud.
- (B) It is ready for production.
- (C) It is useless.
- (D) Very little until you see precision and recall on the fraud class.

K3 **Which evidence is strongest for a client's plan?**

- (A) A polished vendor demo using the client's logo
- (B) A time-boxed pilot on the client's production data with pre-agreed criteria
- (C) A top position on a public leaderboard
- (D) An analyst report naming the vendor a leader

K4 **A vendor says its tool saves "up to 60%" of analyst time, based on a user survey. What is the main concern?**

- (A) Analysts always overstate workload.
- (B) 60% is too high to be credible.
- (C) "Up to" reports a maximum, and the savings are self-reported rather than measured.
- (D) Surveys are illegal under GDPR.

K5 **An AI product is a thin layer over a third-party foundation model. Which risk does this especially create?**

- (A) The product will be more accurate than competitors.
- (B) Upstream (fourth-party) dependency: changes to the foundation model's terms, price or behaviour pass through to the client.
- (C) The product cannot be used in Europe.
- (D) The product has no security risks.

K6 **A compelling start-up vendor has about 15 months of runway. The client wants a 5-year commitment. What is the best advice?**

- (A) Proceed with a shorter term plus exit protections: data export, transition assistance, termination on insolvency or change of control.
- (B) Ask the vendor to confirm it will not fail.
- (C) Reject the vendor outright.
- (D) Sign for 5 years to secure the discount.

K7 **What should come first in an evidence-based advisory note to an executive?**

- (A) Market background
- (B) Methodology
- (C) The recommendation (bottom line up front)
- (D) A list of all vendors considered

K8 **A competitor announces it has "deployed AI agents across operations". How should the advisor treat this?**

- (A) As irrelevant
- (B) As proof the client is behind and must act immediately
- (C) As a reason to copy the competitor's vendor
- (D) As a hypothesis to test: where, what scale, what measured result, and the cost of waiting for better evidence

5.15 Knowledge check: answers and explanations

K1 **Answer: A** Review: 5.2

A vendor claims "97% accuracy". What is the best first question?

✓ A "Under what conditions — task, data, metric, baseline — was this measured?" — **Correct.** Reading the claim back to its measurement conditions comes first.

× B "Is that better than your competitors?" — **Wrong.** Comparison is meaningless until you know what was measured.

× C "Can you guarantee 97% in our contract?" — **Wrong.** Contracting a number before understanding it locks in an ill-defined metric.

× D "How many customers do you have?" — **Wrong.** Customer count is an adoption claim, not evidence of performance.

K2 **Answer: D** Review: 5.4

A fraud model is 99% accurate on data where 1% of transactions are fraud. What can you conclude?

× A It catches 99% of fraud. — **Wrong.** Accuracy is not recall; a model flagging nothing would also score 99%.

× B It is ready for production. — **Wrong.** Neither the metric nor the evidence level supports this.

× C It is useless. — **Wrong.** It could be useful; the figure simply does not tell you.

✓ D Very little until you see precision and recall on the fraud class. — **Correct.** On imbalanced data accuracy is uninformative; minority-class metrics matter.

K3 **Answer: B** Review: 5.3

Which evidence is strongest for a client's plan?

× A A polished vendor demo using the client's logo — **Wrong.** Branding does not change the fact that it is a demo on chosen inputs.

✓ B A time-boxed pilot on the client's production data with pre-agreed criteria — **Correct.** This is closest to real conditions — the top of the evidence ladder.

× C A top position on a public leaderboard — **Wrong.** Benchmarks may not fit the task and can be contaminated.

× D An analyst report naming the vendor a leader — **Wrong.** Analyst positioning is market evidence, not performance on the client's task.

K4 **Answer: C** Review: 5.4

A vendor says its tool saves "up to 60%" of analyst time, based on a user survey. What is the main concern?

× A Analysts always overstate workload. — **Wrong.** This is an unfounded generalisation.

× B 60% is too high to be credible. — **Wrong.** The level is not the problem; the method is.

✓ C "Up to" reports a maximum, and the savings are self-reported rather than measured. — **Correct.** Both features inflate the claim; ask for medians and measured time.

× D Surveys are illegal under GDPR. — **Wrong.** This is false; surveys can be lawful.

K5 **Answer: B** Review: 5.6–5.7

An AI product is a thin layer over a third-party foundation model. Which risk does this especially create?

× A The product will be more accurate than competitors. — **Wrong.** The architecture says nothing about relative accuracy.

✓ B Upstream (fourth-party) dependency: changes to the foundation model's terms, price or behaviour pass through to the client. — **Correct.** The client depends indirectly on a provider it has no contract with.

× C The product cannot be used in Europe. — **Wrong.** Location of use is not determined by this architecture.

× D The product has no security risks. — **Wrong.** Additional parties typically add, not remove, risk.

K6 **Answer: A** Review: 5.7

A compelling start-up vendor has about 15 months of runway. The client wants a 5-year commitment. What is the best advice?

✓ **A** Proceed with a shorter term plus exit protections: data export, transition assistance, termination on insolvency or change of control. — **Correct.** Contracting for failure balances value against durability risk.

✗ B Ask the vendor to confirm it will not fail. — **Wrong.** Assurances are not mitigation.

✗ C Reject the vendor outright. — **Wrong.** This may forgo real value; viability risk can be mitigated.

✗ D Sign for 5 years to secure the discount. — **Wrong.** This maximises exposure to the very risk identified.

K7 **Answer: C** Review: 5.9

What should come first in an evidence-based advisory note to an executive?

✗ A Market background — **Wrong.** Background is rarely what the executive needs first.

✗ B Methodology — **Wrong.** Methodology supports the recommendation; leading with it buries the decision.

✓ **C** The recommendation (bottom line up front) — **Correct.** Leading with the recommendation makes the note decision-ready.

✗ D A list of all vendors considered — **Wrong.** This is supporting detail.

K8 **Answer: D** Review: 5.8

A competitor announces it has "deployed AI agents across operations". How should the advisor treat this?

✗ A As irrelevant — **Wrong.** Competitor moves can matter; they should be examined, not ignored.

✗ B As proof the client is behind and must act immediately — **Wrong.** A press release is a signal, not evidence of scale or results.

✗ C As a reason to copy the competitor's vendor — **Wrong.** Copying a choice without its evidence and context is not due diligence.

✓ **D** As a hypothesis to test: where, what scale, what measured result, and the cost of waiting for better evidence — **Correct.** Signals are treated as questions to investigate, not facts to plan on.

R1 Course at a glance: the master cheat sheet

ALL FIVE MODULES ON ONE PAGE

The AI Transformation Advisor

M1 · WHAT ADVISORY IS

Decide **which** AI problems, **for whom**, **in what order**, **under what rules**. Judgment ≠ build; demo ≠ deployment. Structural uncertainty. Three foundations: **data readiness · rules of use & decision rights · organization ready to act**. Pilots stall on undecided decisions. Ask → problem → decision. One named decision-maker per decision.

M2 · CAPABILITY LANDSCAPE

Capability, not product. **Define before counting.** Classes: rules-based (not AI) · ML · DL · GenAI · agentic. Include shadow, embedded and sprawl. Independent axes (value · performance · data · governance · adoption · risk). Inventory ≠ maturity. Gap conditions: absent · ungoverned · underperforming · unadopted · fragmented · dependency-blocked. Multi-input prioritization → waves.

M3 · ADVISORY TOOLKIT

Arc: opportunity assessment → roadmap → business case → trial report → executive recommendation. Artifact test: **what decision · who holds it · if no?** Governance docs (inventory, ownership, oversight specs, audit logs) are owned by the organization. "Complete" depends on gates, scope, decision rights and commercial form.

M4 · FRAMING & SCOPING

Outcome objectives (baseline, target, guardrail, date). Power/interest grid; engage gatekeepers early. Requirements incl. data & governance. Boundary: inclusions · **exclusions** · assumptions · dependencies. Four-part change test. Glossary. Promise decisions, not performance. Frame & refer outside your authority.

M5 · DUE DILIGENCE

Read claims back to conditions (task, data, metric, baseline, who/when, help, exclusions, spread). Evidence ladder. Statistic red flags. Vendor evaluation incl. exit. Counterparty & fourth-party risk. Viability — contract for failure. Advisory note: BLUF, decision, graded evidence, confidence, conditions, what would change our view.

R2 Template index

Code	Template	Use when
T1.1	Ask-to-Problem Reframing Canvas	Any new AI request
T1.2	Three Foundations Quick Check	Before funding a pilot; at gates
T1.3	Decision Rights Register	Engagement start
T1.4	Advisory vs Build Decision Sort	Sorting open questions
T2.1	Inventory Counting Rules Sheet	Before any landscape count
T2.2	AI Capability Inventory Register	Landscape mapping
T2.3	Independent-Axis Scorecard	Scoring capabilities
T2.4	Gap Classification & Prioritization Matrix	Ranking gaps
T2.5	Build / Buy / Partner Grid	Sourcing decisions
T3.1	Deliverable Decision Card	Before drafting any deliverable

Code	Template	Use when
T3.2	Deliverable Arc Map	Engagement planning
T3.3	Phase Gate Checklist	Before and at gates
T3.4	Governance Documentation Handover Register	Engagement close
T4.1	Engagement Charter & Scope Statement	Before work starts
T4.2	Stakeholder Map & Alignment Tracker	Throughout
T4.3	Requirements Capture Sheet	Discovery
T4.4	Change Request Form & Impact Test	Any new request
T4.5	Authority Boundary Card	Specialist questions
T4.6	Engagement Glossary	Charter
T5.1	Claim Read-Back Worksheet	Any headline claim
T5.2	AI Vendor Due Diligence Questionnaire	Shortlisted vendors
T5.3	Vendor Evaluation Scorecard	Selection
T5.4	Advisory Note	Executive recommendations

R3 Glossary

A

Acceptance criteria
Conditions under which a deliverable is accepted as complete, agreed in advance with the accepting party.

Adoption (axis)
The degree to which a deployed AI capability is actually used as intended in the process.

Advisory artifact
A deliverable that names the decision it enables, who holds that decision and what happens if the answer is no. Contrast: report.

Advisory decision
A decision about whether, what, for whom, in what order or under what rules AI is used. Contrast: build decision.

Agentic AI
A system pattern in which a model plans multi-step work and takes actions through tools with some autonomy; usually built on generative AI.

AI capability
A business-relevant ability to perform a class of task using AI, described independently of the product that delivers it.

AI capability inventory
A descriptive register of AI capabilities and their deployments, owners, scores and gaps.

AI maturity / readiness assessment
An evaluative judgment of an organization's overall ability to adopt and scale AI, usually by dimension against staged levels.

AI system inventory
Governance documentation recording each AI system in use, its purpose, owner, risk class and status; maintained by the organization.

AI tool sprawl
Proliferation of overlapping AI tools for the same capability across units, with inconsistent controls and duplicated cost.

AI-washing
Labelling non-AI automation as AI, or overstating AI content, for marketing or internal prestige.

Assumption
A condition taken as true for planning; must be checkable and owned. Part of the scope boundary.

Audit log
Automatically recorded events of an AI system (inputs, outputs, overrides, changes) enabling investigation and accountability.

B

Baseline
The measured current state against which improvement is assessed.

Benchmark
A standardised test used to compare models or products; may not fit a client's task.

Benchmark contamination
When test data has been seen in training, inflating benchmark scores.

Blended score
A single number combining several dimensions; hides which dimension needs remedy.

BLUF
Bottom line up front: leading a document with the recommendation.

Build decision
A decision about how to make a chosen solution work: architecture, model, pipelines, integration, testing.

Build vs buy vs partner
Strategic sourcing choice for a capability, judged on differentiation, data, speed, skills, control, cost and lock-in.

Business case
Deliverable that justifies funding a specific initiative, including run costs, risks, assumptions and alternatives.

C

Change control
Process for logging, testing, assessing and deciding on changes to agreed scope.

Change-of-control clause
Contract term giving rights (e.g., termination) if the supplier is acquired.

Commercial form
The contractual basis of payment: time and materials, capped T&M, fixed fee, outcome-based or retainer.

Counterparty risk
Risk that a contracted party fails to perform, changes terms or ceases to exist over the commitment period.

D

Data readiness
Data that is available, permitted, of sufficient quality, representative of production and owned.

Decision rights
Who is entitled to approve, accept risk, change rules and stop an AI system or initiative.

Deep learning
Machine learning using multi-layer neural networks, effective on unstructured data such as images, audio and text.

Demonstration
Showing a capability on chosen inputs; evidence of possibility, not of production value.

Dependency
Something others must deliver for the engagement to proceed; named with owner and date.

Dependency-blocked (gap)
Gap condition in which a capability cannot progress until a prerequisite is in place.

Deployer
Under the EU AI Act, an organization using an AI system under its authority (distinct from the provider).

Deployment (production)
An AI system operating with real data, users, integration, ownership and governance.

DPIA
Data Protection Impact Assessment under GDPR, required for processing likely to result in high risk to individuals.

E

Embedded AI
AI features inside existing software (e.g., CRM, office suites), often activated through vendor updates.

Engagement glossary
Agreed definitions of contested terms for a specific engagement.

Escalation path
Defined levels, roles and time limits for resolving issues.

Evidence grade
Rating of evidence strength (e.g., A–E) used in advisory notes.

Evidence ladder
Ordering of evidence from marketing claim to client production results by strength for the client's plan.

Exclusion
Something explicitly out of scope; gives the change-request test its teeth.

Executive recommendation
Final deliverable advising scale, pivot or stop, with evidence, confidence and conditions.

F

Fourth-party risk
Risk arising from a supplier's own suppliers, e.g., the foundation-model provider behind an AI product.

G

Gap condition
Structural classification of a capability shortfall: absent, ungoverned, underperforming, unadopted, fragmented, dependency-blocked.

Gatekeeper
Stakeholder with low day-to-day interest but power to block (e.g., legal, CISO, works council).

Generative AI
AI that produces new content (text, code, images, audio), typically using large pre-trained models.

Governance documentation
Records an organization maintains for the life of AI systems: inventory, ownership, oversight specifications, audit logs.

H

Hallucination
Plausible but false output from a generative model.

Human oversight specification
Document defining who supervises an AI system, which outputs they review, how they can intervene or stop it, and required competence.

L

Learning milestone
Planned point at which uncertainty is reduced by evidence and a decision can be taken.

Lock-in
Switching costs that bind a client to a vendor (formats, workflows, skills, contracts).

M

Machine learning
AI that learns statistical patterns from data rather than following hand-written rules.

Multi-input prioritization
Ranking using several visible inputs (value, risk, feasibility, dependency, fit) with explicit weights and sensitivity testing.

O

Opportunity assessment
Early deliverable screening AI opportunities against problems, value, feasibility and readiness.

Organizational readiness to act
A process, owner, trained users and incentives that turn an AI output into changed behaviour.

Outcome-based contract
Commercial form in which payment depends on measured business results.

P

Phase gate
Pre-agreed decision point with criteria, a decision-holder and outcomes (go, conditions, pivot, extend, stop).

Pilot purgatory
The state in which AI initiatives stall between pilot and production because non-technical decisions are not taken.

Platform fit
How well a capability fits existing data/AI platforms, integration, identity, security and skills.

Power/interest grid
Stakeholder analysis tool positioning stakeholders by their power over and interest in the outcome.

Precision / recall
Precision: share of flagged items that are correct. Recall: share of true items that are flagged.

Proof of concept (PoC)
Test of technical feasibility on a sample; not proof of operational value.

Provider
Under the EU AI Act, the organization that develops an AI system and places it on the market or into service.

R

Roadmap
Deliverable sequencing initiatives and enablers over time, with dependencies and gates.

Rules of use
Defined permitted uses of AI outputs, by whom, with what human review and within what limits.

Rules-based automation
Deterministic automation executing hand-written rules (e.g., RPA, workflow engines); not AI.

S

Scope boundary
Inclusions, exclusions, assumptions, dependencies and constraints precise enough to test change requests.

Scope creep
Uncontrolled expansion of scope without corresponding change to time, cost or decisions.

Scope of authority
Questions an advisor may answer directly versus those that must be framed and referred to specialists.

Shadow AI
Use of AI tools not sanctioned by the organization, often via personal accounts.

Stated ask
The solution-shaped request a sponsor brings, distinct from the underlying business problem.

Structural uncertainty
Uncertainty inherent in AI (probabilistic, data-dependent, hard to know in advance, moving frontier) that project management cannot remove.

Supplier viability
A vendor's financial and operational ability to keep serving the client over the commitment period.

T

Total cost of ownership (TCO)
All costs over the life of a solution: licences, usage, integration, change, run, support and exit.

Trial (pilot) report
Deliverable reporting results of a time-boxed trial against pre-agreed criteria under stated conditions.

U

Unsanctioned tool use
Employees using AI tools that were not approved by the organization.

W

Works council
Elected employee representative body (e.g., Betriebsrat in Germany) with information and co-determination rights.

R4 Mock exam

Format: 38 multiple-choice questions · one best answer each · **Time:** 60 minutes · **Suggested pass mark:** 70% (27 correct)

Questions 1–7 cover Module 1, 8–15 Module 2, 16–21 Module 3, 22–28 Module 4, 29–34 Module 5 and 35–38 are integrated scenarios. Answer every question before checking the answer key, which explains why each option is right or wrong.

Q1 **Which sentence best defines AI transformation advisory?**

- A Designing and building AI models that transform business processes.
- B Selecting the best AI vendors for an enterprise.
- C Deciding which AI problems an enterprise should solve, for whom, in what order and under what rules of use.
- D Advising on corporate strategy with a focus on technology.

Q2 **An insurer's innovation team shows the board a GenAI demo that answers policy questions perfectly. The board asks if it can go live next month. What is the most important point to make?**

- A It should go live to gather data.
- B It can go live if the security team approves.
- C Demos are always misleading and should be ignored.
- D The demo shows possibility; deployment requires evidence on real data, integration, an owner, rules of use and oversight.

Q3 **Which is NOT a source of structural uncertainty in AI decisions?**

- A Capabilities and prices change during an engagement
- B The project team uses a new collaboration tool
- C Performance depends on production data that may differ from test data
- D Outputs are probabilistic

Q4 **A demand-forecasting pilot performed well on a cleaned dataset prepared by the data science team. In production, forecasts deteriorate. Which foundation was overestimated?**

- A Data readiness — representativeness of production data
- B Rules of use
- C Decision rights
- D Organizational readiness to act

Q5 **Which practice most directly prevents pilot purgatory?**

- A Naming a production owner and approving rules of use as conditions of pilot funding
- B Running many pilots in parallel
- C Using the latest model version in the pilot
- D Extending pilots until results are perfect

Q6 **A sponsor asks for "a copilot for every employee". During reframing you learn staff spend hours searching for documents and already use unapproved tools. What have you identified?**

- (A) A build decision about licence counts
- (B) A requirement to ban all AI tools
- (C) That AI is unnecessary
- (D) The underlying problems: knowledge retrieval and ungoverned AI use

Q7 **Who should be the single accountable decision-maker for "may an AI system give customers statements about policy coverage?"**

- (A) The advisor
- (B) The vendor
- (C) A senior role with authority over legal and conduct risk, e.g., Head of Legal, consulting the business owner
- (D) The data science lead

Q8 **Which entry is written as a durable capability?**

- (A) "Vendor Z Smart Inbox v3"
- (B) "Classification and routing of incoming customer correspondence"
- (C) "Our 40 Copilot licences"
- (D) "The AI project from 2024"

Q9 **Two analysts produce AI counts of 38 and 142 for the same company. What is the most likely cause?**

- (A) They did not agree a boundary, unit of analysis and counting rules
- (B) One analyst made calculation errors
- (C) One counted only GenAI
- (D) AI usage changed between the two counts

Q10 **A camera system inspects welds using a neural network trained on 200,000 labelled images. How should it be classified?**

- (A) Rules-based automation
- (B) Deep learning
- (C) Agentic AI
- (D) Generative AI

Q11 **An IT-service assistant reads a ticket, checks the user's permissions in the directory, resets a password and closes the ticket. Which class and which governance emphasis apply?**

- (A) Rules-based automation; rule change control
- (B) Generative AI; content filtering
- (C) Agentic AI; permission scoping, action logging, human checkpoints
- (D) Machine learning; drift monitoring

Q12 **What is the best framing for discovering unsanctioned AI use?**

- (A) Blocking all AI websites first, then asking what people used
- (B) An investigation to identify policy violators
- (C) Relying only on the procurement system
- (D) A non-punitive amnesty treating usage as evidence of demand, agreed with HR, legal and employee representatives

Q13 **Why should capability scores be kept on independent axes?**

(A) Because blending is mathematically impossible
(B) Because regulators require six axes
(C) Because executives prefer more numbers
(D) Because different weak axes call for different remedies, which a blend hides

Q14 **A capability works well technically, but 80% of users bypass it. Which gap condition and remedy fit?**

(A) Underperforming — retrain the model
(B) Absent — buy a tool
(C) Unadopted — process redesign, training, incentives, UX
(D) Fragmented — rationalise tools

Q15 **Which statement correctly distinguishes an inventory from a maturity assessment?**

(A) An inventory describes what capabilities exist and their state; a maturity assessment evaluates organizational ability against levels.
(B) An inventory is evaluative; a maturity assessment is descriptive.
(C) They are the same, with different names.
(D) A maturity assessment lists unsanctioned tools.

Q16 **At which stage of the arc does the business case sit, and which decision does it enable?**

(A) Scale — whether to roll out
(B) Commit — whether to fund a specific initiative on stated assumptions
(C) Explore — which problems to pursue
(D) Prove — whether a trial met its criteria

Q17 **Which question is part of the test that distinguishes an advisory artifact from a report?**

(A) Was it reviewed by a partner?
(B) How many pages does it have?
(C) What happens if the answer is no?
(D) Which template was used?

Q18 **Which item is governance documentation rather than a delivery deliverable?**

(A) Opportunity assessment
(B) Business case
(C) Executive recommendation
(D) Ownership record for a deployed AI system

Q19 **A trial report is accepted, and the decision-holder decides to stop the initiative. Under a fixed-fee contract for the report, which statement is true?**

(A) The deliverable is incomplete because the outcome was negative.
(B) The deliverable is complete; acceptance and the business decision are separate.
(C) The advisor must refund the fee.
(D) The contract converts to outcome-based.

Q20 **Which phase-gate design is strongest?**

(A) Criteria agreed before the phase; one named decision-holder; outcomes include go, pivot, extend and stop
(B) Criteria agreed after results; steering committee decides by consensus
(C) No criteria; sponsor decides on instinct
(D) Criteria set by the vendor

Q21 **An advisor is offered an outcome-based fee tied to "30% cost reduction from AI". What is the key risk to raise?**

(A) The fee may be too low.
(B) Outcome-based fees are illegal.
(C) The outcome depends on factors outside the advisor's control, such as data, adoption and client decisions, creating attribution disputes.
(D) The advisor will have to build the system.

Q22 **Which characteristic of AI work most justifies scoping decisions and learning milestones instead of performance targets?**

(A) AI projects need more stakeholders
(B) Outcomes are uncertain until tested on real data
(C) AI vendors prefer it
(D) Agile methodology requires it

Q23 **Which element is most often missing from scope statements that fail to stop scope creep?**

(A) Explicit exclusions
(B) Inclusions
(C) Project logo
(D) Team biographies

Q24 **A plant director asks whether the predictive-maintenance trial can also detect quality defects. Defect detection is listed as an exclusion. What should the advisor do?**

(A) Agree, since it uses the same sensors
(B) Ask the data scientists to explore it informally
(C) Refuse and close the topic
(D) Log it as a change request touching an exclusion and propose it for a later phase or a formal change decision

Q25 **Which stakeholder is typically high power, low interest in an AI engagement?**

(A) Peripheral functions
(B) The executive sponsor
(C) End users
(D) The CISO who must approve vendor access

Q26 **Which requirement type is most often forgotten in AI engagements and causes late blocks?**

(A) Meeting cadence
(B) Governance and risk requirements (oversight, logging, approvals)
(C) Business requirements
(D) Font preferences

Q27 **A dispute arises because "automate" meant "draft for human review" to the team and "replace the human" to the sponsor. What would have prevented it?**

- (A) An engagement glossary defining levels of automation
- (B) More frequent status meetings
- (C) A different vendor
- (D) A longer contract

Q28 **Which question should the advisor answer directly rather than refer?**

- (A) Whether co-determination rights apply under German law
- (B) Whether the system is robust to prompt injection
- (C) How the decision at Gate 2 should be structured and who should hold it
- (D) Whether a DPIA is legally required

Q29 **A vendor reports "accuracy 94%" measured only on cases where the model was confident (55% of all cases). What is the honest restatement?**

- (A) "94% accurate on the 55% of cases it accepted; the remainder went to humans."
- (B) "94% accurate overall."
- (C) "55% accurate."
- (D) "Accuracy unknown."

Q30 **Which is the weakest evidence for a client's plan?**

- (A) A pilot on client production data
- (B) A production reference in a comparable organization, verified independently
- (C) A PoC on the client's own data
- (D) A vendor's case study stating "up to 70% savings"

Q31 **A leaderboard shows Model A scoring 0.4 points higher than Model B on a general benchmark. How much should this weigh in vendor selection?**

- (A) Decisively — always choose the top model
- (B) Little, unless the benchmark fits the task; test both on client data
- (C) Only if the vendor confirms it
- (D) Not at all — benchmarks are fraudulent

Q32 **Which contract provision most directly mitigates upstream model dependency?**

- (A) A logo-usage clause
- (B) A volume discount
- (C) Disclosure of upstream model providers plus notification and regression-testing rights when models change
- (D) A longer contract term

Q33 **Which pattern is a supplier-viability warning sign?**

- (A) Transparent disclosure of upstream providers
- (B) Negative gross margins sustained by investor funding and short runway
- (C) Diverse customer base with high recurring revenue
- (D) Independent security certification

Q34 **Which section of an advisory note most protects against overconfidence?**

(A) What would change our view
(B) Market background
(C) Appendix of screenshots
(D) Executive summary of vendors

Q35 **Integrated scenario: A retailer's CEO wants to sign a 5-year agentic AI platform deal announced at a conference, "before competitors do". There is no inventory, no defined problem and no trial. Which sequence of advisory moves is best?**

(A) Run a market survey of CEOs
(B) Sign, then define use cases on the platform
(C) Reframe the ask to business problems; map current capabilities including shadow AI; test the vendor's claims and viability; propose a gated trial before any long commitment
(D) Refuse to engage until AI maturity reaches level 4

Q36 **Integrated scenario: A trial met its criteria, but the business unit refuses to own the system in production and Legal has not approved rules of use. What should the executive recommendation say?**

(A) Scale immediately — the trial succeeded.
(B) Extend the trial indefinitely.
(C) Stop — the trial is irrelevant.
(D) Go with conditions: production owner named and rules of use approved before rollout; otherwise defer.

Q37 **Integrated scenario: Halfway through an engagement, the sponsor asks the advisor to "just build the prototype yourselves, it'll be faster". What is the best response?**

(A) Ask the vendor to build it for free.
(B) Agree — speed matters most.
(C) Refuse to discuss prototypes.
(D) Treat it as a change request: clarify which decision a prototype would serve, assess impact on scope and roles, and agree who builds it.

Q38 **Integrated scenario: The board wants "one number" for AI maturity. What is the best advisory response?**

(A) Provide a blended score to satisfy the request.
(B) Offer a short capability profile on independent axes plus top gap conditions, explaining why one number would mislead.
(C) Report the number of AI tools.
(D) Refuse — boards should not see AI metrics.

R5 Mock exam: answer key and explanations

Quick answer grid

1 · C	**2** · D	**3** · B	**4** · A	**5** · A	**6** · D	**7** · C	**8** · B	**9** · A	**10** · B
11 · C	**12** · D	**13** · D	**14** · C	**15** · A	**16** · B	**17** · C	**18** · D	**19** · B	**20** · A
21 · C	**22** · B	**23** · A	**24** · D	**25** · D	**26** · B	**27** · A	**28** · C	**29** · A	**30** · D
31 · B	**32** · C	**33** · B	**34** · A	**35** · C	**36** · D	**37** · D	**38** · B		

Score	Interpretation	Next step
32–38	Exam-ready	Review any missed items; practise templates on a live case
27–31	Pass level	Re-read the cheat sheets for modules where you missed two or more
below 27	Not yet	Re-study weak modules using the “Review” references; retake in one week

Full explanations

Q1 **Answer: C** Review: M1 · 1.1

Which sentence best defines AI transformation advisory?

✗ (A) Designing and building AI models that transform business processes. — **Wrong.** That describes implementation.

✗ (B) Selecting the best AI vendors for an enterprise. — **Wrong.** Vendor selection is one activity, not the definition.

✓ (C) Deciding which AI problems an enterprise should solve, for whom, in what order and under what rules of use. — **Correct.** This is the course definition: a decision discipline.

✗ (D) Advising on corporate strategy with a focus on technology. — **Wrong.** This describes management or technology strategy consulting generally.

Q2 **Answer: D** Review: M1 · 1.3

An insurer’s innovation team shows the board a GenAI demo that answers policy questions perfectly. The board asks if it can go live next month. What is the most important point to make?

✗ (A) It should go live to gather data. — **Wrong.** Going live without rules of use or an owner exposes the insurer to unmanaged risk.

✗ (B) It can go live if the security team approves. — **Wrong.** Security is one of many gaps.

✗ (C) Demos are always misleading and should be ignored. — **Wrong.** Demos are legitimate evidence of possibility; they are not to be ignored.

✓ (D) The demo shows possibility; deployment requires evidence on real data, integration, an owner, rules of use and oversight. — **Correct.** This draws the demonstration-vs-deployment distinction that matters here.

Q3 **Answer: B** Review: M1 · 1.4

Which is NOT a source of structural uncertainty in AI decisions?

✗ (A) Capabilities and prices change during an engagement — **Wrong.** This is a source (moving frontier).

✓ (B) The project team uses a new collaboration tool — **Correct.** This is a delivery detail, not uncertainty inherent in AI.

✗ (C) Performance depends on production data that may differ from test data — **Wrong.** This is a source.

✗ (D) Outputs are probabilistic — **Wrong.** This is a source of structural uncertainty.

Q4 **Answer: A** Review: M1 · 1.5

A demand-forecasting pilot performed well on a cleaned dataset prepared by the data science team. In production, forecasts deteriorate. Which foundation was overestimated?

✓ A Data readiness — representativeness of production data — **Correct.** The pilot data did not represent production conditions.

× B Rules of use — **Wrong.** Rules of use are not implicated by data quality differences.

× C Decision rights — **Wrong.** Nothing suggests confusion about who decides.

× D Organizational readiness to act — **Wrong.** The problem appears before any action is taken.

Q5 **Answer: A** Review: M1 · 1.6

Which practice most directly prevents pilot purgatory?

✓ A Naming a production owner and approving rules of use as conditions of pilot funding — **Correct.** This brings the usually undecided decisions forward.

× B Running many pilots in parallel — **Wrong.** More pilots without decisions produce more stalls.

× C Using the latest model version in the pilot — **Wrong.** Model choice rarely causes the stall.

× D Extending pilots until results are perfect — **Wrong.** This prolongs the stall.

Q6 **Answer: D** Review: M1 · 1.7

A sponsor asks for "a copilot for every employee". During reframing you learn staff spend hours searching for documents and already use unapproved tools. What have you identified?

× A A build decision about licence counts — **Wrong.** Licence counting is premature.

× B A requirement to ban all AI tools — **Wrong.** A ban is a premature solution, and one that tends to drive use underground.

× C That AI is unnecessary — **Wrong.** The problems may well suit AI; that has not been decided yet.

✓ D The underlying problems: knowledge retrieval and ungoverned AI use — **Correct.** Reframing has uncovered problems the ask was standing in for.

Q7 **Answer: C** Review: M1 · 1.8

Who should be the single accountable decision-maker for "may an AI system give customers statements about policy coverage?"

× A The advisor — **Wrong.** The advisor facilitates; they do not own client decisions.

× B The vendor — **Wrong.** The vendor cannot decide the client's rules of use.

✓ C A senior role with authority over legal and conduct risk, e.g., Head of Legal, consulting the business owner — **Correct.** Rules of use with legal consequences belong to a named role with that authority.

× D The data science lead — **Wrong.** This is a legal and business-risk decision, not a technical one.

Q8 **Answer: B** Review: M2 · 2.1

Which entry is written as a durable capability?

× A "Vendor Z Smart Inbox v3" — **Wrong.** This names a product and version.

✓ B "Classification and routing of incoming customer correspondence" — **Correct.** It describes a task class independent of tools.

× C "Our 40 Copilot licences" — **Wrong.** This is a licence count.

× D "The AI project from 2024" — **Wrong.** A project label does not describe a capability.

Q9 **Answer: A** Review: M2 · 2.2

Two analysts produce AI counts of 38 and 142 for the same company. What is the most likely cause?

✓ **A** They did not agree a boundary, unit of analysis and counting rules — **Correct.** Without shared counting rules, numbers diverge widely.

✗ **B** One analyst made calculation errors — **Wrong.** The size of the difference points to different definitions, not arithmetic.

✗ **C** One counted only GenAI — **Wrong.** This is possible but it is a symptom of the root cause: no agreed rules.

✗ **D** AI usage changed between the two counts — **Wrong.** This would not plausibly explain a near fourfold difference.

Q10 **Answer: B** Review: M2 · 2.3

A camera system inspects welds using a neural network trained on 200,000 labelled images. How should it be classified?

✗ **A** Rules-based automation — **Wrong.** It learns from data rather than following written rules.

✓ **B** Deep learning — **Correct.** A neural network on image data is deep learning.

✗ **C** Agentic AI — **Wrong.** It takes no autonomous multi-step actions through tools.

✗ **D** Generative AI — **Wrong.** It classifies; it does not generate content.

Q11 **Answer: C** Review: M2 · 2.3

An IT-service assistant reads a ticket, checks the user's permissions in the directory, resets a password and closes the ticket. Which class and which governance emphasis apply?

✗ **A** Rules-based automation; rule change control — **Wrong.** The multi-step planning by a model goes beyond fixed rules.

✗ **B** Generative AI; content filtering — **Wrong.** It acts, not just generates, so content filtering is insufficient.

✓ **C** Agentic AI; permission scoping, action logging, human checkpoints — **Correct.** Taking actions through tools calls for controls on actions.

✗ **D** Machine learning; drift monitoring — **Wrong.** Drift monitoring is relevant but not the core governance emphasis for an acting system.

Q12 **Answer: D** Review: M2 · 2.4

What is the best framing for discovering unsanctioned AI use?

✗ **A** Blocking all AI websites first, then asking what people used — **Wrong.** Blocking first destroys trust and evidence.

✗ **B** An investigation to identify policy violators — **Wrong.** Fear drives usage underground and corrupts the inventory.

✗ **C** Relying only on the procurement system — **Wrong.** Procurement misses free tools and personal accounts.

✓ **D** A non-punitive amnesty treating usage as evidence of demand, agreed with HR, legal and employee representatives — **Correct.** This maximises disclosure and respects constraints.

Q13 **Answer: D** Review: M2 · 2.6

Why should capability scores be kept on independent axes?

✗ **A** Because blending is mathematically impossible — **Wrong.** Blending is possible, just misleading.

✗ **B** Because regulators require six axes — **Wrong.** No regulation prescribes these axes.

✗ **C** Because executives prefer more numbers — **Wrong.** Preference is not the reason.

✓ **D** Because different weak axes call for different remedies, which a blend hides — **Correct.** Visibility of the specific weakness drives the right action.

Q14 **Answer: C** Review: M2 · 2.8

A capability works well technically, but 80% of users bypass it. Which gap condition and remedy fit?

✗ (A) Underperforming — retrain the model — **Wrong.** Performance is fine.

✗ (B) Absent — buy a tool — **Wrong.** The capability exists.

✓ (C) Unadopted — process redesign, training, incentives, UX — **Correct.** It works but is not used as intended.

✗ (D) Fragmented — rationalise tools — **Wrong.** There is no evidence of duplication.

Q15 **Answer: A** Review: M2 · 2.7

Which statement correctly distinguishes an inventory from a maturity assessment?

✓ (A) An inventory describes what capabilities exist and their state; a maturity assessment evaluates organizational ability against levels. — **Correct.** This is the correct distinction.

✗ (B) An inventory is evaluative; a maturity assessment is descriptive. — **Wrong.** This reverses them.

✗ (C) They are the same, with different names. — **Wrong.** They answer different questions.

✗ (D) A maturity assessment lists unsanctioned tools. — **Wrong.** Maturity assessments do not catalogue specific tools.

Q16 **Answer: B** Review: M3 · 3.1–3.2

At which stage of the arc does the business case sit, and which decision does it enable?

✗ (A) Scale — whether to roll out — **Wrong.** That is the executive recommendation.

✓ (B) Commit — whether to fund a specific initiative on stated assumptions — **Correct.** The business case enables the funding decision.

✗ (C) Explore — which problems to pursue — **Wrong.** That is the opportunity assessment.

✗ (D) Prove — whether a trial met its criteria — **Wrong.** That is the trial report.

Q17 **Answer: C** Review: M3 · 3.3

Which question is part of the test that distinguishes an advisory artifact from a report?

✗ (A) Was it reviewed by a partner? — **Wrong.** Quality review is not the test.

✗ (B) How many pages does it have? — **Wrong.** Length is irrelevant.

✓ (C) What happens if the answer is no? — **Correct.** This is one of the three test questions (with “what decision?” and “who holds it?”).

✗ (D) Which template was used? — **Wrong.** Format is irrelevant.

Q18 **Answer: D** Review: M3 · 3.4

Which item is governance documentation rather than a delivery deliverable?

✗ (A) Opportunity assessment — **Wrong.** This is a delivery deliverable.

✗ (B) Business case — **Wrong.** This is a delivery deliverable.

✗ (C) Executive recommendation — **Wrong.** This is a delivery deliverable.

✓ (D) Ownership record for a deployed AI system — **Correct.** Ownership records persist for the system’s life and are maintained by the organization.

Q19 **Answer: B** Review: M3 · 3.5–3.6

A trial report is accepted, and the decision-holder decides to stop the initiative. Under a fixed-fee contract for the report, which statement is true?

✗ (A) The deliverable is incomplete because the outcome was negative. — **Wrong.** This confuses deliverable with outcome.

✓ (B) The deliverable is complete; acceptance and the business decision are separate. — **Correct.** Fixed fee pays for accepted deliverables, whatever the decision.

✗ (C) The advisor must refund the fee. — **Wrong.** There is no basis for this in a fixed-fee form.

✗ (D) The contract converts to outcome-based. — **Wrong.** Contracts do not change form automatically.

Q20 **Answer: A** Review: M3 · 3.5

Which phase-gate design is strongest?

✓ (A) Criteria agreed before the phase; one named decision-holder; outcomes include go, pivot, extend and stop — **Correct.** Pre-agreed criteria, single accountability and real options make a strong gate.

✗ (B) Criteria agreed after results; steering committee decides by consensus — **Wrong.** Moving goalposts plus diffuse accountability.

✗ (C) No criteria; sponsor decides on instinct — **Wrong.** This produces no evidence discipline.

✗ (D) Criteria set by the vendor — **Wrong.** This creates a conflict of interest.

Q21 **Answer: C** Review: M3 · 3.6

An advisor is offered an outcome-based fee tied to "30% cost reduction from AI". What is the key risk to raise?

✗ (A) The fee may be too low. — **Wrong.** Fee level is secondary.

✗ (B) Outcome-based fees are illegal. — **Wrong.** They are lawful.

✓ (C) The outcome depends on factors outside the advisor's control, such as data, adoption and client decisions, creating attribution disputes. — **Correct.** This is the characteristic risk of outcome-based forms in AI.

✗ (D) The advisor will have to build the system. — **Wrong.** Not necessarily; the risk is attribution and control.

Q22 **Answer: B** Review: M4 · 4.1

Which characteristic of AI work most justifies scoping decisions and learning milestones instead of performance targets?

✗ (A) AI projects need more stakeholders — **Wrong.** This is true, but it is not the reason for this approach.

✓ (B) Outcomes are uncertain until tested on real data — **Correct.** Uncertain outcomes mean you can promise decisions but not performance.

✗ (C) AI vendors prefer it — **Wrong.** Vendor preference is irrelevant.

✗ (D) Agile methodology requires it — **Wrong.** Methodology is not the driver.

Q23 **Answer: A** Review: M4 · 4.5

Which element is most often missing from scope statements that fail to stop scope creep?

✓ (A) Explicit exclusions — **Correct.** Without exclusions, borderline requests cannot be tested.

✗ (B) Inclusions — **Wrong.** Inclusions are usually present.

✗ (C) Project logo — **Wrong.** This is irrelevant.

✗ (D) Team biographies — **Wrong.** These are irrelevant to scope.

Q24 **Answer: D** Review: M4 · 4.5, 4.8

A plant director asks whether the predictive-maintenance trial can also detect quality defects. Defect detection is listed as an exclusion. What should the advisor do?

✗ A Agree, since it uses the same sensors — **Wrong.** That bypasses the boundary.

✗ B Ask the data scientists to explore it informally — **Wrong.** This is a corridor scope change.

✗ C Refuse and close the topic — **Wrong.** The idea may be valuable; refusing outright wastes it.

✓ D Log it as a change request touching an exclusion and propose it for a later phase or a formal change decision — **Correct.** This applies the change test and keeps the idea alive.

Q25 **Answer: D** Review: M4 · 4.3

Which stakeholder is typically high power, low interest in an AI engagement?

✗ A Peripheral functions — **Wrong.** These are low power, low interest.

✗ B The executive sponsor — **Wrong.** The sponsor is high power, high interest.

✗ C End users — **Wrong.** They are usually high interest and lower power.

✓ D The CISO who must approve vendor access — **Correct.** A classic gatekeeper: keep satisfied.

Q26 **Answer: B** Review: M4 · 4.4

Which requirement type is most often forgotten in AI engagements and causes late blocks?

✗ A Meeting cadence — **Wrong.** This is not a requirement type.

✓ B Governance and risk requirements (oversight, logging, approvals) — **Correct.** Omitting them leads to late legal or risk blocks.

✗ C Business requirements — **Wrong.** These are usually captured.

✗ D Font preferences — **Wrong.** These are irrelevant.

Q27 **Answer: A** Review: M4 · 4.6

A dispute arises because "automate" meant "draft for human review" to the team and "replace the human" to the sponsor. What would have prevented it?

✓ A An engagement glossary defining levels of automation — **Correct.** Shared vocabulary prevents disputes that are really about the meaning of words.

✗ B More frequent status meetings — **Wrong.** This might reveal the mismatch earlier but does not prevent it.

✗ C A different vendor — **Wrong.** The problem is internal vocabulary.

✗ D A longer contract — **Wrong.** Length does not fix definitions.

Q28 **Answer: C** Review: M4 · 4.9

Which question should the advisor answer directly rather than refer?

✗ A Whether co-determination rights apply under German law — **Wrong.** This is an employment-law question to refer.

✗ B Whether the system is robust to prompt injection — **Wrong.** This needs security testing specialists.

✓ C How the decision at Gate 2 should be structured and who should hold it — **Correct.** Decision structure and decision rights are core advisory competence.

✗ D Whether a DPIA is legally required — **Wrong.** This belongs to the DPO or privacy counsel.

Q29 **Answer: A** Review: M5 · 5.2

A vendor reports "accuracy 94%" measured only on cases where the model was confident (55% of all cases). What is the honest restatement?

✓ **A** "94% accurate on the 55% of cases it accepted; the remainder went to humans." — **Correct.** This states both accuracy and coverage.

✗ B "94% accurate overall." — **Wrong.** This ignores the excluded 45%.

✗ C "55% accurate." — **Wrong.** This conflates coverage with accuracy.

✗ D "Accuracy unknown." — **Wrong.** Some evidence exists; it just needs honest framing.

Q30 **Answer: D** Review: M5 · 5.3–5.4

Which is the weakest evidence for a client's plan?

✗ A A pilot on client production data — **Wrong.** The strongest of these.

✗ B A production reference in a comparable organization, verified independently — **Wrong.** Fairly strong.

✗ C A PoC on the client's own data — **Wrong.** Moderate: client data, but not operational.

✓ **D** A vendor's case study stating "up to 70% savings" — **Correct.** Vendor-authored marketing with a maximum figure is the weakest.

Q31 **Answer: B** Review: M5 · 5.4

A leaderboard shows Model A scoring 0.4 points higher than Model B on a general benchmark. How much should this weigh in vendor selection?

✗ A Decisively — always choose the top model — **Wrong.** Small differences on general benchmarks rarely transfer to a specific task.

✓ **B** Little, unless the benchmark fits the task; test both on client data — **Correct.** Task fit and client-data testing matter more than small ranking gaps.

✗ C Only if the vendor confirms it — **Wrong.** Vendor confirmation adds nothing.

✗ D Not at all — benchmarks are fraudulent — **Wrong.** Benchmarks have value; the point is their limits.

Q32 **Answer: C** Review: M5 · 5.6

Which contract provision most directly mitigates upstream model dependency?

✗ A A logo-usage clause — **Wrong.** Irrelevant.

✗ B A volume discount — **Wrong.** This is a commercial benefit, not a mitigation.

✓ **C** Disclosure of upstream model providers plus notification and regression-testing rights when models change — **Correct.** This addresses fourth-party and model-change risk.

✗ D A longer contract term — **Wrong.** A longer term increases exposure.

Q33 **Answer: B** Review: M5 · 5.7

Which pattern is a supplier-viability warning sign?

✗ A Transparent disclosure of upstream providers — **Wrong.** A positive transparency signal.

✓ **B** Negative gross margins sustained by investor funding and short runway — **Correct.** Signals of financial fragility.

✗ C Diverse customer base with high recurring revenue — **Wrong.** A positive signal.

✗ D Independent security certification — **Wrong.** A positive security signal, not a viability one.

Q34 **Answer: A** Review: M5 · 5.9

Which section of an advisory note most protects against overconfidence?

✓ A What would change our view — **Correct.** Naming the evidence that would reverse the recommendation makes uncertainty explicit.

× B Market background — **Wrong.** This does not test the recommendation.

× C Appendix of screenshots — **Wrong.** This is supporting material only.

× D Executive summary of vendors — **Wrong.** This is descriptive.

Q35 **Answer: C** Review: M1–M5

Integrated scenario: A retailer's CEO wants to sign a 5-year agentic AI platform deal announced at a conference, "before competitors do". There is no inventory, no defined problem and no trial. Which sequence of advisory moves is best?

× A Run a market survey of CEOs — **Wrong.** Signals are not evidence for this decision.

× B Sign, then define use cases on the platform — **Wrong.** Product before problem: maximum lock-in, least evidence.

✓ C Reframe the ask to business problems; map current capabilities including shadow AI; test the vendor's claims and viability; propose a gated trial before any long commitment — **Correct.** This applies Modules 1, 2, 5 and 3/4 in order.

× D Refuse to engage until AI maturity reaches level 4 — **Wrong.** This is unhelpful, and maturity level is not a gate for problem framing.

Q36 **Answer: D** Review: M1 · 1.5–1.6, M3 · 3.5

Integrated scenario: A trial met its criteria, but the business unit refuses to own the system in production and Legal has not approved rules of use. What should the executive recommendation say?

× A Scale immediately — the trial succeeded. — **Wrong.** This ignores two of the three foundations.

× B Extend the trial indefinitely. — **Wrong.** This creates pilot purgatory.

× C Stop — the trial is irrelevant. — **Wrong.** This discards valid evidence.

✓ D Go with conditions: production owner named and rules of use approved before rollout; otherwise defer. — **Correct.** The evidence supports value, but the foundations must be secured — a conditional recommendation.

Q37 **Answer: D** Review: M1 · 1.2–1.3, M4 · 4.8

Integrated scenario: Halfway through an engagement, the sponsor asks the advisor to "just build the prototype yourselves, it'll be faster". What is the best response?

× A Ask the vendor to build it for free. — **Wrong.** This creates obligations and conflicts of interest.

× B Agree — speed matters most. — **Wrong.** This drifts into implementation without mandate, budget or accountability.

× C Refuse to discuss prototypes. — **Wrong.** A prototype may be useful if it is tied to a decision.

✓ D Treat it as a change request: clarify which decision a prototype would serve, assess impact on scope and roles, and agree who builds it. — **Correct.** This keeps advisory judgment distinct from build and applies change control.

Q38 **Answer: B** Review: M2 · 2.6–2.7

Integrated scenario: The board wants "one number" for AI maturity. What is the best advisory response?

× A Provide a blended score to satisfy the request. — **Wrong.** This hides which remedy is needed.

✓ B Offer a short capability profile on independent axes plus top gap conditions, explaining why one number would mislead. — **Correct.** This meets the board's need for simplicity without hiding the decision-relevant information.

× C Report the number of AI tools. — **Wrong.** A tool count is neither maturity nor capability.

× D Refuse — boards should not see AI metrics. — **Wrong.** This is unhelpful.

R6 Further reading and next steps

The following public frameworks and texts are useful companions to this course. Check for the latest versions, since several are updated regularly.

Resource	Why it is useful
NIST AI Risk Management Framework (AI RMF 1.0) and its Generative AI Profile	Vocabulary and structure (Govern, Map, Measure, Manage) for AI risk across the lifecycle
ISO/IEC 42001 — AI management systems	What a management system for AI expects: roles, policies, records, controls
EU AI Act — Regulation (EU) 2024/1689	Risk-based obligations for providers and deployers; essential context for European clients (confirm current application dates)
OECD AI Principles and OECD definition of an AI system	Widely referenced definitions useful for inventory boundaries
GDPR and national data protection guidance on AI	Purpose limitation, DPIAs and data use constraints that shape data readiness

Practice assignments

1. Take a real or imagined AI request and complete T1.1 and T1.3. Can you name one accountable person for every decision?
2. Build a ten-row capability inventory (T2.2) for an organization you know, including at least two unsanctioned or embedded tools.
3. Pick a published AI vendor claim and complete a Claim Read-Back Worksheet (T5.1); then write a half-page advisory note (T5.4).
4. Write a scope statement (T4.1) with at least five explicit exclusions, then test three hypothetical change requests against it.

CONSULTANT'S TIP · Coming next in the certificate

Later courses build on this foundation, turning the deliverable arc in Module 3 into production methods. Keep your completed templates; you will reuse them.

ENTERPRISE AI TRANSFORMATION AND CONSULTING PROFESSIONAL

COURSE 2 · STUDY BOOK

Identify and Prioritize AI Opportunities

How to find AI opportunities in the way work actually gets done, screen them quickly and fairly, prioritize them with a defensible value-effort lens, and assemble a balanced portfolio that can be funded — or correctly told “not this cycle”.

4 MODULES · CASE STUDIES · TEMPLATES · CHEAT SHEETS · MOCK EXAM

Study edition · 2026

Includes full answer explanations

How to use this book

This study book accompanies **Course 2, Identify and Prioritize AI Opportunities**, of the Enterprise AI Transformation and Consulting Professional certificate. It is written for consultants, business analysts, AI strategy and transformation leads, and portfolio or investment owners who must decide which AI initiatives an enterprise funds first.

Each module follows the same rhythm so that you always know where you are:

Element	What you get	How to use it
Module opener	Learning objectives, section list, study time	Read first; return to it to self-test at the end
Overview & key takeaways	The module on one screen	Skim before, re-read after
Core sections	Concepts, tables, diagrams, worked examples	Study actively: annotate, restate in your own words
Callout boxes	Key ideas, examples, pitfalls, exam traps, consultant tips	Exam traps are the most frequently tested ideas
Case study	A realistic enterprise scenario with model answers	Answer the questions before reading the model answers
Templates	Fill-in consulting tools (T-codes), one per page or less	Copy into your own engagements
Do / Avoid	Best practices and pitfalls side by side	Use as a pre-meeting checklist
Cheat sheet	One-page module summary	Revise from these in the final week
Knowledge check	Scenario questions with explanations for **every** option	Understanding why the wrong answers are wrong matters as much as getting the right one

Callout key

KEY IDEA
The idea to remember if you remember nothing else.

EXAM TRAP
A frequently tested distinction or a tempting wrong answer.

EXAMPLE
A worked illustration of the concept.

CONSULTANT'S TIP
Practical wording or technique from the field.

PITFALL
A common mistake made in real engagements.

DEFINITION
A precise definition worth learning word for word.

> **NOTE · Builds on Course 1**
> This course assumes the vocabulary of Course 1, The AI Transformation Advisor: advisory vs build decisions, the three foundations, demonstration vs deployment, capability inventories, gap conditions and the evidence ladder. Where a concept is reused, the book reminds you briefly rather than re-teaching it.

Suggested study plan (four weeks)

Week	Study	Practise
1	Module 1	Map one real workflow with T1.1 and draft three opportunity cards with ledgers
2	Module 2	Normalize and screen five proposals with T2.1–T2.5
3	Module 3	Set criteria and weights, estimate effort by component, plot a matrix
4	Module 4, then revise all cheat sheets and the glossary	Assemble a portfolio and decision package; take the mock exam (60 min) and review every explanation

> **NOTE · About the examples**
> All organizations in case studies and examples are fictional, and their figures are illustrative. Regulatory and framework references are for orientation only; always check current law and guidance, and refer legal interpretation to qualified specialists.

Contents

FRONT **How to use this book** 2

MODULE 1 **Finding AI Opportunities** 6
Work-first, not technology-first 7
Workflow mapping 8
Process friction analysis 8
Locating candidates by business function and industry 9
AI assistants, copilots and agents 10
Connecting each candidate to a concrete source of value 11
Building the opportunity inventory 11
Estimating value and effort as two separate, arguable judgments 12
Naming the ledger: where the effort goes 13
Case study: Alpenhof Energie 14
Consulting templates 16
Best practices and pitfalls 18
Module 1 cheat sheet 19
Knowledge check 20
Knowledge check: answers and explanations 22

MODULE 2 **Qualifying Candidate Use Cases** 24
Why a fast, structured screen 25
Normalizing a vague proposal 26
Seven separable dimensions 26
Why data readiness is a gate, not a weight 27
Was the outcome ever recorded? The label test 28
Sizing AI risk by use, not by technique 28
Go/no-go criteria and dispositions 29
Selecting good pilots 30
Case study: Kestrel Retail Group 31
Consulting templates 33
Best practices and pitfalls 35
Module 2 cheat sheet 36
Knowledge check 37
Knowledge check: answers and explanations 39

MODULE 3 **Prioritizing with a Value-Effort Lens** 41
Why a value-effort lens 42
Setting criteria and weights before scoring 42
Defining value consistently 43
Estimating effort by component 43
Pricing the direction of error into value 44
The impact-effort matrix 46
From matrix to shortlist with stated dispositions 46
Trade-off transparency and stakeholder alignment 47
AI investment sequencing 47
Case study: Brightwater Bank 48
Consulting templates 50
Best practices and pitfalls 52

Module 3 cheat sheet 53
Knowledge check 54
Knowledge check: answers and explanations 56

MODULE 4 **Building the Prioritized Opportunity Portfolio 58**
A list of good ideas is not a portfolio 59
The tracked AI intake process 59
Classifying risk exposure before weighing value 60
Two-axis scoring against one fixed set of criteria 61
Balancing the portfolio across value, risk and effort 62
Resolving duplicate automation through consolidation 62
Routing the portfolio into an investment and resource allocation decision 63
“Not this cycle” — a correct answer 64
Case study: Vellmar Group 65
Consulting templates 66
Best practices and pitfalls 69
Module 4 cheat sheet 70
Knowledge check 71
Knowledge check: answers and explanations 73

BACK **Course at a glance: the master cheat sheet 75**
Template index 75
Glossary 77
Mock exam 80
Mock exam: answer key and explanations 86
Further reading and next steps 94

MODULE 1

Finding AI Opportunities

Start from how work actually gets done — not from what the technology can do. Map workflows, find friction, tie every candidate to a concrete source of value, and record who picks up the effort that AI moves rather than removes.

YOU WILL BE ABLE TO

- Explain why opportunity discovery should start from work, not technology novelty
- Map a workflow at task level and analyse its friction
- Locate candidates by business function and industry, including AI assistant and agentic categories
- Connect each candidate to a concrete source of business value
- Build an opportunity inventory from solicited input, observed use and a standing audit
- Estimate value and effort as two separate, arguable judgments
- Name the ledger — who verifies, corrects, escalates or waits — and record relocated effort as a moved cost

IN THIS MODULE

1. Work-first, not technology-first
2. Workflow mapping
3. Process friction analysis
4. Candidates by function and industry
5. Assistants, copilots and agents
6. Connecting candidates to value
7. Building the opportunity inventory
8. Value and effort as separate judgments
9. Naming the ledger
10. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4–5 hours

Most enterprises do not lack AI ideas; they lack *good* ones — ideas grounded in real work, tied to real value and owned by someone who will act. This module teaches a discovery method that starts from how work is done, finds where it hurts, and turns each pain point into a candidate with a value hypothesis, a separate effort judgment and an honest ledger of where effort will move.

KEY TAKEAWAYS

1. Discover from **work-pull**, not **technology-push**: map how work is actually done, then look for friction.
2. **Friction types** (searching, re-keying, waiting, rework, judgment bottlenecks, exceptions…) point to specific AI patterns — and to non-AI fixes.
3. Scan systematically by **function** and **industry**, and classify assistant vs agentic candidates by **level of autonomy**.
4. Every candidate needs a **value hypothesis**: AI does X → metric Y changes → worth Z, because…
5. Build the inventory from three channels: **solicited input**, **observed use** and a **standing audit**.
6. Estimate **value** and **effort** separately, as ranges with stated assumptions someone can argue with.
7. **Name the ledger**: who verifies, corrects, escalates or waits. Relocated effort is a **moved cost**, not a saving.

1.1 Work-first, not technology-first

Technology-push discovery starts with a capability ("agents can now browse and act") and hunts for somewhere to apply it. It produces impressive demonstrations and a high stall rate, because the resulting use cases have no natural owner and no measured pain. Work-pull discovery starts with the work: what people do, how long it takes, where it goes wrong — and only then asks which fix, AI or otherwise, removes the friction.

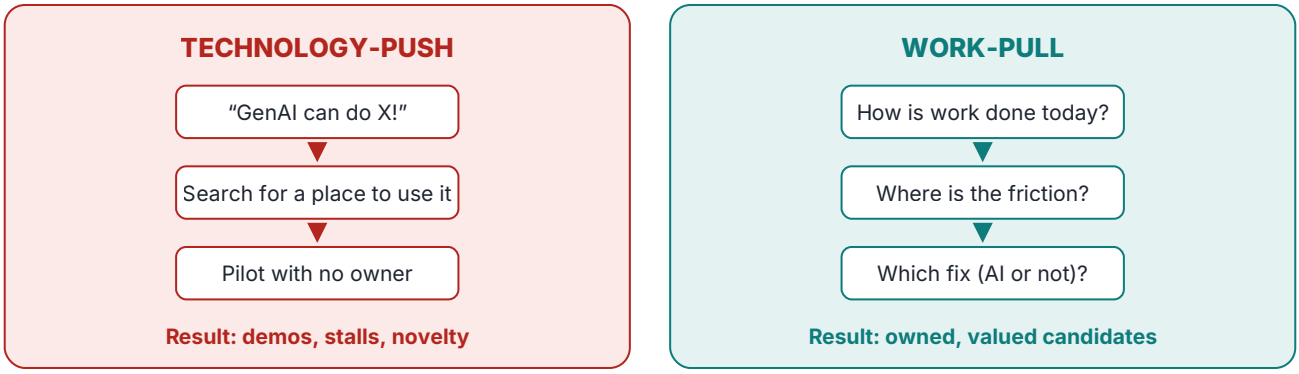


Figure 1.1 — Technology-push vs work-pull discovery

KEY IDEA

A good AI opportunity is a **problem in a workflow** that AI is plausibly the best way to fix. If you cannot point to the step, the people and the measure, you have an idea — not an opportunity.

EXAM TRAP

Exam scenarios often describe a sponsor excited by a new model release. The correct first move is to go to the work (map the process, find the friction) — not to brainstorm applications of the new model.

1.2 Workflow mapping

Workflow mapping makes visible how work actually flows — which is rarely how the procedure manual says it flows. Map at three levels, going only as deep as the opportunity requires:

Level	What you map	Typical tool	Use for
Process	End-to-end flow from trigger to outcome, major steps, systems	SIPOC (suppliers, inputs, process, outputs, customers)	Scoping; choosing where to look closer
Workflow	Steps, roles, hand-offs, decision points, systems, wait states	Swimlane diagram	Locating friction and hand-off costs
Task	What a person does in a step: inputs read, judgments made, outputs produced, time taken	Task inventory / observation notes	Matching friction to AI patterns; sizing

What to capture for each step

Attribute	Why it matters for AI
Volume and frequency	Value scales with volume; low-volume steps rarely justify AI
Handling time and wait time	Separates effort you can save from elapsed time you can shorten
Inputs (structured / unstructured)	Unstructured inputs (text, documents, images, speech) are where modern AI adds most
Judgment involved	Rules-based judgment → automation; pattern judgment → ML; language judgment → GenAI
Error and rework rate	Rework is often the biggest hidden cost
Exceptions and escalations	Shows where autonomy must stop
Systems touched	Drives integration effort
Who acts on the output	Identifies the owner and the ledger (Section 1.9)

> **CONSULTANT'S TIP · Observe, don't only interview**
> People describe the official process; they *perform* the real one. Sit with three to five practitioners for an hour each. The workarounds, sticky notes, spreadsheets and copy-paste between screens you see are the richest source of opportunities.

1.3 Process friction analysis

Friction is any point where work is slower, costlier, riskier or more frustrating than it needs to be. Naming the friction type tells you which AI pattern — if any — fits.

Friction type	Symptom	Candidate AI pattern	Check non-AI fix first
Searching	Staff hunt across systems and documents for information	Retrieval and Q&A over enterprise content	Better filing, single source of truth
Re-keying / extraction	Data copied from documents or e-mails into systems	Document information extraction	E-invoicing, structured web forms

Friction type	Symptom	Candidate AI pattern	Check non-AI fix first
Reading and summarising	Long documents read to find a few facts	Summarisation, classification	Shorter templates at source
Drafting	Repetitive writing of letters, reports, replies	Generative drafting with review	Standard templates
Triage and routing	Items manually sorted to the right queue	Classification and routing	Simple rules if criteria are explicit
Pattern judgment	Experts estimate risk, demand, likelihood	Predictive ML	Clearer policies, checklists
Waiting	Work idle between hand-offs or approvals	Often not AI: workflow redesign; sometimes agentic coordination	Remove approvals, parallelise
Rework	Errors found late and corrected	Validation, anomaly detection at entry	Input validation, training
Exception handling	Non-standard cases consume expert time	Assist experts; AI rarely handles exceptions alone	Reduce exceptions at source
Coordination across systems	Multi-step tasks across tools	Agentic task execution (with controls)	Integration, RPA

PITFALL · Waiting is rarely an AI problem

If a case spends 9 of its 10 days waiting for an approval, making the 30-minute processing step faster with AI changes almost nothing. Elapsed-time friction usually needs process redesign; effort friction is where AI helps most.

1.4 Locating candidates by business function and industry

A systematic scan by function and industry ensures you do not only find the opportunities that the loudest stakeholders bring. Use these as prompts, not a menu — every candidate must still pass through workflow and friction analysis.

Function	Recurring AI opportunity patterns
Customer service	Query triage, agent assist, knowledge retrieval, response drafting, conversation summaries, self-service assistants
Sales & marketing	Lead scoring, account research, proposal drafting, content generation, next-best-action, churn prediction
Finance	Invoice extraction and matching, anomaly detection, close commentary drafting, cash forecasting, audit sampling
Procurement	Contract clause extraction, supplier risk signals, spend classification, RFP drafting and comparison
HR	Job-description drafting, candidate screening support (high-risk!), employee query assistant, learning content
Legal & compliance	Contract review, regulatory change monitoring, policy Q&A, e-discovery triage
IT	Ticket triage and resolution agents, code assistants, incident summarisation, log anomaly detection

Function	Recurring AI opportunity patterns
Operations & supply chain	Demand forecasting, predictive maintenance, quality inspection, route and schedule optimisation
R&D / product	Literature and patent search, experiment design support, code and specification drafting

Industry	Characteristic opportunity areas	Typical constraint
Banking	Fraud, AML alert triage, credit document processing, advisor assist	Model risk management; credit decisions high-risk
Insurance	Claims intake and triage, underwriting assist, fraud, policy Q&A	Fair treatment; explainability
Manufacturing	Predictive maintenance, visual inspection, planning, technical documentation	OT data access; plant variability
Retail & consumer	Demand forecasting, pricing, personalisation, customer service	Data fragmentation across channels
Healthcare	Clinical documentation, coding, scheduling, imaging support	Patient safety; medical-device regulation
Public sector	Case triage, citizen enquiries, document processing	Transparency, fairness, procurement rules
Energy & utilities	Asset maintenance, grid forecasting, billing queries	Critical infrastructure; legacy systems

1.5 AI assistants, copilots and agents

Generative AI has created two categories that are reshaping enterprise work. Classifying candidates by **level of autonomy** — rather than by product label — makes their value, risk and ledger comparable.

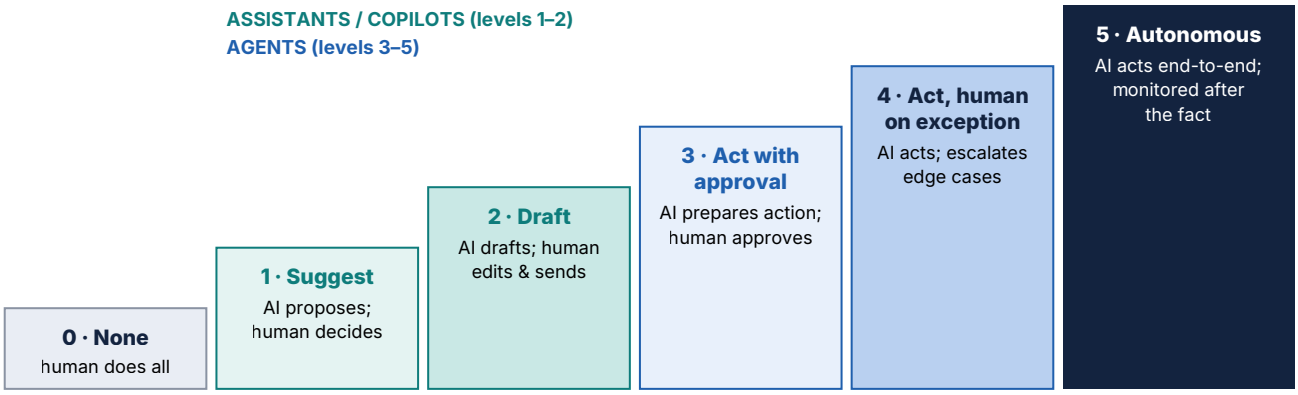


Figure 1.2 — The autonomy ladder

Category	What it does	Where value comes from	What changes in the ledger
AI assistant / copilot	Answers, suggests or drafts inside a person's workflow; the person decides and acts	Faster drafting, searching, summarising; quality uplift	Review time added; responsibility stays with the person
Agentic AI	Plans and executes multi-step tasks across tools, with approval or exception-based oversight	Removes whole tasks or hand-offs; 24/7 throughput	Approvals, exception queues, monitoring and incident handling added

KEY IDEA

Higher autonomy can unlock larger value, but it also moves more effort into verification, exception handling and monitoring — and raises risk. A candidate's autonomy level is a design choice to be argued, not a property of the technology.

1.6 Connecting each candidate to a concrete source of value

"Productivity" is not a value source. Every candidate needs a **value hypothesis** that names the mechanism, the metric and why that metric is worth money or risk reduction to this organization.

AI does **[task change]** → which changes **[business metric]** from **[baseline]** to **[target]** → worth **[€ / risk / strategic value]** because **[mechanism]**

Value lever	Example metric	Converts to value when...
Revenue growth	Conversion rate, basket size, win rate	Extra revenue is not cannibalised and capacity exists to serve it
Cost reduction	Cost per case, FTE need, external spend	Saved capacity is actually removed or redeployed to value
Cost avoidance	Hiring avoided as volume grows	Volume growth is credible
Capacity release	Hours freed per week	Freed time is redirected to named higher-value work
Risk reduction	Losses, fines, incidents	Expected loss (probability × impact) actually falls
Quality	Error rate, rework, complaints	Errors have a cost you can trace
Speed / cycle time	Days to decision, time-to-quote	Speed changes customer behaviour or cash
Working capital	Days sales outstanding, inventory	Cash released has a financing value
Experience	NPS, employee attrition	Linked to retention, revenue or cost

PITFALL · Capacity release is not cash

Saving 20 minutes per person per day across 500 people is a capacity claim. It becomes cost reduction only if headcount, overtime or contractors fall — or value if the time is redirected to work that produces something. State which.

1.7 Building the opportunity inventory

The opportunity inventory is the list of candidates under consideration, each described in a comparable way. Build it from three channels, because each has blind spots the others cover.

Channel	How	Strength	Bias / blind spot
Solicited input	Workshops, idea calls, sponsor and expert interviews	Engages stakeholders; surfaces strategic priorities	Favours the vocal and senior; technology-push ideas; solution-shaped asks

Channel	How	Strength	Bias / blind spot
Observed use	Shadow AI and workarounds, tool usage data, process mining and logs, help-desk tickets	Shows real demand and real friction	Misses needs nobody has tried to solve; privacy sensitivities
Standing audit	A scheduled, systematic review of every function and core process against the friction and pattern lists	Coverage and consistency; repeatable each cycle	Effortful; can become a box-ticking exercise

BEST PRACTICE · Inventory hygiene

- One row per candidate, written in the same structure (Module 2 normalizes it further).
- Record the **channel** each candidate came from — it explains biases later.
- Merge duplicates early but keep the original requesters visible.
- Keep rejected ideas with a reason; they return.

1.8 Estimating value and effort as two separate, arguable judgments

At discovery stage, estimates are rough — and that is fine, provided they are **separate** and **arguable**. Separate, because a single "attractiveness" score lets enthusiasm for value hide the effort, or vice versa. Arguable, because every estimate must state its reasoning so a colleague can challenge a specific assumption rather than a gut feeling.

Gross annual value ≈ volume × time (or cost) per unit × share addressable × expected improvement × value per unit of time

EXAMPLE · An arguable value estimate

Supplier invoices: **180,000/year** (ERP data) × **4 min** manual keying (observed, 12 samples) × **70%** addressable (PDF invoices; paper excluded) × **60%** time reduction (comparable deployments; to be tested) × **€0.75/min** loaded cost ≈ **€227k/year gross capacity** — before the ledger. Range €150–300k. Each number can be challenged on its own.

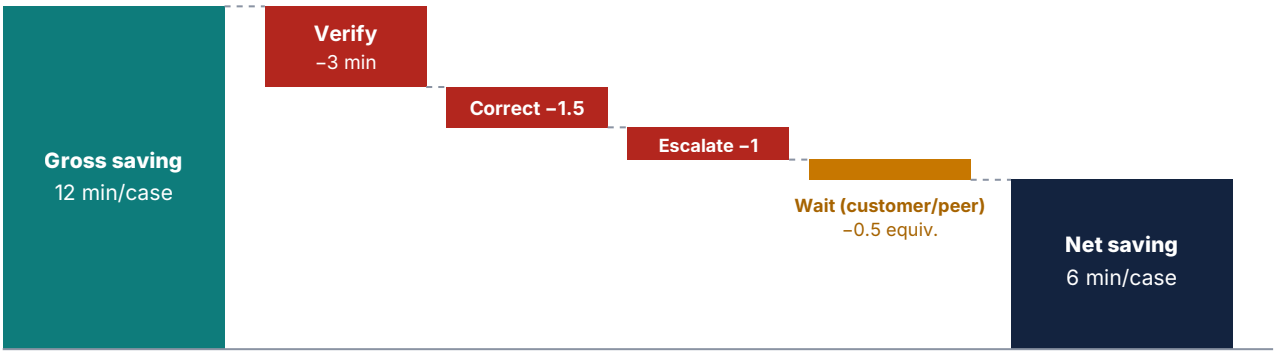
Effort driver (rough, at discovery)	Low	Medium	High
Data	Exists, accessible, permitted	Exists, needs work	Must be created or collected
Integration	Standalone or one system	Two or three systems	Many systems or legacy
Process change	Fits existing step	Changes roles	Redesigns the process
Risk & compliance	Internal, low impact	Customer-facing	Affects rights or safety

NOTE

Module 3 refines effort into components and value into risk-adjusted, error-priced figures. At discovery, the goal is to be roughly right and transparently reasoned.

1.9 Naming the ledger: where the effort goes

AI rarely removes effort cleanly. It moves effort: someone must **verify** outputs, **correct** errors, **escalate** cases the system cannot handle, and someone — often a customer or colleague — **waits** while that happens. If you count only the saving at the step AI touches, you overstate value. The ledger makes the moved effort visible.



The effort did not disappear — part of it moved to reviewers, specialists and customers. The ledger records where.

Figure 1.3 — From gross saving to net saving via the ledger

Ledger line	Question	Example
Verify	Who checks outputs, how often, how long?	Handler reviews each drafted letter: 3 min
Correct	Who fixes errors, at what rate and cost?	15% of drafts need edits of ~10 min
Escalate	Which cases go to whom, how often?	5% go to a specialist: 20 min each
Wait	Who waits for verification, correction or escalation, and what does waiting cost?	Customer waits an extra day for escalated cases
New run work	Who monitors, maintains prompts, handles incidents?	0.5 FTE platform and quality monitoring

Net value = gross value − (verify + correct + escalate + wait costs) − new run costs

KEY IDEA · Moved cost, not saving

If the AI saves 10 minutes for a clerk but adds 4 minutes for a supervisor, record a 10-minute reduction *and* a 4-minute **moved cost** — not a 10-minute saving. Moved costs often land on scarcer, more expensive people, or on customers.

EXAM TRAP

A question may describe a GenAI tool that "saves agents 8 minutes per case" and ask for the most defensible value statement. The right answer accounts for review, correction and escalation effort — the ledger — rather than multiplying the gross saving by volume.

1.10 Case study: Alpenhof Energie

CASE STUDY **From "use GenAI in customer service" to five owned candidates**

Alpenhof Energie (fictional) · regional energy supplier · 1.4m customers

Situation

Alpenhof's board asked for "GenAI in customer service by summer". The advisor proposed a four-week work-first discovery of the billing-enquiry process instead of a technology showcase.

Discovery

- **Workflow mapping** (swimlanes plus 15 hours of observation) showed 420,000 billing enquiries/year across phone, e-mail and web form. E-mail enquiries took 11 minutes on average: 4 minutes finding the account and meter history across three systems, 3 minutes reading attached photos of meters and letters, 4 minutes drafting a reply.
- **Friction:** searching (account lookup), extraction (meter photos), drafting, and waiting — 30% of e-mails waited two days for a meter-reading correction approved by a back-office team.
- **Observed use:** 40 agents used a free online translation tool for Turkish and Polish customers, pasting in customer data.
- **Standing audit** of the billing function surfaced a candidate nobody had raised: predicting which customers would dispute their annual bill, so the bill could be explained before it was sent.

The inventory (excerpt)

Candidate	Autonomy	Value hypothesis	Ledger notes
Account and history retrieval assistant	1	−3 min/e-mail × 250k e-mails	Little verification needed
Meter-photo reading	2	−2 min on 35% of e-mails	Agent confirms reading: +0.5 min
Reply drafting	2	−3 min/e-mail gross	Review 1.5 min; 10% heavy edits; legal wording check
Approved translation service	2	Risk reduction (data leakage) + quality	Replaces shadow tool
Bill-dispute prediction	1	Fewer disputes, less inbound volume	Who calls flagged customers? New outbound work

What changed

The back-office approval wait — the largest elapsed-time problem — was solved by changing approval limits, not AI. The reply-drafting value fell from a gross €690k to an estimated €330k net after the ledger, which re-ordered the list: retrieval and meter-photo reading moved ahead of drafting.

Discussion questions

1. Which candidate came from which discovery channel, and what would have been missed with solicited input alone?
2. Why did the two-day wait not become an AI candidate?
3. Write the ledger lines for the bill-dispute prediction candidate.

4. Why might the translation candidate be prioritized even though its cost saving is small?

Model answers (brief)

1. Drafting and retrieval: solicited/observed; translation: observed use (shadow AI); dispute prediction: standing audit. Solicited input alone would have missed both translation and dispute prediction.
2. It was elapsed-time friction caused by an approval policy; a policy change was cheaper and faster.
3. Verify: who reviews the flagged list; Correct: false flags; Escalate: complex disputes to billing experts; Wait: none for customers, but outbound calls add agent effort; New run: monitoring of model drift each billing cycle.
4. It reduces a live data-leakage risk from unsanctioned use — the value lever is risk reduction, not cost.

1.11 Consulting templates

T1.1 Workflow Friction Map

Purpose: Record a workflow at step level and tag friction. **Use when:** discovery observation and workshops

Process / trigger → outcome: .....

Observed by / date / sample size: .....

#	Step	Role	Volume / yr	Handling time	Wait time	Input type	Friction type	Error / rework %
3	*Find account history*	*Agent*	*250k*	*4 min*	*—*	*3 systems*	*Searching*	*—*

Biggest effort friction: .....

Biggest elapsed-time friction: .....

Non-AI fixes spotted: .....

T1.2 Opportunity Card

Purpose: Describe one candidate with a value hypothesis and ledger.

Candidate name: .....

Workflow step / friction type: .....

Source channel (solicited / observed / audit): .....

Function / industry pattern: .....

Autonomy level (0–5): .....

Value hypothesis — AI does… .....

…which changes metric (baseline → target): .....

…worth (range) because: .....

Effort judgment (L / M / H) and why: .....

Ledger — verify / correct / escalate / wait / new run: .....

Owner who would act on the output: .....

T1.3 Opportunity Inventory Register

Purpose: Keep all candidates comparable in one list.

ID	Candidate	Function	Channel	Friction	Autonomy	Value range (net)	Effort L/ M/H	Owner	Status
OP-07	*Meter-photo reading*	*Billing*	*Observed*	*Extraction*	*2*	*€60–110k*	*M*	*Head of Billing Ops*	*To screen*

T1.4 Value Estimate & Effort Ledger

Purpose: Make value arguable and record moved effort.

GROSS VALUE

Factor	Value	Source / assumption	Low	High
Volume / year				
Time or cost per unit				
Share addressable				
Expected improvement				
Value per unit of time				
Gross value				

LEDGER

Line	Who	Rate	Time / cost per case	Annual moved cost
Verify				
Correct				
Escalate				
Wait				
New run work				
Net value				

1.12 Best practices and pitfalls

✓ DO THIS

- Start every discovery with the work, not the model.
- Observe practitioners; capture workarounds.
- Separate effort friction from elapsed-time friction.
- Write a value hypothesis with a mechanism for every candidate.
- Use all three inventory channels and record the source.
- Estimate value and effort separately, as ranges with assumptions.
- Complete the ledger before quoting any saving.

× AVOID THIS

- Brainstorming uses for the latest model.
- Mapping only the official procedure.
- Applying AI to waiting problems.
- Quoting “productivity” as the value.
- Relying on idea workshops alone.
- One blended “attractiveness” score.
- Counting moved effort as savings.

1.13 Module 1 cheat sheet

MODULE 1 · ONE-PAGE SUMMARY

Finding AI Opportunities

WORK-PULL

Work → friction → fix (AI or not). Not: technology → search for a use. An opportunity = a problem in a workflow AI plausibly fixes best.

WORKFLOW MAPPING

Process (SIPOC) → workflow (swimlane) → task. Capture volume, handling vs wait time, inputs, judgment, errors, exceptions, systems, who acts.

FRICTION → PATTERN

- Searching → retrieval/Q&A
- Re-keying → extraction
- Drafting → GenAI + review
- Triage → classification
- Pattern judgment → ML
- Waiting → usually process redesign

AUTONOMY LADDER

0 none · 1 suggest · 2 draft · 3 act with approval · 4 act, human on exception · 5 autonomous. Assistants 1–2; agents 3–5. More autonomy = more value potential, more risk, different ledger.

VALUE HYPOTHESIS

AI does [task change] → changes [metric] from [baseline] to [target] → worth [value] because [mechanism]. Levers: revenue · cost reduction · cost avoidance · capacity · risk · quality · speed · working capital · experience. Capacity ≠ cash.

INVENTORY CHANNELS

- **Solicited**: workshops, interviews (bias: vocal, tech-push)
- **Observed**: shadow AI, workarounds, logs
- **Standing audit**: systematic, scheduled

SEPARATE & ARGUABLE

Value = volume × unit time × addressable × improvement × €/time (range). Effort: data, integration, process change, risk. Never blended.

THE LEDGER

Verify · correct · escalate · wait · new run work. Net value = gross − moved costs − run costs. Relocated effort is a **moved cost**, not a saving.

1.14 Knowledge check

Answer all questions before turning the page. Explanations for every option follow.

K1 **A sponsor returns from a vendor event asking where the company can use "autonomous agents". What is the best first step?**

- (A) Pilot the vendor's agent on any available process.
- (B) Map priority workflows and identify friction, then assess whether agents fit any of it.
- (C) Run a brainstorm on possible agent use cases.
- (D) Decline, because agents are immature.

K2 **Cases take 10 days end-to-end, of which 9 are spent waiting for an approval; processing takes 30 minutes. What is the most likely best fix for cycle time?**

- (A) Process redesign of the approval step
- (B) A GenAI assistant to speed up processing
- (C) Document extraction on case files
- (D) A predictive model of approval outcomes

K3 **Which statement is a well-formed value hypothesis?**

- (A) "We will deploy the leading document-AI platform."
- (B) "AI will make the finance team more productive."
- (C) "AI is strategic for our future."
- (D) "AI extraction cuts invoice keying from 4 to 1.5 minutes on 70% of 180k invoices, releasing ~5,000 hours, which avoids two planned hires."

K4 **Which discovery channel is most likely to reveal staff pasting customer data into an unapproved translation tool?**

- (A) The annual budget process
- (B) Solicited input via idea workshops
- (C) Observed use
- (D) A vendor briefing

K5 **Why should value and effort be estimated as two separate judgments?**

- (A) Because regulators require it.
- (B) Because effort cannot be estimated at discovery stage.
- (C) Because value is always more important than effort.
- (D) Because a blended score lets enthusiasm for one hide the other, and separate judgments can each be argued.

K6 **A drafting assistant saves agents 8 minutes per case. Supervisors now spend 3 minutes checking each draft, and 10% of cases go to specialists for 15 minutes. How should this be recorded?**

- (A) As an 8-minute saving per case
- (B) As no saving, because effort moved
- (C) As an 8-minute reduction with moved costs of 3 minutes plus 1.5 minutes per case, giving about 3.5 minutes net
- (D) As an 11-minute saving because quality improves

K7 **An assistant prepares a refund and a human approves it before it is issued. Which autonomy level is this?**

- (A) Level 2 — draft
- (B) Level 3 — act with approval
- (C) Level 1 — suggest
- (D) Level 5 — autonomous

K8 **A team claims a tool "frees 20 minutes per employee per day for 500 staff". What is the most important follow-up?**

- (A) Ask whether the freed time will reduce cost or be redirected to named higher-value work.
- (B) Accept it, since time savings are always valuable.
- (C) Multiply by salary cost to get the saving.
- (D) Ask which model the tool uses.

1.15 Knowledge check: answers and explanations

K1 **Answer: B** Review: 1.1

A sponsor returns from a vendor event asking where the company can use "autonomous agents". What is the best first step?

- ✗ (A) Pilot the vendor's agent on any available process. — **Wrong.** A pilot without a defined problem is likely to stall.
- ✓ (B) Map priority workflows and identify friction, then assess whether agents fit any of it. — **Correct.** Work-pull discovery starts from how work is done and only then matches technology.
- ✗ (C) Run a brainstorm on possible agent use cases. — **Wrong.** This is technology-push discovery; it yields ideas without owners or measured pain.
- ✗ (D) Decline, because agents are immature. — **Wrong.** This dismisses the request without investigating real opportunities.

K2 **Answer: A** Review: 1.3

Cases take 10 days end-to-end, of which 9 are spent waiting for an approval; processing takes 30 minutes. What is the most likely best fix for cycle time?

- ✓ (A) Process redesign of the approval step — **Correct.** The friction is elapsed-time waiting, which usually calls for process or policy change.
- ✗ (B) A GenAI assistant to speed up processing — **Wrong.** Processing is a small fraction of elapsed time; speeding it barely changes the cycle.
- ✗ (C) Document extraction on case files — **Wrong.** Extraction addresses re-keying, not waiting.
- ✗ (D) A predictive model of approval outcomes — **Wrong.** Prediction does not remove the wait unless the approval itself changes.

K3 **Answer: D** Review: 1.6

Which statement is a well-formed value hypothesis?

- ✗ (A) "We will deploy the leading document-AI platform." — **Wrong.** This names a product, not value.
- ✗ (B) "AI will make the finance team more productive." — **Wrong.** No mechanism, metric, baseline or value.
- ✗ (C) "AI is strategic for our future." — **Wrong.** This is not testable or concrete.
- ✓ (D) "AI extraction cuts invoice keying from 4 to 1.5 minutes on 70% of 180k invoices, releasing ~5,000 hours, which avoids two planned hires." — **Correct.** It names the task change, metric, baseline, target and why it is worth money.

K4 **Answer: C** Review: 1.7

Which discovery channel is most likely to reveal staff pasting customer data into an unapproved translation tool?

- ✗ (A) The annual budget process — **Wrong.** Budgets do not show free tools.
- ✗ (B) Solicited input via idea workshops — **Wrong.** Staff rarely volunteer unsanctioned use in workshops.
- ✓ (C) Observed use — **Correct.** Observing real work, usage data and workarounds surfaces shadow AI.
- ✗ (D) A vendor briefing — **Wrong.** Vendors do not know internal usage.

K5 **Answer: D** Review: 1.8

Why should value and effort be estimated as two separate judgments?

- ✗ (A) Because regulators require it. — **Wrong.** No regulation requires this.
- ✗ (B) Because effort cannot be estimated at discovery stage. — **Wrong.** Rough effort can and should be estimated.
- ✗ (C) Because value is always more important than effort. — **Wrong.** Both matter; separation does not imply ranking.
- ✓ (D) Because a blended score lets enthusiasm for one hide the other, and separate judgments can each be argued. — **Correct.** Separation keeps both visible and challengeable.

K6 **Answer: C** Review: 1.9

A drafting assistant saves agents 8 minutes per case. Supervisors now spend 3 minutes checking each draft, and 10% of cases go to specialists for 15 minutes. How should this be recorded?

✗ (A) As an 8-minute saving per case — **Wrong.** This ignores the effort moved to supervisors and specialists.

✗ (B) As no saving, because effort moved — **Wrong.** Some effort is genuinely saved; the net figure is positive.

✓ (C) As an 8-minute reduction with moved costs of 3 minutes plus 1.5 minutes per case, giving about 3.5 minutes net — **Correct.** The ledger records relocated effort as moved cost: 8 − 3 − (0.1 × 15) = 3.5.

✗ (D) As an 11-minute saving because quality improves — **Wrong.** Quality effects must be evidenced separately, not added to time.

K7 **Answer: B** Review: 1.5

An assistant prepares a refund and a human approves it before it is issued. Which autonomy level is this?

✗ (A) Level 2 — draft — **Wrong.** Drafting produces content for a human to send; here an action is prepared.

✓ (B) Level 3 — act with approval — **Correct.** The system prepares an action that executes only after human approval.

✗ (C) Level 1 — suggest — **Wrong.** It does more than suggest; it prepares the action.

✗ (D) Level 5 — autonomous — **Wrong.** A human approves each action, so it is not autonomous.

K8 **Answer: A** Review: 1.6

A team claims a tool "frees 20 minutes per employee per day for 500 staff". What is the most important follow-up?

✓ (A) Ask whether the freed time will reduce cost or be redirected to named higher-value work. — **Correct.** Capacity release becomes value only when it is removed or redeployed.

✗ (B) Accept it, since time savings are always valuable. — **Wrong.** Unredirected time produces no measurable value.

✗ (C) Multiply by salary cost to get the saving. — **Wrong.** This treats capacity as cash without checking what happens to the time.

✗ (D) Ask which model the tool uses. — **Wrong.** The model choice does not determine value.

MODULE 2

Qualifying Candidate Use Cases

A fast, structured screen that separates use cases worth deeper study from attractive-sounding ones that will stall — normalizing vague proposals, judging separable dimensions, gating on data and sizing risk by use.

YOU WILL BE ABLE TO

- Explain the purpose and design of a fast qualification screen
- Normalize a vague proposal into a defined, comparable use case
- Judge a use case on seven separable dimensions instead of one blended impression
- Explain why data readiness works as a gate rather than a weight
- Test whether the outcome a model would predict was ever recorded
- Size AI risk by what a system is used for rather than by its technique
- Apply go/no-go criteria and select good first pilots

IN THIS MODULE

1. Why a fast screen
2. Normalizing a proposal
3. Seven separable dimensions
4. Data readiness as a gate
5. Was the outcome ever recorded?
6. Sizing risk by use, not technique
7. Go/no-go criteria and dispositions
8. Selecting pilots
9. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4 hours

An opportunity inventory always contains more attractive ideas than an organization can study. Qualification is the fast, structured screen — hours per use case, not weeks — that decides which candidates earn deeper analysis. Done well, it kills the use cases that would stall *before* they consume a pilot budget, and it does so with reasons a sponsor can accept.

KEY TAKEAWAYS

1. Qualification is a **fast, cheap, structured screen** — not a business case.
2. First **normalize** each proposal into a defined, comparable object: user, trigger, input, AI task, output, decision fed, human role, volume, metric.
3. Judge **seven separable dimensions**: business value, workflow fit, data readiness, technical feasibility, risk, sponsorship, measurement clarity.
4. Treat **data readiness as a gate**, not a weight: high value must not be allowed to mask missing data.
5. For prediction use cases, test whether the **outcome was ever recorded** — completely, consistently, in time, and without selection bias.
6. Size **risk by use** (who is affected, consequence, reversibility, autonomy), not by technique.
7. End every screen with a **disposition and a reason**: advance, advance with conditions, park, redirect or reject.

2.1 Why a fast, structured screen

Without a screen, organizations either study everything superficially or study a few favourites deeply — usually the ones with the most senior sponsor. A good screen is **fast** (a few hours per candidate), **structured** (the same questions for every candidate), **separable** (dimensions judged independently) and **decisive** (every candidate leaves with a disposition).

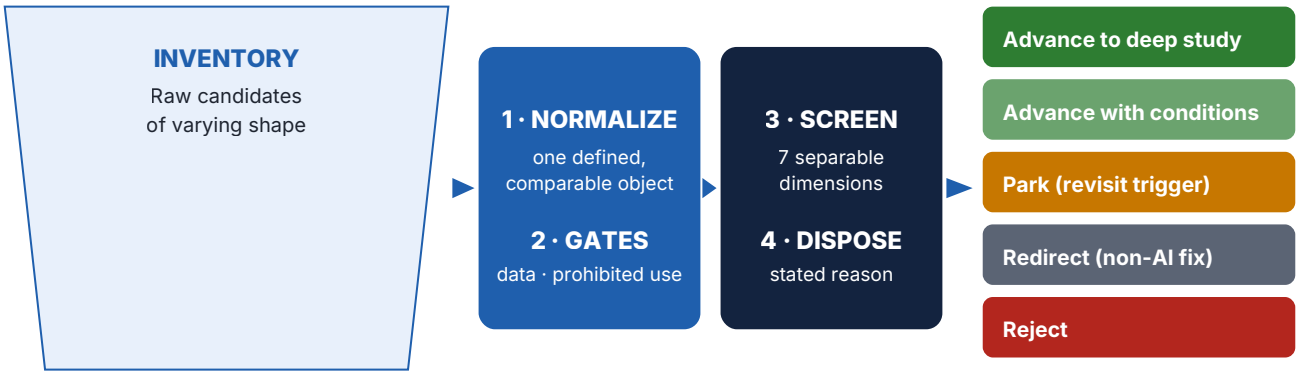


Figure 2.1 — The qualification funnel

KEY IDEA

Qualification asks: **"Is this worth deeper study?"** — not "Is this worth funding?" Funding decisions come later, after prioritization (Module 3) and portfolio construction (Module 4).

2.2 Normalizing a vague proposal

Proposals arrive in every shape: “AI for procurement”, “a chatbot for HR”, “use agents to speed up onboarding”. They cannot be compared until each is rewritten in the same structure. Normalization often reveals that one proposal is really three use cases — or that it is not an AI use case at all.

Field	Question	Example — “AI for procurement” normalized
User / actor	Who uses the output?	Category buyers (35 people)
Trigger	What starts the task?	Receipt of a supplier contract for renewal
Input	What does the AI receive?	Contract PDF and the company’s clause playbook
AI task	What does the AI do (one verb)?	Extract and compare clauses against the playbook
Output	What does it produce?	Deviation list with clause references
Decision or action fed	What is decided or done with the output?	Buyer decides which clauses to renegotiate
Human role / autonomy	Who checks or acts, at which level?	Level 1–2: buyer reviews every deviation
Volume / frequency	How often?	≈1,200 contracts per year
Success metric	How will we know it worked?	Review time per contract; deviations missed (sampled)
Process location	Where in which workflow?	Contract renewal, step 3 (review)

CONSULTANT'S TIP · The one-verb rule

If the AI task needs “and” to describe it (“extract *and* negotiate *and* approve”), split the proposal into separate use cases. Each part has different data, risk and feasibility.

2.3 Seven separable dimensions

Judging a use case as a single impression (“this one feels strong”) lets one vivid feature dominate. Judge each dimension separately, on its own evidence, using simple anchors.

Dimension	Core question	Strong signal	Red flag
Business value	Is there a credible net value hypothesis (incl. ledger)?	Measured baseline; clear mechanism; material size	“Productivity”; no baseline; value only gross
Workflow fit	Does the output land where someone acts on it?	Fits an existing step; owner wants it	Output has no consumer; requires new behaviour nobody owns
Data readiness	Is the needed data available, permitted, sufficient, representative, owned? (gate)	Data exists and has been sampled	“We must have that somewhere”
Technical feasibility	Can current AI do this task to the needed standard?	Comparable tasks in production; quick test passed	Requires reasoning or accuracy beyond evidence
Risk	What harm could use cause, to whom? (sized by use)	Internal, reversible, human-reviewed	Affects rights, safety, money; autonomous; irreversible

Dimension	Core question	Strong signal	Red flag
Sponsorship	Is there a sponsor with budget and an owner who will run it?	Named sponsor and process owner	Only an innovation team is interested
Measurement clarity	Can success be measured soon, against a baseline?	Metric exists, baseline recorded	Benefit only visible years later or unmeasurable

EXAM TRAP

Questions may present a use case with outstanding value and enthusiastic sponsorship but no clear way to measure success. The correct reading is that it is **not yet qualified** (or qualified only with a condition to establish measurement) — strengths on other dimensions do not compensate for an unmeasurable outcome.

2.4 Why data readiness is a gate, not a weight

In a weighted scoring model, every dimension can compensate for every other. That is reasonable for value and feasibility, but not for data: a use case with superb value and no usable data is not a weaker use case — it is a use case that **cannot happen** until the data problem is solved. Treating data as a weight lets the most attractive ideas float to the top of the list while being impossible to build.

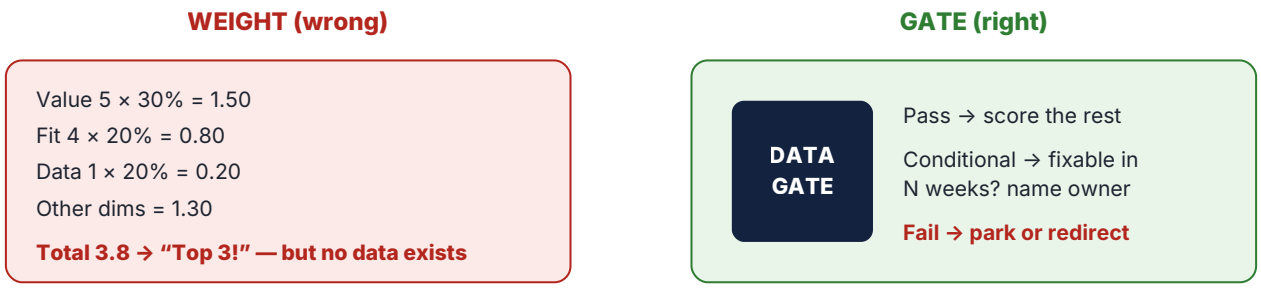


Figure 2.2 — Data readiness as weight vs gate

Gate test	Pass when…
Exists	The data needed is actually captured somewhere today (or reference content exists for GenAI)
Accessible	It can be extracted at the needed frequency and latency
Permitted	There is a legal, contractual and policy basis to use it for this purpose
Sufficient	There is enough volume and history for the task
Representative	It reflects the cases the system will meet in production
Owned	Someone is accountable for it

Gate outcomes: Pass score the other dimensions · Conditional fixable within a defined time by a named owner — advance with that condition · Fail park with a trigger (e.g., "when the new CRM captures X") or redirect.

NOTE

Other dimensions can also act as gates where the consequence is absolute — for example, a use that would be **prohibited** or clearly unacceptable under law or policy is rejected regardless of score.

2.5 Was the outcome ever recorded? The label test

Predictive use cases learn from historical examples of the outcome they predict. If that outcome was never recorded — or was recorded only for some cases — there is nothing to learn from. This is the most common silent killer of attractive ML proposals.

Step	Question	Example failure
1 · Name the target	Exactly what outcome would the model predict?	"Good candidates" — never defined
2 · Find where it is stored	Which system records it, from when?	Reason for customer churn exists only in agents' heads
3 · Check coverage	Is it recorded for all cases, or only some?	Fraud labelled only for claims that were investigated
4 · Check consistency	Was it recorded the same way over time and across units?	Defect codes changed in 2023; plants use different codes
5 · Check timing	Would the inputs be known at prediction time? (no leakage)	"Days to payment" uses a field filled in after payment
6 · Check selection bias	Do recorded outcomes reflect past decisions rather than reality?	Credit defaults known only for applicants who were approved

EXAMPLE · The investigated-claims trap

An insurer wants to predict fraud. Fraud labels exist only for the 3% of claims investigated, selected by today's rules. A model trained on them learns to reproduce today's selection, not to find fraud the rules miss. The outcome was recorded, but not for the population that matters.

CONSULTANT'S TIP · For generative use cases

The equivalent question is: **"Is 'good' recorded anywhere?"** Are there approved example answers, accepted letters, resolved tickets or expert-reviewed summaries that can serve as reference and evaluation material? If quality has never been written down, measurement clarity and feasibility are both at risk.

2.6 Sizing AI risk by use, not by technique

The same model can be harmless in one use and high-risk in another. A large language model drafting internal meeting summaries and the same model drafting rejection letters for credit applicants share a technique; they do not share a risk. Risk follows the **intended purpose and context of use** — the principle underlying the EU AI Act's risk-based approach, where classification depends on the use (for example, employment, creditworthiness, access to essential services) rather than on the algorithm.

Risk factor	Lower risk	Higher risk
Who is affected	Internal staff, aggregated data	Customers, applicants, citizens, vulnerable people
Consequence of error	Minor inconvenience	Money, rights, health, safety, legal position
Reversibility	Easily corrected	Hard or impossible to undo
Autonomy	Suggests to an expert	Acts or decides without review

Risk factor	Lower risk	Higher risk
Scale	Few cases	Many people, systematically
Regulated domain	None	Employment, credit, insurance pricing, health, education, public services
Transparency to affected people	Clearly disclosed	Hidden or unexpected

Screen tier	Typical uses	Screen consequence
Low	Internal drafting, search, summarisation with human use	Proceed with standard controls
Moderate	Customer-facing content with review; operational predictions	Proceed; define oversight and testing in deep study
High	Decisions about people's access to jobs, credit, services; safety-relevant uses	Proceed only with specialist (legal/risk) involvement; higher effort and evidence bar
Escalate / stop	Potentially prohibited or clearly unacceptable uses	Refer to legal immediately; do not advance

EXAM TRAP

"It's just a GenAI drafting tool, so it's low risk" is a classic wrong answer. Ask what the draft is **used for**. If it becomes a decision letter affecting a person's rights, the use is high-risk even though the technique is the same as for meeting notes.

2.7 Go/no-go criteria and dispositions

Agree the go/no-go criteria before screening, apply them to every candidate, and record the disposition with a reason. The point is a decision the sponsor can understand and accept.

Disposition	When	What you record
Advance to deep study	Gates passed; no dimension weak enough to stall it	Questions for deep study; owner
Advance with conditions	Promising but one fixable issue (e.g., data access, measurement baseline)	The condition, its owner and deadline
Park	Blocked by something outside the team's control now	Revisit trigger (event or date)
Redirect	Real problem, but a non-AI fix is better or the use case belongs elsewhere	Where it was sent
Reject	Low value, unacceptable risk, or infeasible	Reason — so it is not re-proposed unchanged

BEST PRACTICE · Kill criteria

Publish explicit kill criteria, for example: no data path within 6 months · no owner willing to run it · net value below €50k/year · use would be prohibited · success unmeasurable within 12 months. Kill criteria protect the team from political pressure to advance favourites.

2.8 Selecting good pilots

Among advanced candidates, the first pilots deserve special care: they teach the organization how AI delivery works and shape its confidence.

Pilot criterion	Why
Meaningful but bounded value	Visible enough to matter; small enough to fail safely
Data passes the gate	Avoids a pilot that is really a data project
Low-to-moderate use risk	Lets the organization learn controls before high-risk uses
Engaged sponsor and production owner	Prevents pilot purgatory
Clear, short-cycle measurement	Results within weeks, against a baseline
Reversible	Easy to switch off
Reusable learning	Builds patterns, platforms or skills other use cases need

PITFALL · The flagship-first trap

Choosing the most ambitious, high-risk, high-visibility use case as the first pilot maximises the chance of a public stall. Start where learning is cheap and success measurable, then scale ambition.

2.9 Case study: Kestrel Retail Group

CASE STUDY **Twelve proposals, one afternoon**

Kestrel Retail Group (fictional) · fashion and home retailer · 260 stores

Situation

After an internal idea campaign, Kestrel has 12 AI proposals and budget to study four in depth. The advisor runs a structured screen with the leadership team.

Selected results

Proposal (normalized)	Key findings	Disposition
Predict which returns are fraudulent	Label test: fraud only recorded for returns stores chose to check (selection bias); data gate conditional	Park — trigger: 6 months of random-sample return audits
Draft product descriptions from supplier specs	Data passes; low risk; measurable (time per listing, conversion); merchandising owner keen	Advance — pilot candidate
Rank job applicants for store roles	Use is high-risk (employment); feasibility unproven; works council not consulted	Advance with conditions: legal/risk review and works-council engagement first
Store-level demand forecast for perishables	Data passes; value high; owner exists; measurement clear	Advance
“AI to improve customer experience”	Could not be normalized: no user, trigger or task	Return to proposer for normalization
Chatbot for store staff HR questions	Value small; HR policy pages out of date (no reference content)	Redirect — fix the policy pages first
Agent to reorder stock autonomously	Overlaps the forecast; autonomy level 5 premature	Merge into forecast as later phase

Outcome

Four candidates advanced; the leadership team accepted the parked and redirected items because each had a stated reason and trigger. The CEO’s favourite — return fraud — was parked, but with a concrete path back.

Discussion questions

1. Why is the return-fraud proposal parked rather than rejected?
2. Explain why the applicant-ranking proposal is high-risk even though the technique is ordinary ML.
3. Which proposal would you choose as the first pilot, and why?
4. What did normalization reveal about the “customer experience” and “reorder agent” proposals?

Model answers (brief)

1. The value is real; the blocker is a fixable data problem (selection bias). Random audits create unbiased labels — a clear revisit trigger.

2. Risk is sized by use: ranking people for employment affects their access to work, a regulated, high-consequence use.
3. Product-description drafting: bounded value, data passes, low risk, clear measurement, keen owner, reversible.
4. The first was not a use case yet; the second was a higher-autonomy extension of an existing candidate, so it was merged rather than screened separately.

2.10 Consulting templates

T2.1 Use-Case Normalization Sheet

Purpose: Rewrite any proposal as a defined, comparable object. **Use when:** before any screening

Original proposal (verbatim) / proposer: .....

User / actor: .....

Trigger: .....

Input: .....

AI task (one verb): .....

Output: .....

Decision or action fed: .....

Human role / autonomy level: .....

Volume / frequency: .....

Success metric and baseline: .....

Process location: .....

Split into further use cases? ☐ no ☐ yes → .....

T2.2 Qualification Screen Scorecard

Purpose: Judge seven dimensions separately, with evidence.

Use case: .....

Data gate result (Pass / Conditional / Fail): .....

Prohibited-use check (clear / escalate): .....

Dimension	Rating (Strong / Adequate / Weak)	Evidence	Condition or question for deep study
Business value (net)			
Workflow fit			
Technical feasibility			
Risk (by use)			
Sponsorship & owner			
Measurement clarity			

Disposition (advance / conditions / park / redirect / reject): .....

Reason: .....

Condition owner & deadline, or revisit trigger: .....

T2.3 Data Readiness Gate

Purpose: Apply the gate before scoring other dimensions.

Test	Finding	Pass / Cond. / Fail	Fix, owner, by when
Exists			
Accessible			
Permitted (legal, contract, policy)			
Sufficient (volume, history)			
Representative of production			
Owned			

Gate result: .....

If conditional — fixable within (weeks): .....

T2.4 Outcome-Recorded (Label) Test

Purpose: Confirm a predictive use case has something to learn from.

Target outcome, precisely defined: .....

System and field where recorded / since when: .....

- ☐ Coverage — recorded for all relevant cases (not only some)
- ☐ Consistency — same definition over time and across units
- ☐ Timing — all inputs available at prediction time (no leakage)
- ☐ Selection bias — outcomes not only known for cases past decisions selected
- ☐ GenAI equivalent — reference examples of “good” exist

Result and implication: .....

T2.5 Use-Based Risk Screen

Purpose: Size risk by what the system is used for.

Intended use and decision affected: .....

Factor	Lower / Higher	Notes
Who is affected		
Consequence of error		
Reversibility		
Autonomy level		
Scale		
Regulated domain		
Transparency to affected people		

Screen tier (Low / Moderate / High / Escalate): .....

Specialists to involve: .....

2.11 Best practices and pitfalls

✓ DO THIS

- Normalize every proposal before comparing it.
- Judge each dimension separately, with evidence.
- Apply the data gate before scoring.
- Run the label test on every predictive proposal.
- Size risk by intended use and context.
- Publish go/no-go and kill criteria in advance.
- Give every candidate a disposition and reason.

× AVOID THIS

- Comparing proposals of different shapes and sizes.
- One blended “feels strong” impression.
- Letting high value compensate for missing data.
- Assuming the outcome “must be in the system”.
- Calling a use low-risk because the technique is common.
- Advancing the most senior sponsor’s favourites.
- Choosing the flagship as the first pilot.

2.12 Module 2 cheat sheet

MODULE 2 · ONE-PAGE SUMMARY

Qualifying Candidate Use Cases

SCREEN DESIGN

Fast · structured · separable · decisive. Question: “worth deeper study?” — not “worth funding?”

NORMALIZE

User · trigger · input · AI task (one verb) · output · decision fed · human role/autonomy · volume · metric · process location.

SEVEN SEPARABLE DIMENSIONS

Business value (net) · **workflow fit** · **data readiness** (gate) · **technical feasibility** · **risk** (by use) · **sponsorship** (sponsor + owner) · **measurement clarity**.

DATA GATE

Exists · accessible · permitted · sufficient · representative · owned. Pass / conditional (owner, deadline) / fail (park or redirect). Value never compensates.

LABEL TEST

1. Name the target
2. Where stored?
3. Coverage
4. Consistency
5. Timing / leakage
6. Selection bias

GenAI: is “good” recorded?

RISK BY USE

Who affected · consequence · reversibility · autonomy · scale · regulated domain · transparency. Same model, different use, different risk.

DISPOSITIONS

Advance · advance with conditions · park (trigger) · redirect (non-AI) · reject (reason). Publish kill criteria.

GOOD FIRST PILOT

Bounded but meaningful value · data passes gate · low–moderate risk · sponsor + production owner · short-cycle measurement · reversible · reusable learning.

2.13 Knowledge check

Answer all questions before turning the page.

K1 **What is the main purpose of a qualification screen?**

- A To produce a detailed business case for each idea
- B To decide quickly and consistently which candidates are worth deeper study
- C To decide which use cases to fund
- D To select vendors

K2 **A proposal reads "use AI to extract, negotiate and approve supplier contracts". What does normalization suggest?**

- A Send it straight to a vendor for a demo.
- B Score it as one use case to save time.
- C Split it into separate use cases, one AI task each.
- D Reject it as too ambitious.

K3 **A use case scores very high on value and sponsorship, but the required data is not captured anywhere. How should a well-designed screen treat it?**

- A Apply the data gate: it fails or is conditional, so it is parked or advanced only with a defined data fix.
- B Lower its value score.
- C Rank it highly, since value outweighs data.
- D Pilot it to find out.

K4 **A bank wants to predict loan default for all applicants. Defaults are recorded only for applicants who were approved. Which label-test issue is this?**

- A Inconsistency
- B Insufficient volume
- C Leakage
- D Selection bias

K5 **A model predicting late payment uses a field that is filled in only after payment occurs. What is the problem?**

- A Timing / leakage: the input is not available at prediction time
- B Selection bias
- C The outcome was never recorded
- D Risk is too high

K6 **The same LLM is used to summarise internal meetings and to draft decisions on social-benefit applications. How should risk be sized?**

- A Neither is risky because a human reviews outputs
- B Higher for meetings, because volume is larger
- C Higher for the benefit decisions, because they affect people's access to essential services
- D Identically, because the technique is the same

K7 **Which candidate makes the best first pilot?**

- (A) An autonomous agent approving refunds for all customers
- (B) An applicant-ranking model for recruitment
- (C) An idea with high value but no data
- (D) Drafting product descriptions with merchandiser review: data available, metric clear, owner keen

K8 **A proposal has strong value and sponsorship but no agreed way to measure success within a year. What is the right disposition?**

- (A) Increase its value score to compensate.
- (B) Advance with a condition to define a metric and baseline, or park until one exists.
- (C) Advance unconditionally — measurement can be sorted out later.
- (D) Reject permanently.

2.14 Knowledge check: answers and explanations

K1 **Answer: B** Review: 2.1

What is the main purpose of a qualification screen?

- ✗ (A) To produce a detailed business case for each idea — **Wrong.** A screen is deliberately faster and lighter than a business case.
- ✓ (B) To decide quickly and consistently which candidates are worth deeper study — **Correct.** Qualification separates candidates that merit analysis from those that would stall.
- ✗ (C) To decide which use cases to fund — **Wrong.** Funding comes after prioritization and portfolio construction.
- ✗ (D) To select vendors — **Wrong.** Vendor selection is a later activity.

K2 **Answer: C** Review: 2.2

A proposal reads "use AI to extract, negotiate and approve supplier contracts". What does normalization suggest?

- ✗ (A) Send it straight to a vendor for a demo. — **Wrong.** Normalization comes before any external engagement.
- ✗ (B) Score it as one use case to save time. — **Wrong.** The parts have different data, risk and feasibility; one score would mislead.
- ✓ (C) Split it into separate use cases, one AI task each. — **Correct.** The one-verb rule: multiple verbs signal multiple use cases.
- ✗ (D) Reject it as too ambitious. — **Wrong.** Some parts may be viable; splitting reveals which.

K3 **Answer: A** Review: 2.4

A use case scores very high on value and sponsorship, but the required data is not captured anywhere. How should a well-designed screen treat it?

- ✓ (A) Apply the data gate: it fails or is conditional, so it is parked or advanced only with a defined data fix. — **Correct.** Data readiness is a gate; value cannot compensate for absent data.
- ✗ (B) Lower its value score. — **Wrong.** Value is real; the problem is data, which should be handled as a gate.
- ✗ (C) Rank it highly, since value outweighs data. — **Wrong.** This is the weighting error the gate prevents.
- ✗ (D) Pilot it to find out. — **Wrong.** A pilot without data cannot produce evidence.

K4 **Answer: D** Review: 2.5

A bank wants to predict loan default for all applicants. Defaults are recorded only for applicants who were approved. Which label-test issue is this?

- ✗ (A) Inconsistency — **Wrong.** Nothing suggests definitions changed.
- ✗ (B) Insufficient volume — **Wrong.** Volume is not the stated problem.
- ✗ (C) Leakage — **Wrong.** Leakage is about using information unavailable at prediction time.
- ✓ (D) Selection bias — **Correct.** Outcomes are known only for cases past decisions selected, so the model cannot learn about rejected applicants.

K5 **Answer: A** Review: 2.5

A model predicting late payment uses a field that is filled in only after payment occurs. What is the problem?

- ✓ (A) Timing / leakage: the input is not available at prediction time — **Correct.** The model would look accurate in testing but fail in use.
- ✗ (B) Selection bias — **Wrong.** This is not about which cases have labels.
- ✗ (C) The outcome was never recorded — **Wrong.** The outcome is recorded; the input is the issue.
- ✗ (D) Risk is too high — **Wrong.** This is a data validity problem, not a use-risk issue.

K6 **Answer: C** Review: 2.6

The same LLM is used to summarise internal meetings and to draft decisions on social-benefit applications. How should risk be sized?

✗ (A) Neither is risky because a human reviews outputs — **Wrong.** Human review mitigates but does not remove use-based risk.

✗ (B) Higher for meetings, because volume is larger — **Wrong.** Scale matters, but consequence to people dominates here.

✓ (C) Higher for the benefit decisions, because they affect people's access to essential services — **Correct.** Use-based sizing reflects consequence, who is affected and regulated domain.

✗ (D) Identically, because the technique is the same — **Wrong.** Risk follows use, not technique.

K7 **Answer: D** Review: 2.8

Which candidate makes the best first pilot?

✗ (A) An autonomous agent approving refunds for all customers — **Wrong.** High autonomy and customer impact make it a poor first pilot.

✗ (B) An applicant-ranking model for recruitment — **Wrong.** This is a high-risk use; not a learning pilot.

✗ (C) An idea with high value but no data — **Wrong.** It fails the data gate.

✓ (D) Drafting product descriptions with merchandiser review: data available, metric clear, owner keen — **Correct.** Bounded value, low risk, clear measurement, reversible — a good first pilot.

K8 **Answer: B** Review: 2.3, 2.7

A proposal has strong value and sponsorship but no agreed way to measure success within a year. What is the right disposition?

✗ (A) Increase its value score to compensate. — **Wrong.** Adjusting one dimension to offset another is the blending error.

✓ (B) Advance with a condition to define a metric and baseline, or park until one exists. — **Correct.** Measurement clarity is a separate dimension that strengths elsewhere do not compensate for.

✗ (C) Advance unconditionally — measurement can be sorted out later. — **Wrong.** Unmeasurable success undermines gates and later decisions.

✗ (D) Reject permanently. — **Wrong.** The idea may become viable once measurement is defined.

MODULE 3

Prioritizing with a Value-Effort Lens

Making the funding order defensible: criteria and weights fixed before scoring, effort estimated by component, the direction of error priced into value, and an impact-effort matrix read into a shortlist with stated dispositions.

YOU WILL BE ABLE TO

- Set prioritization criteria and weights before scoring
- Define value so that it means the same thing to every stakeholder
- Estimate effort by component: data readiness, integration, compliance, adoption resistance and operational capacity
- Price the direction of error into business value
- Plot candidates on an impact-effort matrix and read the groups
- Turn the matrix into a shortlist with stated dispositions
- Make trade-offs transparent and sequence investments

IN THIS MODULE

1. Why a value-effort lens
2. Criteria and weights before scoring
3. Defining value consistently
4. Effort by component
5. Pricing the direction of error
6. The impact-effort matrix
7. From matrix to shortlist
8. Trade-off transparency and alignment
9. Investment sequencing
10. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4–5 hours

Prioritization decides which AI initiatives an enterprise funds first. Every stakeholder believes their use case is high-value and low-effort; the value-effort lens makes the decision defensible by fixing the rules before anyone scores, decomposing effort into its real components, pricing in the cost of being wrong, and showing the result in a form executives can debate.

KEY TAKEAWAYS

1. Set **criteria, definitions and weights before scoring** — otherwise the scores will be bent to fit favourites.
2. Define **value** with anchors so "high impact" means the same to everyone; express it net, as a range, with confidence.
3. Estimate **effort by component**: data readiness, integration, compliance burden, adoption resistance, operational capacity to absorb AI workload.
4. **Price the direction of error**: model errors are asymmetric (false positives ≠ false negatives) and estimates err in a predictable direction (optimism).
5. Plot on an **impact-effort matrix** with ranges and risk visible; read quadrants as quick wins, strategic bets, fill-ins and money pits.
6. Turn groups into a **shortlist with stated dispositions**: fund now, prepare, bundle, defer, drop — each with a reason.
7. Show **trade-offs and sensitivity** openly, and **sequence** by dependency and capacity.

3.1 Why a value-effort lens

A value-effort lens is simple enough for executives to grasp in a minute and rigorous enough to defend. Its power comes not from the two-by-two picture but from the discipline behind the axes: what counts as value, how effort is built up, and how uncertainty and error are treated. Without that discipline, the matrix becomes a drawing of opinions.

KEY IDEA

A prioritization is **defensible** when a stakeholder who disagrees can see exactly which criterion, weight or estimate produced the result — and can argue about that, rather than about the whole ranking.

3.2 Setting criteria and weights before scoring

When criteria and weights are set after scores are visible, people — usually unconsciously — adjust weights until their preferred use case wins. The sequence matters:

1. **Choose criteria** under value and effort (keep each list short: 3–5 criteria).
2. **Write anchors** for each criterion: what a 1, 3 and 5 look like in observable terms.
3. **Agree weights** with the decision-makers — for example by allocating 100 points, or by pairwise comparison.
4. **Freeze and publish** criteria, anchors and weights before any scoring.
5. **Score**, ideally by more than one person, then calibrate differences.
6. **Change only through a recorded decision** — and re-score everything if weights change.

Weighting method	How it works	Good for
100-point allocation	Each decision-maker distributes 100 points across criteria; average	Quick consensus; visible differences in priorities
Pairwise comparison	Compare criteria two at a time; derive weights	Surfacing real trade-offs when leaders disagree
Equal weights	All criteria count equally	Early screens; avoiding false precision

EXAM TRAP

If a question describes a team that adjusts weights after seeing which use case ranks first, the issue is **post-hoc weighting** — the remedy is to set and freeze criteria and weights before scoring.

3.3 Defining value consistently

"High impact" means revenue to Sales, risk reduction to Compliance and hours saved to Operations. Anchored value criteria make these comparable.

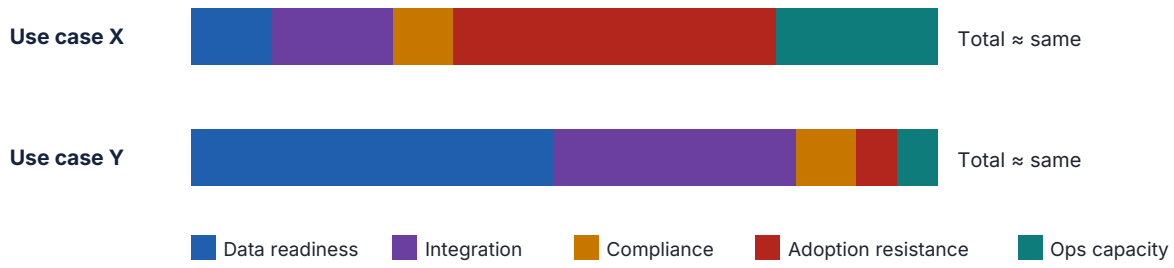
Value criterion	1 — low	3 — medium	5 — high
Financial value (net, annual, after ledger)	< €50k	€250k–1m	> €3m
Strategic alignment	Not linked to a strategic objective	Supports one objective	Directly enables a top-3 priority
Risk reduction	Negligible	Reduces a moderate known risk	Reduces a material, rated risk
Customer / employee experience	No measurable change	Measurable change in one journey	Material change in a core journey
Time to value	> 18 months	6–12 months	< 3 months

CONSULTANT'S TIP · Carry confidence alongside value

Record each value estimate as a range with a confidence level (high/medium/low) based on the evidence ladder from Course 1. Two use cases with the same midpoint but different confidence are not equally attractive.

3.4 Estimating effort by component

A single effort number ("medium", "€400k") hides *what kind* of work is required — and different kinds of work draw on different, often scarce, resources. Estimate each component separately, then combine.



Same single number, completely different work: X is a change-management challenge, Y is a data and integration project.

Figure 3.1 — Same total effort, different work

Effort component	What it covers	1 — low	5 — high
Data readiness	Work to make data available, permitted, clean, labelled, representative	Data ready and governed	Data must be created, labelled or legally enabled
Integration	Connecting to systems of record, identity, workflow tools, logging	Standalone or one modern API	Many legacy systems, real-time, write-back
Compliance burden	Risk assessment, DPIA, legal review, documentation, oversight design, audits	Internal, low-risk use	High-risk use with regulatory obligations
Adoption resistance	Training, process redesign, role change, trust building, works-council engagement	Users ask for it	Threatens roles or professional judgment; strong scepticism
Operational capacity	Capacity to absorb new AI workload: review queues, exception handling, monitoring, support, incident response	Existing teams can absorb	New team or shifts needed; scarce experts

PITFALL · Forgetting operational capacity

A fraud model that flags 3× more cases is only valuable if investigators can work them. If the review team is already at capacity, the "effort" of the use case includes hiring or reprioritising investigators — or the value never materialises. This is the ledger from Module 1, priced as effort.

NOTE

Effort components can be scored on 1–5 anchors and weighted, or estimated in money and time where data allows. Either way, keep the components visible: they also tell you **who** must provide the effort.

3.5 Pricing the direction of error into value

Two kinds of error must be priced into expected value.

1 · Model errors are asymmetric

An AI system will make mistakes, and the two directions of mistake rarely cost the same. A missed fraud (false negative) may cost €5,000; a false alarm (false positive) may cost 20 minutes of investigator time and some customer irritation. Value must reflect the expected mix and cost of each.

Expected value = (benefit of correct outcomes) − (# false positives × cost per FP) − (# false negatives × cost per FN) − run costs

EXAMPLE · Same accuracy, different value

	Model A	Model B
Fraud cases caught (of 1,000)	800	700
False alarms (of 100,000 legitimate)	4,000	1,000
Value of caught fraud @ €5,000	€4.0m	€3.5m
Cost of false alarms @ €40 review + €60 customer impact	−€0.40m	−€0.10m
Expected value (before run costs)	**€3.60m**	**€3.40m**

Now suppose investigators can review only 2,000 alerts per year. Model A's 4,800 alerts cannot be worked; its realisable value collapses — capacity and error direction interact.

2 · Estimates err in a predictable direction

Early estimates of AI initiatives are systematically optimistic: value is overestimated, effort and time underestimated (optimism bias, planning fallacy). Treat this as a known direction of error and correct for it explicitly.

Technique	How
Haircut by evidence level	Discount value by confidence, e.g., vendor claim −50%, PoC −30%, pilot on own data −10%
Downside scenario	Show value if adoption is half the plan and error costs double
Reference-class check	Compare with outcomes of similar past initiatives, internal or external
Effort contingency by component	Larger contingency on components with low certainty (data, adoption)
Asymmetric ranges	Ranges skewed toward lower value and higher effort, not symmetric ±

EXAM TRAP

"Use the model with the highest accuracy" is often the wrong answer. The right answer prices **which** errors each model makes and what they cost — and checks whether downstream capacity can act on the outputs.

3.6 The impact-effort matrix

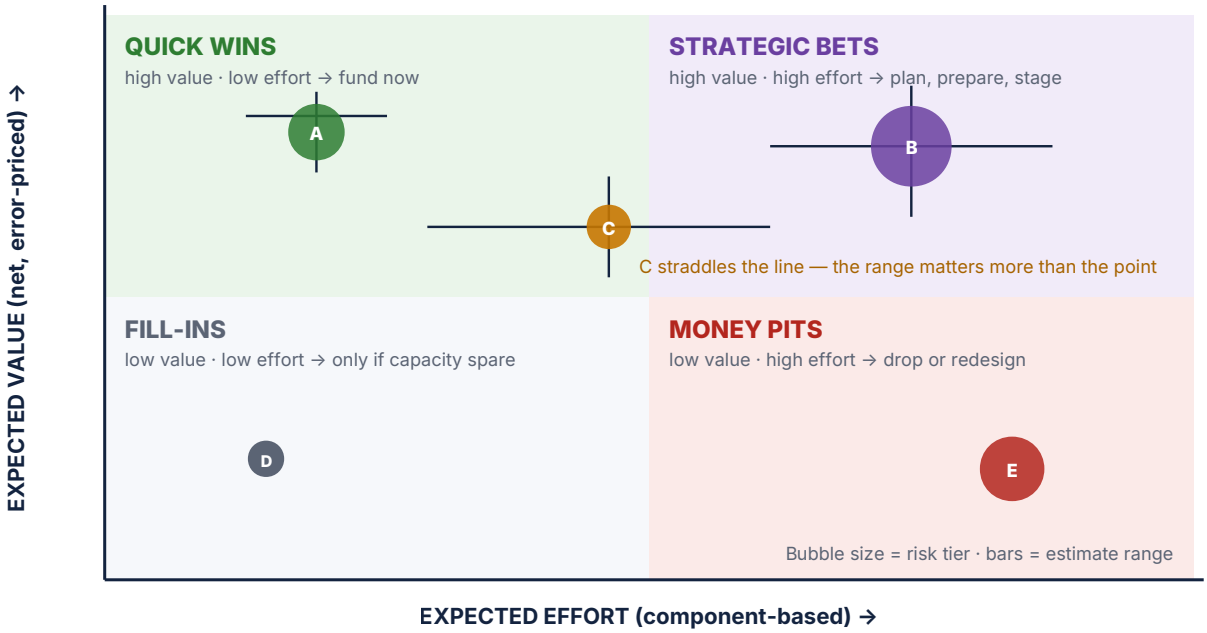


Figure 3.2 — Impact-effort matrix with ranges and risk

Quadrant	Reading	Typical disposition
Quick wins	High value, low effort	Fund now — but check they are not only "quick" because effort components were missed
Strategic bets	High value, high effort	Prepare enablers; stage funding through gates
Fill-ins	Low value, low effort	Do only with spare capacity or as part of a bundle
Money pits	Low value, high effort	Drop or redesign

BEST PRACTICE · Make uncertainty visible

- Plot ranges (bars) not just points; a candidate straddling a boundary needs more evidence, not a confident label.
- Use bubble size or colour for risk tier from Module 2.
- Mark dependencies (arrows) between candidates.
- Never let the quadrant label replace the reasoning.

3.7 From matrix to shortlist with stated dispositions

The matrix is a reading aid; the output is a **shortlist** in which every candidate has a disposition and a one-line reason.

Disposition	Meaning	Example reason
Fund now	Proceed to delivery this cycle	Quick win; data ready; owner committed; net value €0.4–0.7m
Prepare	Invest in enablers first (data, platform, policy)	Strategic bet blocked by master-data quality; fix in Q1

Disposition	Meaning	Example reason
Bundle	Deliver together with related candidates	Shares retrieval platform with two others; marginal effort low
Defer	Not this cycle; revisit on trigger	Capacity: review team cannot absorb more alerts until Q3
Drop	Remove from active list	Money pit after error pricing; value under €50k net

Disposition statement: **[Use case]** — **[disposition]** because **[decisive criterion or evidence]**; revisit when **[trigger]** (if applicable)

3.8 Trade-off transparency and stakeholder alignment

- **Show what you gave up.** For each "fund now", name what was not funded as a consequence.
- **Run a sensitivity test.** Does the top of the list change if the main weight moves ±20%? If not, say so — it builds confidence. If it does, the decision is really about that weight.
- **Calibrate scores in a session.** Where scorers differ by two or more points, discuss the evidence, not the people.
- **Use agreed tie-breakers.** E.g., lower risk first, then shorter time to value, then reusable learning.
- **Handle the HiPPO** (highest-paid person's opinion) by pointing to the published criteria: "If we want to fund it anyway, which criterion should change, and should it change for everyone?"

CONSULTANT'S TIP · Script for an overruled favourite

> *"Your use case scores high on strategic value. What holds it back is data readiness and the review team's capacity. If we fix master data in Q1, it moves into the top group — so the disposition is 'prepare', not 'drop'."*

3.9 AI investment sequencing

Sequencing principle	Application
Dependencies first	Enablers (data, platform, policy) before the use cases they unblock
Learning order	Low-risk use cases before high-risk ones of the same pattern
Capacity limits	Do not start more than delivery and operational teams can absorb
Reuse	Sequence use cases that share components so later ones are cheaper
Value early	Include at least one quick win to fund and legitimise the programme
Gated funding	Fund strategic bets in stages tied to evidence

3.10 Case study: Brightwater Bank

CASE STUDY **Nine qualified use cases, budget for three**

Brightwater Bank (fictional) · regional bank · 900k customers

Situation

Nine use cases passed qualification. The executive committee has budget and delivery capacity for three this year. Before any scoring, the advisor facilitates agreement on criteria and weights: value — net financial (40), risk reduction (25), strategic fit (20), time to value (15); effort by five components with anchors.

Key moments

- **Effort by component** revealed that the mortgage-document extraction use case, labelled "low effort" by its sponsor, carried high integration effort (three legacy systems) and high compliance burden (credit process). It moved from quick win to strategic bet.
- **Error pricing** changed the AML alert-triage case. Model A deprioritised more alerts (bigger saving) but missed more true positives; with regulatory penalty costs for missed suspicious activity, Model B's more conservative configuration delivered higher expected value.
- **Operational capacity:** the fraud-detection upgrade would generate 60% more alerts; the investigation team was full. Disposition: *defer* until two investigators are hired, or bundle with the alert-triage case that frees investigator time.
- **Sensitivity:** moving the risk-reduction weight ±20% did not change the top three — reported explicitly to the committee.

Shortlist (excerpt)

Use case	Disposition	Reason
Advisor meeting-notes assistant	Fund now	Quick win; low risk; data ready; owner keen
AML alert triage (conservative config)	Fund now	Highest error-priced value; frees investigators
Fraud-detection upgrade	Bundle	Run after alert triage frees capacity
Mortgage-document extraction	Prepare	Strategic bet; integration and compliance work first
Branch footfall forecasting	Drop	Low value after ledger; branches closing

Discussion questions

1. Why did criteria and weights have to be agreed before scoring in this committee?
2. What did the component view reveal that a single effort score would have hidden?
3. Explain in two sentences why the higher-saving AML configuration had lower expected value.
4. Why is "bundle" a better disposition than "fund now" for the fraud upgrade?

Model answers (brief)

1. Several sponsors were present; post-hoc weights would have been bent toward favourites and the result would not be defensible.

2. That the “low effort” mortgage case required heavy integration and compliance work — different teams, longer time.
3. It missed more genuinely suspicious activity, and each miss carries a large regulatory and financial cost. Pricing false negatives outweighed the extra time saved.
4. Its value depends on investigator capacity, which the alert-triage case creates; sequencing them together makes the value realisable.

3.11 Consulting templates

T3.1 Criteria & Weights Charter

Purpose: Fix the rules before anyone scores. **Use when:** before the first scoring session

Axis	Criterion	Anchor 1	Anchor 3	Anchor 5	Weight
Value					
Value					
Value					
Value					
Effort	*Data readiness*				
Effort	*Integration*				
Effort	*Compliance burden*				
Effort	*Adoption resistance*				
Effort	*Operational capacity*				

Weighting method used: .....

Tie-breakers (in order): .....

Agreed and frozen by / date: .....

T3.2 Component Effort Estimator

Purpose: Estimate effort by component and name who supplies it.

Use case: .....

Component	Score 1–5	Key driver	Who supplies the effort	Certainty H/M/L	Contingency
Data readiness					
Integration					
Compliance burden					
Adoption resistance					
Operational capacity					
Weighted total					

T3.3 Error-Direction Value Adjuster

Purpose: Price model errors and estimate optimism into value.

MODEL ERROR PRICING

Item	Count / yr	Unit value / cost	Annual €
Correct positive outcomes			
False positives			
False negatives			
Capacity limit check (can outputs be acted on?)			
Expected value			

ESTIMATE-ERROR CORRECTION

Evidence level and haircut %: .....

Downside scenario value: .....

Reference-class comparison: .....

Adjusted value range: .....

T3.4 Shortlist Disposition Register

Purpose: Record the reading of the matrix as decisions with reasons.

Use case	Value (range)	Effort (range)	Quadrant	Risk tier	Disposition	Decisive reason	Revisit trigger

Sensitivity test result (±20% on main weight): .....

What was not funded as a consequence: .....

3.12 Best practices and pitfalls

✓ DO THIS

- Freeze criteria, anchors and weights before scoring.
- Anchor value so every stakeholder means the same thing.
- Estimate effort by five components and name who supplies it.
- Price false positives and false negatives separately.
- Haircut value by evidence level; show a downside case.
- Plot ranges and risk on the matrix.
- Give every candidate a disposition and reason.

× AVOID THIS

- Adjusting weights after seeing results.
- “High impact” without definitions.
- A single effort number.
- Choosing the model with the best headline accuracy.
- Symmetric ± ranges on optimistic midpoints.
- Letting quadrant labels replace reasoning.
- Hiding what was not funded.

3.13 Module 3 cheat sheet

MODULE 3 · ONE-PAGE SUMMARY

Prioritizing with a Value-Effort Lens

RULES FIRST

Criteria → anchors → weights (100 points / pairwise / equal) → freeze & publish → score → calibrate. Change only by recorded decision; re-score all.

VALUE CRITERIA

Net financial · strategic fit · risk reduction · experience · time to value. Range + confidence.

EFFORT COMPONENTS

Data readiness · integration · compliance burden · adoption resistance · operational capacity to absorb AI workload (reviews, exceptions, monitoring). Name who supplies each.

PRICE MODEL ERROR

EV = benefit − FP×cost − FN×cost − run cost. Check capacity to act on outputs. Highest accuracy ≠ highest value.

PRICE ESTIMATE ERROR

Optimism is predictable: haircut by evidence level, downside scenario, reference class, component contingency, skewed ranges.

MATRIX

Quick wins · strategic bets · fill-ins · money pits. Plot ranges; bubble = risk; arrows = dependencies.

DISPOSITIONS

Fund now · prepare · bundle · defer · drop — each with decisive reason and trigger.

DEFENSIBILITY & SEQUENCING

Show what you gave up · sensitivity ±20% · calibration · tie-breakers · answer the HiPPO with criteria. Sequence: dependencies first, low-risk before high-risk, within capacity, reuse, at least one quick win, gated funding for bets.

3.14 Knowledge check

Answer all questions before turning the page.

K1 **Why must criteria and weights be agreed before scoring?**

(A) Because weights set after seeing scores tend to be bent toward preferred use cases, making the result indefensible.
(B) Because scoring takes less time that way.
(C) Because weights cannot be changed later.
(D) Because regulators require pre-set weights.

K2 **Two use cases both have "medium" total effort. One is mainly adoption resistance; the other mainly data and integration. Why does this matter?**

(A) Data effort should always be weighted zero.
(B) It doesn't; total effort is what counts.
(C) The adoption-heavy one should be dropped.
(D) They draw on different teams and risks, which affects sequencing, staffing and likelihood of success.

K3 **A fraud model would triple alerts, but the investigation team is at full capacity. Which effort component is most directly affected?**

(A) Compliance burden
(B) Integration
(C) Operational capacity to absorb new AI workload
(D) Data readiness

K4 **Model A catches more fraud but raises four times as many false alarms as Model B. How should you compare them?**

(A) Choose B, because it has fewer false alarms.
(B) Price true positives, false positives and false negatives, check capacity to act, and compare expected value.
(C) Choose A, because it catches more fraud.
(D) Choose whichever has higher overall accuracy.

K5 **Which technique corrects for the predictable optimism of early AI estimates?**

(A) Ignoring uncertainty to keep it simple
(B) Symmetric ±10% ranges
(C) Haircutting value by evidence level and showing a downside scenario
(D) Using the sponsor's estimate

K6 **A use case is high value and high effort. Which disposition is typically most appropriate?**

(A) Treat as a fill-in
(B) Prepare enablers and fund in gated stages
(C) Drop
(D) Fund fully now

K7 **A candidate's value-effort range straddles the boundary between "quick win" and "money pit". What should you conclude?**

(A) Drop it immediately.
(B) Label it a quick win to keep momentum.
(C) Average its position and use that quadrant.
(D) It needs more evidence before a confident disposition; the range matters more than the midpoint.

K8 **The CFO insists on funding a use case that ranks ninth. What is the most effective advisory response?**

(A) Ask which criterion should change, and whether it should change for every use case — then re-score transparently if agreed.
(B) Refuse to include it.
(C) Remove the CFO from the process.
(D) Change its scores so it ranks higher.

3.15 Knowledge check: answers and explanations

K1 **Answer: A** Review: 3.2

Why must criteria and weights be agreed before scoring?

✓ **A** Because weights set after seeing scores tend to be bent toward preferred use cases, making the result indefensible. — **Correct.** Fixing rules first prevents motivated, post-hoc weighting.

✗ B Because scoring takes less time that way. — **Wrong.** Speed is not the reason.

✗ C Because weights cannot be changed later. — **Wrong.** They can change, by recorded decision with re-scoring.

✗ D Because regulators require pre-set weights. — **Wrong.** No regulation requires this.

K2 **Answer: D** Review: 3.4

Two use cases both have "medium" total effort. One is mainly adoption resistance; the other mainly data and integration. Why does this matter?

✗ A Data effort should always be weighted zero. — **Wrong.** Data effort is a real component.

✗ B It doesn't; total effort is what counts. — **Wrong.** The totals hide different resource needs and risks.

✗ C The adoption-heavy one should be dropped. — **Wrong.** Adoption effort is manageable; it just needs the right resources.

✓ **D** They draw on different teams and risks, which affects sequencing, staffing and likelihood of success. — **Correct.** Component effort reveals what kind of work and whose effort is needed.

K3 **Answer: C** Review: 3.4

A fraud model would triple alerts, but the investigation team is at full capacity. Which effort component is most directly affected?

✗ A Compliance burden — **Wrong.** Compliance may matter, but the scenario is about workload.

✗ B Integration — **Wrong.** Integration concerns system connections, not review capacity.

✓ **C** Operational capacity to absorb new AI workload — **Correct.** Alerts must be worked by people; without capacity the value is not realised.

✗ D Data readiness — **Wrong.** Data is not the stated constraint.

K4 **Answer: B** Review: 3.5

Model A catches more fraud but raises four times as many false alarms as Model B. How should you compare them?

✗ A Choose B, because it has fewer false alarms. — **Wrong.** This ignores false-negative cost.

✓ **B** Price true positives, false positives and false negatives, check capacity to act, and compare expected value. — **Correct.** Pricing the direction of error and capacity gives the defensible comparison.

✗ C Choose A, because it catches more fraud. — **Wrong.** This ignores false-positive cost and capacity limits.

✗ D Choose whichever has higher overall accuracy. — **Wrong.** Accuracy hides the asymmetric costs of each error type.

K5 **Answer: C** Review: 3.5

Which technique corrects for the predictable optimism of early AI estimates?

✗ A Ignoring uncertainty to keep it simple — **Wrong.** This hides the direction of error.

✗ B Symmetric ±10% ranges — **Wrong.** Symmetric ranges assume errors are equally likely in both directions, which they are not.

✓ **C** Haircutting value by evidence level and showing a downside scenario — **Correct.** These explicitly correct for the known direction of estimation error.

✗ D Using the sponsor's estimate — **Wrong.** Sponsor estimates are often the most optimistic.

K6 **Answer: B** Review: 3.6–3.7

A use case is high value and high effort. Which disposition is typically most appropriate?

✗ (A) Treat as a fill-in — **Wrong.** Fill-ins are low value, low effort.

✓ (B) Prepare enablers and fund in gated stages — **Correct.** Strategic bets need preparation and evidence-based staging.

✗ (C) Drop — **Wrong.** Dropping wastes high value; drop suits money pits.

✗ (D) Fund fully now — **Wrong.** High effort and uncertainty argue for staged funding and enablers first.

K7 **Answer: D** Review: 3.6

A candidate's value-effort range straddles the boundary between "quick win" and "money pit". What should you conclude?

✗ (A) Drop it immediately. — **Wrong.** The evidence does not yet support dropping.

✗ (B) Label it a quick win to keep momentum. — **Wrong.** This hides real uncertainty.

✗ (C) Average its position and use that quadrant. — **Wrong.** Averaging hides the spread that drives the decision.

✓ (D) It needs more evidence before a confident disposition; the range matters more than the midpoint. — **Correct.** Uncertainty spanning quadrants calls for evidence, not a label.

K8 **Answer: A** Review: 3.8

The CFO insists on funding a use case that ranks ninth. What is the most effective advisory response?

✓ (A) Ask which criterion should change, and whether it should change for every use case — then re-score transparently if agreed. — **Correct.** This answers the HiPPO with the published criteria while respecting authority.

✗ (B) Refuse to include it. — **Wrong.** Decision rights belong to the executives; the advisor ensures transparency.

✗ (C) Remove the CFO from the process. — **Wrong.** This is neither feasible nor appropriate.

✗ (D) Change its scores so it ranks higher. — **Wrong.** This destroys the integrity of the method.

MODULE 4

Building the Prioritized Opportunity Portfolio

A list of good ideas is not yet a portfolio. Track every candidate through one intake, classify risk before weighing value, score on two fixed axes, balance and consolidate — and route the result into an investment decision that can say “not this cycle”.

YOU WILL BE ABLE TO

- Distinguish a portfolio from a list of good ideas
- Design a tracked AI intake process that captures every visible candidate
- Classify regulatory and AI risk exposure against recognized frameworks before weighing value
- Score the full set on expected value and expected effort against one fixed set of criteria
- Group a ranked shortlist into a portfolio balanced across value, risk and effort
- Resolve duplicate automation through consolidation
- Route the portfolio into an investment and resource allocation decision that can return “not this cycle”

IN THIS MODULE

1. List vs portfolio
2. The tracked AI intake process
3. Risk classification before value
4. Two-axis scoring on fixed criteria
5. Balancing the portfolio
6. Consolidating duplicate automation
7. Routing to the investment decision
8. “Not this cycle”
9. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4 hours

Choosing among AI opportunities is where advisory work becomes accountable. This module turns the screened, prioritized candidates of Modules 2 and 3 into a **portfolio**: a set of initiatives selected *together*, tracked from intake to decision, classified for risk before anyone gets attached to the value, scored on one fixed set of criteria, balanced, de-duplicated, and presented as an investment decision in which "not this cycle" is a legitimate, correct answer.

KEY TAKEAWAYS

1. A **portfolio** is selected as a whole — balancing value, risk, effort, dependencies and capacity. A list is not.
2. Run every candidate the organization can see through **one tracked intake** — requests, discoveries, shadow AI, vendor proposals, embedded SaaS features.
3. **Classify risk first**, against recognized frameworks (EU AI Act tiers, NIST AI RMF, ISO/IEC 23894 / 42001), before value is weighed.
4. Score the **full set** on two separate axes — **expected value** and **expected effort** — using **one fixed set of criteria**.
5. Group the shortlist into a portfolio **balanced across value, risk and effort**, within capacity.
6. Resolve **duplicate automation by consolidation**: one capability, one solution, many consumers.
7. Route to an **investment and resource allocation decision** with options — including **"not this cycle"**.

4.1 A list of good ideas is not a portfolio

	List of good ideas	Portfolio
Selection	Each idea judged on its own merit	Ideas selected together, as a set
Risk	Risk of each idea	Aggregate risk exposure and concentration
Capacity	Ignored or assumed	Fits delivery and operational capacity
Dependencies	Invisible	Shared enablers and sequences explicit
Duplication	Same capability built several times	Consolidated
Decision	"Approve the top ten"	Investment and resource allocation with options
Accountability	Diffuse	Owner per initiative; portfolio owner for the whole

KEY IDEA

The top ten individually best ideas are often **not** the best portfolio: they may all be high-risk, all depend on the same scarce data team, or duplicate one another.

4.2 The tracked AI intake process

A portfolio can only be as complete as the intake that feeds it. If business units, IT, vendors and individual employees each start AI initiatives through different routes, the portfolio owner sees only part of the picture — and duplicates, shadow deployments and unassessed risks multiply.

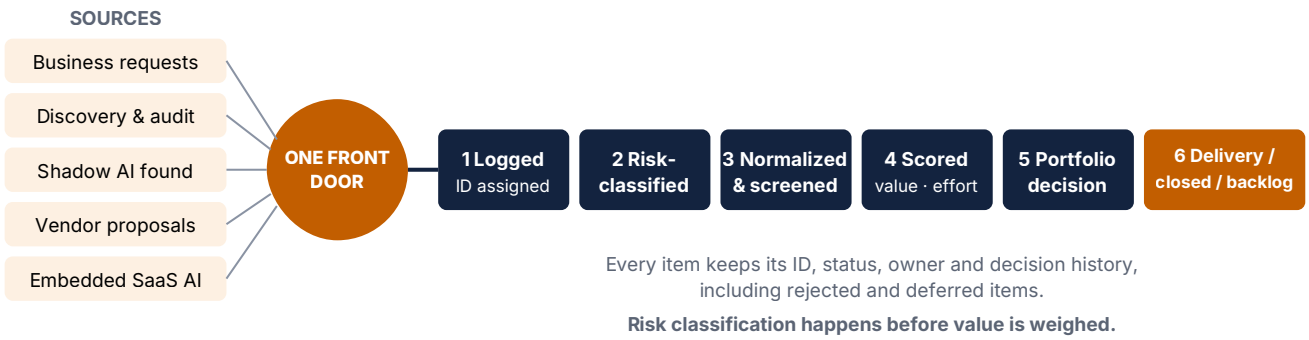


Figure 4.1 — A tracked AI intake process

Design element	Good practice
Single front door	One intake route for all AI candidates, whatever the source; publicised and easy
Minimum intake form	Short: problem, users, data, intended use, autonomy, sponsor (detail comes at screening)
Unique ID and status	Every item tracked through stages: logged → risk-classified → screened → scored → decided → delivery / closed / backlog
Ownership	Intake owner (e.g., AI portfolio office); named requester and business owner per item
Service levels	E.g., acknowledgement in 2 days; screening decision in 4 weeks
Coverage	Includes discoveries, shadow AI found, vendor-initiated proposals and AI features switching on in existing software
Decision history	Dispositions and reasons kept, including rejections — so re-proposals are recognised
Link to inventory	Approved items enter the AI system inventory (Course 1) on deployment

PITFALL · Intake as bureaucracy

If the intake is slow and heavy, people route around it, and shadow AI grows. Keep the first step light, publish service levels, and make the benefit visible: help with data, legal and vendor review that a team could not get alone.

4.3 Classifying risk exposure before weighing value

Classify each candidate's regulatory and AI risk exposure **before** value is weighed, for three reasons:

- **Some uses must not proceed at all** — identifying them early avoids wasted analysis.
- **Risk class drives effort** — high-risk uses carry compliance, documentation and oversight obligations that belong in the effort estimate.
- **Anchoring** — once a team has fallen in love with a value number, risk findings tend to be minimised.

Framework	What it offers for classification
EU AI Act (Regulation (EU) 2024/1689)	Risk tiers by intended purpose: prohibited practices; high-risk systems (e.g., uses listed in Annex III such as employment, credit scoring, access to essential services, education); transparency obligations (e.g., chatbots, synthetic content); minimal risk. Obligations differ for providers and deployers. Confirm current application dates.

Framework	What it offers for classification
NIST AI Risk Management Framework (AI RMF 1.0)	Voluntary framework with four functions — Govern, Map, Measure, Manage. "Map" establishes context and categorises the system and its impacts; useful for consistent risk identification outside the EU.
ISO/IEC 23894:2023	Guidance on AI risk management, applying ISO 31000 principles to AI.
ISO/IEC 42001:2023	AI management system standard: requires AI risk assessment and impact assessment within an organizational management system.
Sector rules & internal taxonomy	E.g., model risk management in banking, medical-device rules in healthcare, GDPR (DPIA triggers), the organization's own risk appetite.

Portfolio risk class (example internal mapping)	Typical triggers	Portfolio consequence
Stop	Possible prohibited practice; outside risk appetite	Not scored; referred to legal
High	EU AI Act high-risk use; decisions on people's rights; safety	Scored, with full compliance burden in effort; specialist sign-off required
Limited / transparency	Customer-facing chatbots, generated content	Scored; disclosure and content controls in effort
Minimal	Internal productivity, low-impact uses	Scored; standard controls

> **NOTE**
> The advisor's role is to make classification happen consistently and early, using the organization's chosen framework, and to involve legal and risk specialists for formal determinations — not to give legal opinions (Course 1, scope of authority).

4.4 Two-axis scoring against one fixed set of criteria

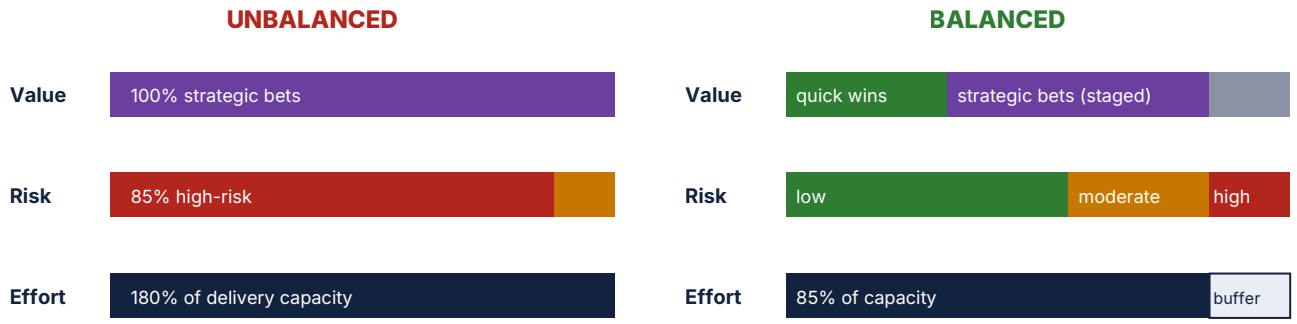
Once risk is classified, score **every** remaining candidate on two separate axes, using the same criteria, anchors and weights fixed as in Module 3.

Principle	Why
Two separate axes: expected value and expected effort	Keeps the trade-off visible; blending hides it
"Expected" = probability-weighted, error-priced, haircut by evidence	Compares realistic outcomes, not best cases
One fixed set of criteria for the full set	Comparability; no special criteria invented for favourites
Score the full set, not a pre-selected few	Portfolio balance and consolidation need the whole picture
Same scorers or calibrated scorers	Reduces inconsistency between units
Risk class carried alongside, not hidden in a score	Needed for balancing (next section)

EXAM TRAP

A sponsor proposes an extra criterion — "innovation visibility" — that only their use case scores well on. Adding criteria for one candidate breaks comparability. The answer is to evaluate any new criterion for **all** candidates, through a recorded decision, or not at all.

4.5 Balancing the portfolio across value, risk and effort



Balance is not equal shares — it is a mix the organization can deliver, absorb and govern, with value arriving over time.

Figure 4.2 — An unbalanced vs a balanced portfolio

Balance dimension	Check	Typical guideline (adapt to context)
Value timing	Do some initiatives deliver value soon while others build for later?	At least one quick win per cycle; strategic bets staged
Value type	Mix of cost, revenue, risk reduction and enabling value?	Avoid all-cost portfolios that erode support
Risk	Is aggregate exposure within appetite? Too many high-risk uses at once?	Limit concurrent high-risk initiatives to what governance can oversee
Effort / capacity	Does total effort fit delivery and operational capacity — by component?	Plan to ~80–85% of capacity; keep buffer
Concentration	Does too much depend on one vendor, platform or team?	Diversify or plan mitigation
Spread across units	Are some functions over- or under-served — and is that intended?	Explain imbalance; do not force equality

CONSULTANT'S TIP · Balance by component

Check capacity per effort component: a portfolio can fit total budget yet need three times the available data-engineering time, or flood the same review team with outputs from several use cases at once.

4.6 Resolving duplicate automation through consolidation

When intake is complete, duplicates appear: three units want "a document assistant", two vendors propose overlapping extraction tools, and an embedded SaaS feature already does part of what a new project plans to build. Duplicate automation multiplies cost, fragments data and controls, and confuses users.

Step	Action
1 · Detect	Normalize candidates to capability level (Course 1, Module 2) and look for the same AI task on similar inputs
2 · Test for real differences	Different data, risk class, latency, language or user needs may justify separate solutions
3 · Choose a consolidation pattern	See table below
4 · Re-score the consolidated item	Value adds up across consumers; effort is shared; risk follows the highest-risk use
5 · Communicate	Explain to each requester what they get, when and why

Consolidation pattern	When
Merge	Requests are essentially the same — one initiative, multiple beneficiaries
Shared platform, separate configurations	Same capability, different content or rules per unit
Lead unit, then roll out	One unit pilots; others adopt after evidence
Use what exists	An existing or embedded tool already covers the need — retire the duplicate request
Keep separate (documented)	Genuine differences justify it; record why

> **PITFALL · Consolidation by decree**
> Forcing every unit onto one solution without testing real differences produces a tool nobody likes — and new shadow AI. Consolidate where needs truly overlap and explain the trade-off.

4.7 Routing the portfolio into an investment and resource allocation decision

The portfolio is complete only when it reaches the body that allocates money and people, in a form that body can decide on. A good decision package offers **options**, not a single take-it-or-leave-it list.

Decision package element	Content
Recommendation	Preferred option, in one paragraph
Options	E.g., A: full portfolio · B: core portfolio (quick wins + one prepared bet) · C: not this cycle — invest only in enablers
Portfolio view	Initiatives with value and effort ranges, risk class, dependencies, dispositions
Resource request	Budget and people by component (data, integration, compliance, change, operations)
Capacity check	Evidence that the organization can deliver and absorb the portfolio
Risk profile	Aggregate exposure; high-risk items and their governance
Gates and funding stages	What is funded now, what is released at which gate on which evidence
What is not funded	Deferred and dropped items with reasons and triggers

Decision package element	Content
Decision required	Who decides, by when; what happens on each option

BEST PRACTICE

Fund strategic bets in stages: release money for the next stage when the previous gate's evidence criteria are met. This turns a large, uncertain commitment into a sequence of smaller, evidence-based decisions.

4.8 "Not this cycle" — a correct answer

A portfolio process that can only say "yes" is not a decision process. "Not this cycle" is the correct outcome when:

- No candidate clears the value threshold after error pricing and the ledger.
- Foundations are missing (data, platform, governance), so the right investment is in enablers.
- Delivery or operational capacity is already committed.
- Risk exposure would exceed appetite or governance capacity.
- Better evidence is expected soon (e.g., a running pilot reports next quarter).

KEY IDEA

"Not this cycle" must be **recorded** with reasons and revisit triggers, and the candidates must stay in the tracked intake backlog. That makes it a decision, not a rejection — and protects the credibility of the process for the next cycle.

EXAM TRAP

A scenario may present pressure to fund "at least something" to show AI progress. If no candidate meets the criteria, the advisory answer is to recommend enabler investment or "not this cycle" with triggers — not to lower the bar for one cycle.

4.9 Case study: Vellmar Group

CASE STUDY **From 41 intake items to a fundable portfolio**

Vellmar Group (fictional) · industrial conglomerate · four divisions · 22,000 staff

Situation

Vellmar launched a group AI intake six months ago. The portfolio office now holds 41 items: 23 business requests, 8 from discovery, 5 shadow-AI tools found in an audit, 3 vendor proposals and 2 AI features that went live in existing HR and CRM software through updates.

What the advisor did

1. **Risk first:** classified all 41 against the EU AI Act tiers and the group risk taxonomy. One item (emotion recognition of employees in video calls) was flagged as a likely prohibited practice and referred to legal; six were high-risk (including the HR software's new candidate-ranking feature, already switched on).
2. **Screen and score:** 28 passed screening; all 28 were scored on expected value and expected effort against one fixed criteria set agreed with the group executive team.
3. **Consolidate:** 7 requests for "document assistants" across divisions became one shared retrieval platform with divisional content; two overlapping vendor extraction proposals were reduced to one; one request was retired because the embedded CRM feature already covered it.
4. **Balance:** the initial top-ten list contained five high-risk items and needed 170% of data-engineering capacity. The balanced portfolio held two high-risk items with full governance, three quick wins, two staged strategic bets and the shared platform, at 85% of capacity.
5. **Decision package:** three options. The executive committee chose option B (core portfolio) and also decided to *switch off* the HR candidate-ranking feature until its risk assessment was complete.

Discussion questions

1. Why was it important that the embedded HR and CRM features entered the intake?
2. What would have happened if value had been scored before risk classification?
3. How did consolidation change value, effort and risk for the document assistants?
4. Under what circumstances should the committee have chosen option C, "not this cycle"?

Model answers (brief)

1. Embedded features can bring high-risk uses in without procurement; without intake, the candidate-ranking feature would have run unassessed.
2. Teams would have anchored on value; the likely prohibited and high-risk items might have consumed analysis or been under-scrutinised.
3. Value aggregated across seven consumers; effort was shared (one platform); risk was set by the most sensitive content domain, so controls were designed once to that level.
4. If capacity was already committed, if foundations such as data or governance were missing, or if no candidate met the value threshold — with triggers recorded for the next cycle.

4.10 Consulting templates

T4.1 AI Intake Form (front door)

Purpose: Capture every candidate quickly and consistently. **Use when:** every new AI candidate, whatever its source

Requester / unit / date: .....

Source (request / discovery / shadow AI / vendor / embedded feature): .....

Problem in one sentence: .....

Users and people affected: .....

Intended use — what decision or action does the output feed? .....

Autonomy level (0–5): .....

Data involved (incl. personal data?): .....

Existing tool or vendor involved: .....

Sponsor and business owner: .....

Urgency or deadline: .....

T4.2 Intake Tracking Log

Purpose: One row per item, from intake to decision.

ID	Item	Source	Risk class	Stage	Owner	Disposition	Reason	Trigger / next date	Duplicate of
AI-031	*HR candidate ranking (embedded)*	*Embedded*	*High*	*Decided*	*CHRO*	*Switch off pending RA*	*High-risk, unassessed*	*RA done Q2*	*—*

T4.3 Risk Classification Record

Purpose: Classify regulatory and AI risk exposure before value is weighed.

Item ID / name: .....

Intended purpose and context of use: .....

Framework	Classification	Basis / notes	Specialist consulted
EU AI Act tier			
NIST AI RMF — Map: context & impacts			
Sector rules / GDPR (DPIA trigger?)			
Internal risk taxonomy & appetite			

Portfolio risk class (Stop / High / Limited / Minimal): .....

Compliance effort implications: .....

Classified by / date: .....

T4.4 Two-Axis Portfolio Scoring Sheet

Purpose: Score the full set on one fixed criteria set.

Criteria charter reference (version / date): .....

ID	Item	Risk class	Expected value (range)	Value confidence	Expected effort (range)	Main effort component	Dependencies	Consolidation group

T4.5 Portfolio Balance & Consolidation Check

Purpose: Confirm the portfolio is balanced, within capacity and free of duplicates.

BALANCE

Check	Status	Notes / action
At least one quick win; bets staged		
Value types mixed		
High-risk items ≤ governance capacity		
Capacity by component ≤ ~85%		
Vendor / platform / team concentration		
Unit spread intended and explained		

CONSOLIDATION

Capability	Items requesting it	Real differences?	Pattern (merge / shared / lead-unit / use existing / separate)	Re-scored?

T4.6 Investment Decision Package

Purpose: Route the portfolio to the funding body with real options.

Decision forum / decision-holder / date: .....

Recommendation (one paragraph): .....

Option A — full portfolio (cost, value, risk): .....

Option B — core portfolio: .....

Option C — not this cycle / enablers only: .....

Resource request by component: .....

Capacity evidence: .....

Aggregate risk profile: .....

Funding stages and gate criteria: .....

Not funded — items, reasons, triggers: .....

4.11 Best practices and pitfalls

✓ DO THIS

- Operate one light, tracked front door for all AI candidates.
- Include shadow AI, vendor proposals and embedded features.
- Classify risk before anyone weighs value.
- Score the full set against one fixed criteria set.
- Balance by value timing, risk and capacity per component.
- Consolidate duplicates where needs genuinely overlap.
- Offer options, staged funding and a real “not this cycle”.

× AVOID THIS

- Separate intake routes per unit or vendor.
- Letting embedded SaaS AI bypass review.
- Scoring value first and “checking risk later”.
- Inventing criteria for one favoured item.
- Funding the top ten without a portfolio view.
- Consolidating by decree.
- Lowering the bar to show progress.

4.12 Module 4 cheat sheet

MODULE 4 · ONE-PAGE SUMMARY

Building the Prioritized Opportunity Portfolio

LIST VS PORTFOLIO

Portfolio = selected together: aggregate risk, capacity, dependencies, no duplicates, options, accountable owners.

TRACKED INTAKE

One front door · light form · ID + status · owner · SLAs · all sources (requests, discovery, shadow, vendors, embedded) · decision history.

RISK BEFORE VALUE

Frameworks: **EU AI Act** tiers (prohibited · high-risk · transparency · minimal) · **NIST AI RMF** (Govern, Map, Measure, Manage) · **ISO/IEC 23894** (AI risk guidance) · **ISO/IEC 42001** (AI management system) · sector rules, GDPR. Why first: stop prohibited uses early · risk drives effort · avoid anchoring on value.

TWO-AXIS SCORING

Expected value and expected effort · separate · probability-weighted, error-priced · one fixed criteria set · full set scored · risk class carried alongside.

BALANCE

Value timing & type · risk within governance capacity · effort ≤ ~85% capacity **per component** · concentration · unit spread.

CONSOLIDATE

Detect at capability level → test real differences → merge / shared platform / lead unit / use existing / keep separate → re-score → communicate.

DECISION PACKAGE

Recommendation · options A/B/C · portfolio view · resources by component · capacity · risk · staged funding · not funded · decision required.

"NOT THIS CYCLE"

Correct when value threshold not met, foundations missing, capacity committed, risk over appetite, or better evidence imminent. Record reasons + triggers; keep items in backlog. Never lower the bar to show progress.

4.13 Knowledge check

Answer all questions before turning the page.

K1 **What most distinguishes a portfolio from a list of the ten best-scoring ideas?**

- (A) A portfolio only includes low-risk items.
- (B) A portfolio contains more items.
- (C) A portfolio is approved by IT.
- (D) A portfolio is selected as a set, considering aggregate risk, capacity, dependencies and duplication.

K2 **An HR software update switches on an AI candidate-ranking feature. How should the portfolio process handle it?**

- (A) Ignore it — it is part of licensed software.
- (B) Bring it through the tracked intake and classify its risk like any other candidate.
- (C) Add it to next year's budget.
- (D) Let HR decide informally.

K3 **Why should risk classification come before value is weighed?**

- (A) To stop prohibited uses early, build compliance burden into effort, and avoid anchoring on attractive value numbers.
- (B) Because high-risk uses always have low value.
- (C) Because value cannot be estimated for risky uses.
- (D) Because regulators require scoring to be done by risk teams.

K4 **Which framework is organized around the functions Govern, Map, Measure and Manage?**

- (A) EU AI Act
- (B) GDPR
- (C) NIST AI Risk Management Framework
- (D) ISO/IEC 42001

K5 **A sponsor wants to add "innovation visibility" as a criterion that only their use case scores well on. What is the correct response?**

- (A) Add it but give it a small weight.
- (B) Consider it as a change for all candidates through a recorded decision and re-scoring — or not at all.
- (C) Ignore the request silently.
- (D) Add it for that use case only.

K6 **The proposed portfolio fits the total budget but needs three times the available data-engineering capacity. What does this show?**

- (A) Balance must be checked by effort component, not just total budget.
- (B) Data engineering should be removed from estimates.
- (C) The portfolio is balanced because budget fits.
- (D) All data-heavy items should be dropped.

K7 **Seven divisions request “a document assistant”, with different content but the same AI task. Which consolidation pattern fits best?**

(A) Force one division’s configuration on all
(B) Reject all seven
(C) Seven separate projects
(D) A shared platform with separate divisional configurations or content

K8 **No candidate clears the value threshold after error pricing, and data foundations are weak. The CEO wants to “fund something to show progress”. What should the decision package recommend?**

(A) Delay the decision indefinitely.
(B) Fund the highest-scoring candidate anyway.
(C) “Not this cycle” for use cases, with investment in data enablers and recorded revisit triggers.
(D) Fund all candidates at a reduced level.

4.14 Knowledge check: answers and explanations

K1 **Answer: D** Review: 4.1

What most distinguishes a portfolio from a list of the ten best-scoring ideas?

- ✗ (A) A portfolio only includes low-risk items. — **Wrong.** Portfolios can include high-risk items within appetite and governance capacity.
- ✗ (B) A portfolio contains more items. — **Wrong.** Size is not the distinction.
- ✗ (C) A portfolio is approved by IT. — **Wrong.** Approval body is not the defining feature.
- ✓ (D) A portfolio is selected as a set, considering aggregate risk, capacity, dependencies and duplication. — **Correct.** Portfolio thinking evaluates the combination, not just the parts.

K2 **Answer: B** Review: 4.2

An HR software update switches on an AI candidate-ranking feature. How should the portfolio process handle it?

- ✗ (A) Ignore it — it is part of licensed software. — **Wrong.** Embedded AI can introduce high-risk uses without review.
- ✓ (B) Bring it through the tracked intake and classify its risk like any other candidate. — **Correct.** All visible AI candidates, including embedded features, belong in intake.
- ✗ (C) Add it to next year's budget. — **Wrong.** The feature is live now; it needs immediate classification.
- ✗ (D) Let HR decide informally. — **Wrong.** Informal decisions bypass risk classification and tracking.

K3 **Answer: A** Review: 4.3

Why should risk classification come before value is weighed?

- ✓ (A) To stop prohibited uses early, build compliance burden into effort, and avoid anchoring on attractive value numbers. — **Correct.** These are the three reasons for risk-first classification.
- ✗ (B) Because high-risk uses always have low value. — **Wrong.** High-risk uses can be very valuable.
- ✗ (C) Because value cannot be estimated for risky uses. — **Wrong.** Value can be estimated; the issue is sequence and anchoring.
- ✗ (D) Because regulators require scoring to be done by risk teams. — **Wrong.** No such general requirement exists.

K4 **Answer: C** Review: 4.3

Which framework is organized around the functions Govern, Map, Measure and Manage?

- ✗ (A) EU AI Act — **Wrong.** The Act is a regulation with risk tiers, not these four functions.
- ✗ (B) GDPR — **Wrong.** GDPR is data protection law.
- ✓ (C) NIST AI Risk Management Framework — **Correct.** AI RMF 1.0 is structured around these four functions.
- ✗ (D) ISO/IEC 42001 — **Wrong.** 42001 is a management-system standard with clauses and controls.

K5 **Answer: B** Review: 4.4

A sponsor wants to add "innovation visibility" as a criterion that only their use case scores well on. What is the correct response?

- ✗ (A) Add it but give it a small weight. — **Wrong.** Even a small, bespoke criterion biases the result.
- ✓ (B) Consider it as a change for all candidates through a recorded decision and re-scoring — or not at all. — **Correct.** One fixed set of criteria applies to the full set.
- ✗ (C) Ignore the request silently. — **Wrong.** Requests should be handled transparently.
- ✗ (D) Add it for that use case only. — **Wrong.** This breaks comparability across the set.

K6 **Answer: A** Review: 4.5

The proposed portfolio fits the total budget but needs three times the available data-engineering capacity. What does this show?

✓ **A** Balance must be checked by effort component, not just total budget. — **Correct.** Capacity per component determines whether the portfolio can be delivered.

✗ B Data engineering should be removed from estimates. — **Wrong.** This hides real effort.

✗ C The portfolio is balanced because budget fits. — **Wrong.** Budget fit hides component capacity constraints.

✗ D All data-heavy items should be dropped. — **Wrong.** Re-sequencing or adding capacity may be better than dropping.

K7 **Answer: D** Review: 4.6

Seven divisions request "a document assistant", with different content but the same AI task. Which consolidation pattern fits best?

✗ A Force one division's configuration on all — **Wrong.** This ignores genuine content differences.

✗ B Reject all seven — **Wrong.** There is real demand and value.

✗ C Seven separate projects — **Wrong.** This duplicates cost, controls and effort.

✓ **D** A shared platform with separate divisional configurations or content — **Correct.** Same capability, different content — one platform, configured per unit.

K8 **Answer: C** Review: 4.8

No candidate clears the value threshold after error pricing, and data foundations are weak. The CEO wants to "fund something to show progress". What should the decision package recommend?

✗ A Delay the decision indefinitely. — **Wrong.** Indefinite delay is not a decision; "not this cycle" with triggers is.

✗ B Fund the highest-scoring candidate anyway. — **Wrong.** This lowers the bar and undermines the process.

✓ **C** "Not this cycle" for use cases, with investment in data enablers and recorded revisit triggers. — **Correct.** This is a correct, legitimate portfolio outcome.

✗ D Fund all candidates at a reduced level. — **Wrong.** Spreading thin funds weak candidates without fixing foundations.

R1 Course at a glance: the master cheat sheet

ALL FOUR MODULES ON ONE PAGE

Identify and Prioritize AI Opportunities

M1 · FINDING OPPORTUNITIES

Work-pull: map workflows (process → workflow → task), find friction (searching, re-keying, drafting, triage, pattern judgment; waiting = usually not AI). Scan by function and industry; autonomy ladder 0–5 (assistants 1–2, agents 3–5). Value hypothesis with mechanism. Inventory from **solicited · observed · standing audit**. Value and effort separate and arguable. **Ledger**: verify · correct · escalate · wait · new run → moved cost, not saving.

M2 · QUALIFYING USE CASES

Fast, structured screen. **Normalize** (user, trigger, input, one-verb AI task, output, decision, human role, volume, metric, location). Seven separable dimensions. **Data readiness = gate**. **Label test**: target, where stored, coverage, consistency, timing/leakage, selection bias. **Risk by use**, not technique. Dispositions: advance · conditions · park · redirect · reject. Good first pilots.

M3 · VALUE-EFFORT LENS

Criteria and weights before scoring; anchored value. **Effort by component**: data · integration · compliance · adoption · operational capacity. **Price direction of error**: FP vs FN costs, capacity to act; optimism haircuts, downside case. Matrix: quick wins · strategic bets · fill-ins · money pits, with ranges and risk. Dispositions: fund now · prepare · bundle · defer · drop. Sensitivity, trade-offs, sequencing.

M4 · PORTFOLIO

Portfolio ≠ list. **Tracked intake** (all sources incl. shadow, vendor, embedded). **Risk before value** (EU AI Act tiers, NIST AI RMF, ISO/IEC 23894, 42001). **Two axes, one fixed criteria set**, full set. **Balance** value timing, risk, capacity per component. **Consolidate** duplicates. **Decision package** with options and staged funding. "**Not this cycle**" is correct when warranted — recorded with triggers.

R2 Template index

Code	Template	Use when
T1.1	Workflow Friction Map	Discovery observation
T1.2	Opportunity Card	Describing a candidate
T1.3	Opportunity Inventory Register	Keeping candidates comparable
T1.4	Value Estimate & Effort Ledger	Sizing net value
T2.1	Use-Case Normalization Sheet	Before screening
T2.2	Qualification Screen Scorecard	Screening
T2.3	Data Readiness Gate	Before scoring dimensions
T2.4	Outcome-Recorded (Label) Test	Predictive use cases
T2.5	Use-Based Risk Screen	Screening risk
T3.1	Criteria & Weights Charter	Before scoring
T3.2	Component Effort Estimator	Effort estimation
T3.3	Error-Direction Value Adjuster	Pricing error

Code	Template	Use when
T3.4	Shortlist Disposition Register	Reading the matrix
T4.1	AI Intake Form	Every new candidate
T4.2	Intake Tracking Log	Portfolio office
T4.3	Risk Classification Record	Before value scoring
T4.4	Two-Axis Portfolio Scoring Sheet	Scoring the full set
T4.5	Portfolio Balance & Consolidation Check	Assembling the portfolio
T4.6	Investment Decision Package	Funding decision

R3 Glossary

A

Adoption resistance
Effort component covering training, role change, trust building and stakeholder engagement needed for people to use an AI output.

Agentic AI
AI that plans and executes multi-step tasks across tools with some autonomy (autonomy levels 3–5).

AI assistant / copilot
AI that answers, suggests or drafts inside a person's workflow while the person decides and acts (levels 1–2).

AI intake process
A single, tracked route through which every AI candidate enters portfolio consideration.

Anchors
Observable descriptions of what each score level means for a criterion.

Autonomy ladder
Levels 0–5 describing how far AI acts without human decision: none, suggest, draft, act with approval, act with human on exception, autonomous.

B

Bundle (disposition)
Deliver a candidate together with related candidates that share components or create its capacity.

C

Calibration session
Meeting in which scorers discuss divergent scores on the evidence to align their use of anchors.

Capacity release
Time freed by AI; becomes value only if removed as cost or redirected to named higher-value work.

Compliance burden
Effort component covering risk assessment, legal review, documentation, oversight design and audits.

Consolidation
Resolving duplicate automation by merging, sharing a platform, leading with one unit, using an existing tool or documenting separation.

D

Data readiness gate
Pass/conditional/fail test (exists, accessible, permitted, sufficient, representative, owned) applied before other dimensions are scored.

Decision package
Document routing a portfolio to the funding body: recommendation, options, resources, capacity, risk, stages, not-funded items.

Defer (disposition)
Not this cycle; revisit on a stated trigger.

Direction of error
The asymmetry in costs of false positives vs false negatives, and the predictable optimism of early estimates.

Disposition
The recorded outcome for a candidate with its reason, e.g., advance, park, redirect, reject, fund now, prepare, bundle, defer, drop.

Duplicate automation
Several initiatives automating the same capability separately.

E

Effort by component
Estimating effort separately for data readiness, integration, compliance burden, adoption resistance and operational capacity.

Elapsed-time friction
Delay where work waits (queues, approvals); usually solved by process redesign rather than AI.

EU AI Act
Regulation (EU) 2024/1689 establishing risk-based obligations by intended purpose: prohibited practices, high-risk systems, transparency obligations, minimal risk.

Expected effort
Effort estimate adjusted for uncertainty and contingency, by component.

Expected value
Value estimate weighted for probability, priced for error direction and haircut for evidence strength.

F

False negative
A case the system should have flagged but did not.

False positive
A case the system flagged that did not warrant it.

Fill-in
Low-value, low-effort candidate; do only with spare capacity or as part of a bundle.

Friction
Any point where work is slower, costlier, riskier or more frustrating than necessary.

Front door
The single entry point of the AI intake process.

H

Haircut
Percentage discount on estimated value according to the strength of evidence.

HiPPO
Highest-paid person's opinion; pressure to override criteria by seniority.

I

ISO/IEC 23894
International guidance on AI risk management, building on ISO 31000.

ISO/IEC 42001
International standard for AI management systems.

K

Kill criteria
Pre-published conditions under which a candidate is rejected.

L

Label test
Checking whether the outcome a model would predict was recorded — coverage, consistency, timing and selection bias.

Leakage
Using input information that would not be available at prediction time, inflating test performance.

Ledger
Record of who verifies, corrects, escalates or waits after an AI change, and of new run work.

M

Money pit
Low-value, high-effort candidate; drop or redesign.

Moved cost
Effort relocated by AI to another person, team or customer; recorded as a cost, not a saving.

N

Net value
Gross value minus ledger (moved) costs and new run costs.

NIST AI RMF
NIST AI Risk Management Framework 1.0, structured around Govern, Map, Measure and Manage.

Normalization
Rewriting a proposal into a standard structure: user, trigger, input, AI task, output, decision fed, human role, volume, metric, location.

Not this cycle
A legitimate portfolio decision to fund no (or only enabling) AI use cases now, recorded with reasons and triggers.

O

Observed use
Inventory channel based on actual behaviour: shadow AI, workarounds, logs, usage data.

Operational capacity
Effort component: the organization's ability to absorb AI workload — reviews, exceptions, monitoring, support.

Opportunity inventory
List of AI candidates described comparably, with source channel, value hypothesis, effort and owner.

Optimism bias
Systematic tendency to overestimate benefits and underestimate effort and time.

P

Park (disposition)
Hold a candidate blocked by an external factor, with a revisit trigger.

Portfolio
A set of AI initiatives selected together, balanced for value, risk and effort within capacity.

Post-hoc weighting
Setting or changing weights after seeing scores; undermines defensibility.

Prepare (disposition)
Invest in enablers before funding the use case itself.

Q

Quick win
High-value, low-effort candidate; fund now after checking effort components.

R

Redirect (disposition)
Send a real problem to a better non-AI fix or another owner.

Reference class
Comparable past initiatives used to sanity-check estimates.

Risk by use
Sizing risk by intended purpose and context — who is affected, consequence, reversibility, autonomy — rather than technique.

S

Selection bias
Outcomes recorded only for cases chosen by past decisions, misrepresenting the full population.

Sensitivity test
Checking whether rankings change when weights or estimates move within plausible ranges.

Separable dimensions
Qualification criteria judged independently: value, workflow fit, data readiness, feasibility, risk, sponsorship, measurement clarity.

Shortlist
Ranked candidates with stated dispositions after the value-effort reading.

Solicited input
Inventory channel based on workshops, idea calls and interviews.

Staged funding
Releasing funds in tranches tied to gate evidence.

Standing audit
Inventory channel: scheduled, systematic review of functions and processes for opportunities.

Strategic bet
High-value, high-effort candidate; prepare and fund in stages.

Swimlane diagram
Workflow map showing steps by role, with hand-offs and wait states.

T

Technology-push
Discovery that starts from a technology and looks for uses; contrasted with work-pull.

V

Value hypothesis
Statement linking an AI task change to a metric change and to value, with a mechanism.

W

Work-pull
Discovery that starts from how work is done and where friction lies.

R4 Mock exam

Format: 36 multiple-choice questions · one best answer each · **Time:** 60 minutes · **Suggested pass mark:** 70% (25 correct)

Questions 1–8 cover Module 1, 9–16 Module 2, 17–24 Module 3, 25–32 Module 4 and 33–36 are integrated scenarios. Answer every question before checking the answer key, which explains why each option is right or wrong.

Q1 **Which approach best reflects work-pull discovery?**

- A Copying a competitor's announced AI projects
- B Listing the capabilities of the newest models and matching them to departments
- C Observing how claims are processed, measuring where time goes, and asking which fix would remove the biggest friction
- D Asking vendors for their top use cases

Q2 **Which attribute of a workflow step best distinguishes friction AI can reduce from friction it usually cannot?**

- A Number of systems involved
- B The seniority of the people involved
- C The department's budget
- D Handling time versus waiting time

Q3 **Agents copy data from emailed PDF forms into the CRM. Which AI pattern fits, and what non-AI fix should be checked first?**

- A Generative drafting; templates
- B Document information extraction; structured web forms at source
- C Predictive ML; better reporting
- D Agentic AI; more staff

Q4 **An assistant drafts replies that agents edit and send. Which autonomy level is this?**

- A Level 2 — draft
- B Level 1 — suggest
- C Level 5 — autonomous
- D Level 4 — act, human on exception

Q5 **Which channel is designed to provide systematic coverage independent of who speaks up?**

- A Solicited input
- B Observed use
- C Vendor briefings
- D Standing audit

Q6 **An AI tool saves clerks 10 minutes per case; supervisors now spend 4 minutes checking each case. How should value be stated?**

(A) 14 minutes saved because quality improves

(B) 10 minutes reduced, 4 minutes moved to supervisors: about 6 minutes net — noting supervisors' time may cost more

(C) 10 minutes saved per case

(D) No value, since effort moved

Q7 **Which is the most arguable (challengeable) value estimate?**

(A) "€500k, based on the vendor's benchmark."

(B) "Strategic — not quantifiable."

(C) "140k cases × 6 min × 60% addressable × 50% reduction × €0.80/min ≈ €200k (range €130–260k)."

(D) "High value — everyone agrees."

Q8 **A team claims "capacity release of 8,000 hours" as cost reduction. What must be shown?**

(A) That cost actually falls (fewer hires, contractors, overtime) or that time is redirected to named value

(B) That the hours are recorded in timesheets

(C) That the model is accurate

(D) That staff like the tool

Q9 **Which set of fields best normalizes a use case?**

(A) Model type, parameters, GPU needs

(B) User, trigger, input, AI task, output, decision fed, human role, volume, metric, process location

(C) Vendor, price, contract length

(D) Department, budget, headcount

Q10 **Why is data readiness treated as a gate rather than a weight?**

(A) Because high scores elsewhere cannot compensate for data that does not exist or cannot be used.

(B) Because data is the most important dimension.

(C) Because regulations require it.

(D) Because data cannot be scored.

Q11 **A retailer wants to predict which returns are fraudulent, but fraud is recorded only for returns staff chose to inspect. What is the issue and a remedy?**

(A) Inconsistency; standardise codes

(B) Leakage; remove late fields

(C) Selection bias; create unbiased labels, e.g., random-sample audits

(D) Insufficient volume; wait longer

Q12 **For a generative AI use case drafting regulatory responses, what is the equivalent of the label test?**

(A) Checking the vendor's benchmark

(B) Checking GPU availability

(C) Checking that the model is large enough

(D) Checking that examples of "good" — approved past responses or reference answers — exist for evaluation

Q13 **Which use is highest risk when sized by use?**

- A) An LLM drafting product descriptions reviewed by merchandisers
- B) An LLM summarising internal project updates
- C) An LLM-based system scoring applicants' eligibility for social benefits
- D) An LLM translating internal manuals

Q14 **Which disposition fits a strong use case blocked until a new CRM captures the needed field next year?**

- A) Redirect to a non-AI fix
- B) Park with a revisit trigger tied to the CRM release
- C) Advance unconditionally
- D) Reject

Q15 **What is the main reason to publish kill criteria before screening?**

- A) To protect the process from pressure to advance favourites, by making rejection reasons known in advance
- B) To reduce the number of proposals
- C) To satisfy auditors only
- D) To reduce paperwork

Q16 **Which is the weakest choice for a first pilot?**

- A) Meeting-summary assistant for one team
- B) Product-description drafting with review
- C) Internal knowledge retrieval with clear time metrics
- D) A fully autonomous agent granting credit-limit increases to customers

Q17 **A steering group sets weights after seeing that its preferred use case ranks fourth. What is the problem?**

- A) Post-hoc weighting undermines defensibility
- B) The scores are too low
- C) The matrix has the wrong axes
- D) Too few criteria

Q18 **Which list correctly names the five effort components used in this course?**

- A) People, process, technology, data, culture
- B) Budget, timeline, vendor, model, licence
- C) Data readiness, integration, compliance burden, adoption resistance, operational capacity
- D) Value, risk, fit, feasibility, sponsorship

Q19 **A document-review AI would route 25% of documents to legal experts who are already overloaded. Which component captures this?**

- A) Data readiness
- B) Integration
- C) Adoption resistance
- D) Operational capacity to absorb new AI workload

Q20 **Missing a defect costs €20,000 (recall, warranty); a false alarm costs €50 of inspection. How should a configuration be chosen?**

(A) Maximise overall accuracy
(B) Price both error types and choose the configuration with the highest expected value, checking inspection capacity
(C) Minimise false alarms
(D) Choose the vendor default

Q21 **Which practice best corrects for optimism bias in value estimates?**

(A) Removing ranges for clarity
(B) Using the most enthusiastic sponsor's numbers
(C) Doubling effort for every use case
(D) Applying evidence-based haircuts and showing a downside scenario

Q22 **A candidate is low value and high effort after error pricing. Which disposition fits?**

(A) Fund now
(B) Drop or redesign
(C) Bundle
(D) Prepare

Q23 **What does a sensitivity test add to a prioritization?**

(A) A final decision
(B) A vendor shortlist
(C) Evidence of whether the ranking is robust to plausible changes in weights or estimates
(D) More criteria

Q24 **Which sequencing principle applies when a strategic bet depends on cleaned master data?**

(A) Dependencies first: fund the master-data enabler before the use case
(B) Reuse
(C) Random order
(D) Value early

Q25 **Which belongs in a tracked AI intake?**

(A) Only centrally approved projects
(B) Only GenAI projects
(C) Only projects over €100k
(D) Business requests, discovery findings, shadow AI found, vendor proposals and embedded SaaS features

Q26 **A candidate may involve emotion recognition of employees at work. What should happen at intake?**

(A) Score value first to see if it is worth it
(B) Classify as Stop and refer to legal before any further analysis
(C) Pilot it with volunteers
(D) Let the works council decide alone

Q27 **Which statement about the EU AI Act is accurate for portfolio risk classification?**

(A) It classifies by intended purpose, with tiers such as prohibited, high-risk, transparency obligations and minimal risk.

(B) Risk depends mainly on the algorithm used.

(C) It applies only to companies outside the EU.

(D) It replaces GDPR.

Q28 **What does "one fixed set of scoring criteria" prevent?**

(A) Fast scoring

(B) Risk classification

(C) Bespoke criteria that favour particular candidates and break comparability

(D) Consolidation

Q29 **Which portfolio is best balanced?**

(A) Six high-risk strategic bets using 160% of capacity

(B) One huge transformation programme

(C) Ten fill-ins

(D) Two quick wins, one staged strategic bet, one shared platform; two high-risk items within governance capacity; ~85% of capacity

Q30 **Two vendors propose overlapping extraction tools for finance and procurement, with similar documents and risk. What should the advisor recommend?**

(A) Consolidate onto one solution after confirming there are no material differences, then re-score

(B) Let each unit decide independently

(C) Reject both

(D) Buy both to keep units happy

Q31 **What should a decision package include to allow a real decision?**

(A) A technical architecture

(B) Options (including "not this cycle"), resources by component, capacity evidence, risk profile, staged funding and the items not funded

(C) A single list to approve

(D) Only the vendor quotes

Q32 **When is "not this cycle" the correct portfolio outcome?**

(A) Never — some AI should always be funded

(B) Only if the budget is zero

(C) When no candidate clears the threshold, foundations are missing, capacity is committed or risk exceeds appetite — recorded with triggers

(D) Whenever the CEO is uncertain

Q33 **Integrated: A business unit submits "AI for HR". In the correct order, what happens?**

(A) Log at intake → classify risk → normalize and screen (data gate, label test) → score value and effort → portfolio decision

(B) Pilot → measure → normalize

(C) Consolidate → reject

(D) Score value → pick vendor → classify risk

Q34 **Integrated: An attractive GenAI drafting case shows €900k gross value. Review adds 40% of the saved time back, and the reviewers are scarce senior staff. What happens to its position?**

(A) Nothing — gross value is what counts

(B) Risk class changes to prohibited

(C) It becomes a money pit automatically

(D) Net value falls (ledger) and operational capacity effort rises; it may move from quick win toward strategic bet or defer

Q35 **Integrated: Three divisions each rank "contract clause extraction" as their top priority. What is the portfolio move?**

(A) Pick one division at random

(B) Fund three separate projects

(C) Consolidate into one initiative, aggregate value, share effort and design controls to the highest risk class, then re-score

(D) Reject because of duplication

Q36 **Integrated: A predictive use case has strong value, an eager sponsor and data that exists — but the outcome is recorded only for 5% of cases selected by current rules. Where should it land?**

(A) Add it to the portfolio with high effort

(B) Advance with conditions or park: create representative labels first; it cannot be valued reliably until then

(C) Drop permanently

(D) Fund now as a quick win

R5 Mock exam: answer key and explanations

Quick answer grid

1 · C	**2** · D	**3** · B	**4** · A	**5** · D	**6** · B	**7** · C	**8** · A	**9** · B	**10** · A
11 · C	**12** · D	**13** · C	**14** · B	**15** · A	**16** · D	**17** · A	**18** · C	**19** · D	**20** · B
21 · D	**22** · B	**23** · C	**24** · A	**25** · D	**26** · B	**27** · A	**28** · C	**29** · D	**30** · A
31 · B	**32** · C	**33** · A	**34** · D	**35** · C	**36** · B				

Score	Interpretation	Next step
31–36	Exam-ready	Review missed items; apply the templates to a live portfolio
25–30	Pass level	Re-read cheat sheets for modules where you missed two or more
below 25	Not yet	Re-study weak modules using the “Review” references; retake in one week

Full explanations

Q1 **Answer: C** Review: M1 · 1.1

Which approach best reflects work-pull discovery?

- ✗ (A) Copying a competitor’s announced AI projects — **Wrong.** This imports others’ choices without understanding one’s own work.
- ✗ (B) Listing the capabilities of the newest models and matching them to departments — **Wrong.** This is technology-push.
- ✓ (C) Observing how claims are processed, measuring where time goes, and asking which fix would remove the biggest friction — **Correct.** Work-pull starts from the work and its friction.
- ✗ (D) Asking vendors for their top use cases — **Wrong.** Vendor lists are technology-push and sales-driven.

Q2 **Answer: D** Review: M1 · 1.2–1.3

Which attribute of a workflow step best distinguishes friction AI can reduce from friction it usually cannot?

- ✗ (A) Number of systems involved — **Wrong.** Relevant to integration, but not the key distinction.
- ✗ (B) The seniority of the people involved — **Wrong.** Seniority does not determine friction type.
- ✗ (C) The department’s budget — **Wrong.** Budget is irrelevant to friction type.
- ✓ (D) Handling time versus waiting time — **Correct.** AI mainly reduces handling effort; waiting usually needs process redesign.

Q3 **Answer: B** Review: M1 · 1.3

Agents copy data from emailed PDF forms into the CRM. Which AI pattern fits, and what non-AI fix should be checked first?

- ✗ (A) Generative drafting; templates — **Wrong.** Drafting addresses writing, not data capture.
- ✓ (B) Document information extraction; structured web forms at source — **Correct.** Re-keying friction maps to extraction; collecting data structured at source may remove it entirely.
- ✗ (C) Predictive ML; better reporting — **Wrong.** Prediction does not address re-keying.
- ✗ (D) Agentic AI; more staff — **Wrong.** Over-engineered; staffing is not a fix for friction.

Q4 **Answer: A** Review: M1 · 1.5

An assistant drafts replies that agents edit and send. Which autonomy level is this?

✓ A Level 2 — draft — **Correct.** AI drafts; the human edits and sends.

✗ B Level 1 — suggest — **Wrong.** It produces a full draft, not just a suggestion.

✗ C Level 5 — autonomous — **Wrong.** Nothing is sent without a human.

✗ D Level 4 — act, human on exception — **Wrong.** The human handles every case, not only exceptions.

Q5 **Answer: D** Review: M1 · 1.7

Which channel is designed to provide systematic coverage independent of who speaks up?

✗ A Solicited input — **Wrong.** This depends on who participates and speaks.

✗ B Observed use — **Wrong.** This depends on existing behaviour and workarounds.

✗ C Vendor briefings — **Wrong.** These depend on vendor offerings.

✓ D Standing audit — **Correct.** A scheduled, systematic review of all functions provides coverage.

Q6 **Answer: B** Review: M1 · 1.9

An AI tool saves clerks 10 minutes per case; supervisors now spend 4 minutes checking each case. How should value be stated?

✗ A 14 minutes saved because quality improves — **Wrong.** Adding time is incorrect; quality needs separate evidence.

✓ B 10 minutes reduced, 4 minutes moved to supervisors: about 6 minutes net — noting supervisors' time may cost more — **Correct.** The ledger records moved costs, which often fall on more expensive people.

✗ C 10 minutes saved per case — **Wrong.** This ignores the moved cost.

✗ D No value, since effort moved — **Wrong.** Some effort is genuinely saved.

Q7 **Answer: C** Review: M1 · 1.8

Which is the most arguable (challengeable) value estimate?

✗ A "€500k, based on the vendor's benchmark." — **Wrong.** Single-source and not decomposed.

✗ B "Strategic — not quantifiable." — **Wrong.** This avoids estimation.

✓ C "140k cases × 6 min × 60% addressable × 50% reduction × €0.80/min ≈ €200k (range €130–260k)." — **Correct.** Each factor and its source can be challenged separately.

✗ D "High value — everyone agrees." — **Wrong.** This cannot be challenged on specifics.

Q8 **Answer: A** Review: M1 · 1.6

A team claims "capacity release of 8,000 hours" as cost reduction. What must be shown?

✓ A That cost actually falls (fewer hires, contractors, overtime) or that time is redirected to named value — **Correct.** Capacity is value only when removed or redeployed.

✗ B That the hours are recorded in timesheets — **Wrong.** Recording does not make them cash.

✗ C That the model is accurate — **Wrong.** Accuracy does not convert time into money.

✗ D That staff like the tool — **Wrong.** Satisfaction does not prove cost reduction.

Q9 **Answer: B** Review: M2 · 2.2

Which set of fields best normalizes a use case?

✗ (A) Model type, parameters, GPU needs — **Wrong.** These are build details.

✓ (B) User, trigger, input, AI task, output, decision fed, human role, volume, metric, process location — **Correct.** These make use cases defined and comparable.

✗ (C) Vendor, price, contract length — **Wrong.** These are procurement attributes.

✗ (D) Department, budget, headcount — **Wrong.** These describe the organization, not the use case.

Q10 **Answer: A** Review: M2 · 2.4

Why is data readiness treated as a gate rather than a weight?

✓ (A) Because high scores elsewhere cannot compensate for data that does not exist or cannot be used. — **Correct.** A gate prevents value from masking an impossibility.

✗ (B) Because data is the most important dimension. — **Wrong.** Importance is not the logic; compensability is.

✗ (C) Because regulations require it. — **Wrong.** The reason is logical, not regulatory.

✗ (D) Because data cannot be scored. — **Wrong.** Data readiness can be assessed; it is applied as pass/conditional/fail.

Q11 **Answer: C** Review: M2 · 2.5

A retailer wants to predict which returns are fraudulent, but fraud is recorded only for returns staff chose to inspect. What is the issue and a remedy?

✗ (A) Inconsistency; standardise codes — **Wrong.** Not the stated problem.

✗ (B) Leakage; remove late fields — **Wrong.** Leakage concerns timing of inputs.

✓ (C) Selection bias; create unbiased labels, e.g., random-sample audits — **Correct.** Labels reflect past selection; random audits create a representative sample.

✗ (D) Insufficient volume; wait longer — **Wrong.** Volume does not cure selection bias.

Q12 **Answer: D** Review: M2 · 2.5

For a generative AI use case drafting regulatory responses, what is the equivalent of the label test?

✗ (A) Checking the vendor's benchmark — **Wrong.** Benchmarks do not show the organization's own standard of good.

✗ (B) Checking GPU availability — **Wrong.** This is infrastructure.

✗ (C) Checking that the model is large enough — **Wrong.** Size is not the question.

✓ (D) Checking that examples of "good" — approved past responses or reference answers — exist for evaluation — **Correct.** Recorded examples of quality make feasibility and measurement possible.

Q13 **Answer: C** Review: M2 · 2.6

Which use is highest risk when sized by use?

✗ (A) An LLM drafting product descriptions reviewed by merchandisers — **Wrong.** Customer-facing but reviewed and low consequence.

✗ (B) An LLM summarising internal project updates — **Wrong.** Internal, low consequence, easily corrected.

✓ (C) An LLM-based system scoring applicants' eligibility for social benefits — **Correct.** It affects access to essential services, with high consequence and regulatory exposure.

✗ (D) An LLM translating internal manuals — **Wrong.** Low consequence.

Q14 **Answer: B** Review: M2 · 2.7

Which disposition fits a strong use case blocked until a new CRM captures the needed field next year?

✗ (A) Redirect to a non-AI fix — **Wrong.** There may be no non-AI substitute; the need remains.

✓ (B) Park with a revisit trigger tied to the CRM release — **Correct.** Parking with a trigger preserves the candidate.

✗ (C) Advance unconditionally — **Wrong.** It cannot proceed without the data.

✗ (D) Reject — **Wrong.** The block is temporary and external.

Q15 **Answer: A** Review: M2 · 2.7

What is the main reason to publish kill criteria before screening?

✓ (A) To protect the process from pressure to advance favourites, by making rejection reasons known in advance — **Correct.** Pre-published criteria make dispositions defensible.

✗ (B) To reduce the number of proposals — **Wrong.** Fewer proposals is not the aim.

✗ (C) To satisfy auditors only — **Wrong.** Auditors may benefit, but that is not the main purpose.

✗ (D) To reduce paperwork — **Wrong.** Not the main reason.

Q16 **Answer: D** Review: M2 · 2.8

Which is the weakest choice for a first pilot?

✗ (A) Meeting-summary assistant for one team — **Wrong.** Low risk; easy to measure.

✗ (B) Product-description drafting with review — **Wrong.** A reasonable first pilot.

✗ (C) Internal knowledge retrieval with clear time metrics — **Wrong.** A reasonable first pilot.

✓ (D) A fully autonomous agent granting credit-limit increases to customers — **Correct.** High autonomy, high-risk use and customer impact — poor as a learning pilot.

Q17 **Answer: A** Review: M3 · 3.2

A steering group sets weights after seeing that its preferred use case ranks fourth. What is the problem?

✓ (A) Post-hoc weighting undermines defensibility — **Correct.** Weights must be fixed before scoring.

✗ (B) The scores are too low — **Wrong.** Score level is not the issue.

✗ (C) The matrix has the wrong axes — **Wrong.** The axes are not the issue.

✗ (D) Too few criteria — **Wrong.** Not the issue.

Q18 **Answer: C** Review: M3 · 3.4

Which list correctly names the five effort components used in this course?

✗ (A) People, process, technology, data, culture — **Wrong.** A generic framework, not the effort components.

✗ (B) Budget, timeline, vendor, model, licence — **Wrong.** These are not the components taught.

✓ (C) Data readiness, integration, compliance burden, adoption resistance, operational capacity — **Correct.** These are the five components.

✗ (D) Value, risk, fit, feasibility, sponsorship — **Wrong.** These are qualification dimensions.

Q19 **Answer: D** Review: M3 · 3.4

A document-review AI would route 25% of documents to legal experts who are already overloaded. Which component captures this?

✗ (A) Data readiness — **Wrong.** Data is not the constraint.

✗ (B) Integration — **Wrong.** Not about system connections.

✗ (C) Adoption resistance — **Wrong.** The experts may welcome it; the issue is capacity.

✓ (D) Operational capacity to absorb new AI workload — **Correct.** Expert review capacity limits realisable value.

Q20 **Answer: B** Review: M3 · 3.5

Missing a defect costs €20,000 (recall, warranty); a false alarm costs €50 of inspection. How should a configuration be chosen?

✗ (A) Maximise overall accuracy — **Wrong.** Accuracy treats both errors equally.

✓ (B) Price both error types and choose the configuration with the highest expected value, checking inspection capacity — **Correct.** Error-direction pricing with a capacity check is correct.

✗ (C) Minimise false alarms — **Wrong.** This ignores the much larger cost of misses.

✗ (D) Choose the vendor default — **Wrong.** Defaults are not tuned to these costs.

Q21 **Answer: D** Review: M3 · 3.5

Which practice best corrects for optimism bias in value estimates?

✗ (A) Removing ranges for clarity — **Wrong.** This hides uncertainty.

✗ (B) Using the most enthusiastic sponsor's numbers — **Wrong.** This increases bias.

✗ (C) Doubling effort for every use case — **Wrong.** Uniform doubling is arbitrary; contingency should follow certainty by component.

✓ (D) Applying evidence-based haircuts and showing a downside scenario — **Correct.** These explicitly correct the predictable direction of error.

Q22 **Answer: B** Review: M3 · 3.6–3.7

A candidate is low value and high effort after error pricing. Which disposition fits?

✗ (A) Fund now — **Wrong.** Low value and high effort do not justify funding.

✓ (B) Drop or redesign — **Correct.** Money pits are dropped or redesigned.

✗ (C) Bundle — **Wrong.** Bundling helps when marginal effort is low.

✗ (D) Prepare — **Wrong.** Preparing invests in enablers for something worth doing.

Q23 **Answer: C** Review: M3 · 3.8

What does a sensitivity test add to a prioritization?

✗ (A) A final decision — **Wrong.** Decision-holders decide; the test informs them.

✗ (B) A vendor shortlist — **Wrong.** Unrelated.

✓ (C) Evidence of whether the ranking is robust to plausible changes in weights or estimates — **Correct.** It shows whether the decision hinges on a specific assumption.

✗ (D) More criteria — **Wrong.** It does not add criteria.

Q24 **Answer: A** Review: M3 · 3.9

Which sequencing principle applies when a strategic bet depends on cleaned master data?

✓ **A** Dependencies first: fund the master-data enabler before the use case — **Correct.** Enablers precede what they unblock.

✗ B Reuse — **Wrong.** Not the main principle here.

✗ C Random order — **Wrong.** Not a principle.

✗ D Value early — **Wrong.** Relevant generally, but not the issue here.

Q25 **Answer: D** Review: M4 · 4.2

Which belongs in a tracked AI intake?

✗ A Only centrally approved projects — **Wrong.** This misses shadow, vendor and embedded AI.

✗ B Only GenAI projects — **Wrong.** All AI candidates belong.

✗ C Only projects over €100k — **Wrong.** Thresholds hide small but risky uses.

✓ **D** Business requests, discovery findings, shadow AI found, vendor proposals and embedded SaaS features — **Correct.** All visible candidates belong in intake.

Q26 **Answer: B** Review: M4 · 4.3

A candidate may involve emotion recognition of employees at work. What should happen at intake?

✗ A Score value first to see if it is worth it — **Wrong.** Risk must be classified first; a possible prohibited practice is not scored.

✓ **B** Classify as Stop and refer to legal before any further analysis — **Correct.** Potentially prohibited uses are stopped early.

✗ C Pilot it with volunteers — **Wrong.** Volunteering does not remove the risk.

✗ D Let the works council decide alone — **Wrong.** The council should be involved, but legal classification comes first.

Q27 **Answer: A** Review: M4 · 4.3

Which statement about the EU AI Act is accurate for portfolio risk classification?

✓ **A** It classifies by intended purpose, with tiers such as prohibited, high-risk, transparency obligations and minimal risk. — **Correct.** This reflects the Act's risk-based approach.

✗ B Risk depends mainly on the algorithm used. — **Wrong.** Classification depends on intended purpose and context.

✗ C It applies only to companies outside the EU. — **Wrong.** It applies to providers and deployers in the EU market.

✗ D It replaces GDPR. — **Wrong.** The two apply alongside each other.

Q28 **Answer: C** Review: M4 · 4.4

What does "one fixed set of scoring criteria" prevent?

✗ A Fast scoring — **Wrong.** It does not prevent speed.

✗ B Risk classification — **Wrong.** Risk classification happens alongside it.

✓ **C** Bespoke criteria that favour particular candidates and break comparability — **Correct.** One set keeps the full portfolio comparable.

✗ D Consolidation — **Wrong.** It supports consolidation.

Q29 **Answer: D** Review: M4 · 4.5

Which portfolio is best balanced?

✗ (A) Six high-risk strategic bets using 160% of capacity — **Wrong.** Over capacity, with risk concentrated.

✗ (B) One huge transformation programme — **Wrong.** Concentrated risk and no early value.

✗ (C) Ten fill-ins — **Wrong.** Low value overall.

✓ (D) Two quick wins, one staged strategic bet, one shared platform; two high-risk items within governance capacity; ~85% of capacity — **Correct.** A mix of timing, risk and capacity with a buffer.

Q30 **Answer: A** Review: M4 · 4.6

Two vendors propose overlapping extraction tools for finance and procurement, with similar documents and risk. What should the advisor recommend?

✓ (A) Consolidate onto one solution after confirming there are no material differences, then re-score — **Correct.** Consolidate where needs genuinely overlap.

✗ (B) Let each unit decide independently — **Wrong.** This perpetuates duplication.

✗ (C) Reject both — **Wrong.** The need is real.

✗ (D) Buy both to keep units happy — **Wrong.** Duplicate automation increases cost and fragmentation.

Q31 **Answer: B** Review: M4 · 4.7

What should a decision package include to allow a real decision?

✗ (A) A technical architecture — **Wrong.** Build detail, not a decision package.

✓ (B) Options (including "not this cycle"), resources by component, capacity evidence, risk profile, staged funding and the items not funded — **Correct.** These give the funding body a real choice.

✗ (C) A single list to approve — **Wrong.** Without options the decision is take-it-or-leave-it.

✗ (D) Only the vendor quotes — **Wrong.** Insufficient.

Q32 **Answer: C** Review: M4 · 4.8

When is "not this cycle" the correct portfolio outcome?

✗ (A) Never — some AI should always be funded — **Wrong.** This removes the decision.

✗ (B) Only if the budget is zero — **Wrong.** There are many other legitimate reasons.

✓ (C) When no candidate clears the threshold, foundations are missing, capacity is committed or risk exceeds appetite — recorded with triggers — **Correct.** These are legitimate grounds, and they must be recorded.

✗ (D) Whenever the CEO is uncertain — **Wrong.** Uncertainty alone is not a ground; evidence is.

Q33 **Answer: A** Review: M1–M4

Integrated: A business unit submits "AI for HR". In the correct order, what happens?

✓ (A) Log at intake → classify risk → normalize and screen (data gate, label test) → score value and effort → portfolio decision — **Correct.** This is the course sequence.

✗ (B) Pilot → measure → normalize — **Wrong.** Piloting before qualification invites stalls.

✗ (C) Consolidate → reject — **Wrong.** No basis yet.

✗ (D) Score value → pick vendor → classify risk — **Wrong.** Risk comes before value; normalization comes before scoring.

Q34 **Answer: D** Review: M1 · 1.9, M3 · 3.4

Integrated: An attractive GenAI drafting case shows €900k gross value. Review adds 40% of the saved time back, and the reviewers are scarce senior staff. What happens to its position?

✗ A Nothing — gross value is what counts — **Wrong.** Gross ignores the ledger.

✗ B Risk class changes to prohibited — **Wrong.** Review effort does not change the risk tier.

✗ C It becomes a money pit automatically — **Wrong.** Not necessarily; it depends on the net figures.

✓ D Net value falls (ledger) and operational capacity effort rises; it may move from quick win toward strategic bet or defer — **Correct.** Both axes change when moved costs and capacity are priced.

Q35 **Answer: C** Review: M4 · 4.6

Integrated: Three divisions each rank "contract clause extraction" as their top priority. What is the portfolio move?

✗ A Pick one division at random — **Wrong.** No rationale.

✗ B Fund three separate projects — **Wrong.** This is duplicate automation.

✓ C Consolidate into one initiative, aggregate value, share effort and design controls to the highest risk class, then re-score — **Correct.** Consolidation changes the scores and is better for the portfolio.

✗ D Reject because of duplication — **Wrong.** Duplication calls for consolidation, not rejection.

Q36 **Answer: B** Review: M2 · 2.5

Integrated: A predictive use case has strong value, an eager sponsor and data that exists — but the outcome is recorded only for 5% of cases selected by current rules. Where should it land?

✗ A Add it to the portfolio with high effort — **Wrong.** It is not ready to score reliably.

✓ B Advance with conditions or park: create representative labels first; it cannot be valued reliably until then — **Correct.** The label test fails on selection bias; fix it before prioritization.

✗ C Drop permanently — **Wrong.** The problem is fixable.

✗ D Fund now as a quick win — **Wrong.** The label problem makes the value unrealisable as specified.

R6 Further reading and next steps

Useful public companions to this course. Check for the latest versions.

Resource	Why it is useful
EU AI Act — Regulation (EU) 2024/1689, incl. Annex III	Intended-purpose risk tiers that drive portfolio risk classification (confirm current application dates)
NIST AI Risk Management Framework (AI RMF 1.0) and Playbook	Govern–Map–Measure–Manage structure for consistent risk identification
ISO/IEC 23894:2023 and ISO/IEC 42001:2023	AI risk management guidance and AI management-system requirements
Literature on optimism bias and reference-class forecasting	Background for pricing the direction of estimation error
Process mapping references (SIPOC, swimlanes, value-stream mapping)	Techniques for work-first discovery

Practice assignments

1. Observe one real workflow for two hours; complete T1.1 and three opportunity cards (T1.2) with ledgers.
2. Normalize and screen five proposals with T2.1–T2.5; give each a disposition and reason.
3. Agree criteria and weights with a colleague (T3.1) before scoring; estimate effort by component (T3.2) and price error (T3.3).
4. Assemble a portfolio of the survivors: run the balance and consolidation check (T4.5) and write a three-option decision package (T4.6), one option being “not this cycle”.

ENTERPRISE AI TRANSFORMATION AND CONSULTING PROFESSIONAL

COURSE 3 · STUDY BOOK

Assess AI Readiness

How to diagnose whether an organization can actually use a specific AI capability: the seven dimensions of readiness, data and technology foundations, organizational operating conditions, and a decision-ready report that can say “not yet”.

4 MODULES · CASE STUDIES · TEMPLATES · CHEAT SHEETS · MOCK EXAM

Study edition · 2026

Includes full answer explanations

How to use this book

This study book accompanies **Course 3, Assess AI Readiness**, of the Enterprise AI Transformation and Consulting Professional certificate. It is written for consultants, advisors, transformation, change and enablement leads, and governance professionals who must judge whether an organization is ready to deploy a specific AI capability — before investment is committed.

Each module follows the same rhythm so that you always know where you are:

Element	What you get	How to use it
Module opener	Learning objectives, section list, study time	Read first; return to it to self-test at the end
Overview & key takeaways	The module on one screen	Skim before, re-read after
Core sections	Concepts, tables, diagrams, worked examples	Study actively: annotate, restate in your own words
Callout boxes	Key ideas, examples, pitfalls, exam traps, consultant tips	Exam traps are the most frequently tested ideas
Case study	A realistic enterprise scenario with model answers	Answer the questions before reading the model answers
Templates	Fill-in consulting tools (T-codes), one per page or less	Copy into your own engagements
Do / Avoid	Best practices and pitfalls side by side	Use as a pre-meeting checklist
Cheat sheet	One-page module summary	Revise from these in the final week
Knowledge check	Scenario questions with explanations for **every** option	Understanding why the wrong answers are wrong matters as much as getting the right one

Callout key

KEY IDEA
The idea to remember if you remember nothing else.

EXAM TRAP
A frequently tested distinction or a tempting wrong answer.

EXAMPLE
A worked illustration of the concept.

CONSULTANT'S TIP
Practical wording or technique from the field.

PITFALL
A common mistake made in real engagements.

DEFINITION
A precise definition worth learning word for word.

> **NOTE · Builds on Courses 1 and 2**
> This course assumes the vocabulary of Course 1 (the three foundations, demonstration vs deployment, capability inventories, the evidence ladder, scope of authority) and Course 2 (the data readiness gate, the label test, risk by use, the tracked intake). Course 2 asked *which* AI opportunities to pursue; this course asks whether the organization is *ready* to make a chosen one work. Reused concepts are recalled briefly rather than re-taught.

Suggested study plan (four weeks)

Week	Study	Practise
1	Module 1	Bound and classify a use-case inventory (T1.1–T1.3) for an organization you know
2	Module 2	Run the data readiness worksheets on one use case and write a finding with T2.5
3	Module 3	Map roles and control points; test operating conditions; write a proceed / phase / not yet finding
4	Module 4, then revise all cheat sheets and the glossary	Grade evidence and draft a readiness report; take the mock exam (60 min) and review every explanation

> **NOTE · About the examples**
> All organizations in case studies and examples are fictional, and their figures are illustrative. Regulatory and framework references are for orientation only; always check current law and guidance, and refer legal interpretation to qualified specialists.

Contents

FRONT **How to use this book** 2

MODULE 1 **The Dimensions of Readiness** 6
What an AI readiness assessment is — and is not 7
The seven dimensions of readiness 8
Why one gap stalls a capability that demonstrates well 9
Model-level risk vs organization-level risk 9
Trustworthiness as a property of a system in context 10
The opening move: bounding the AI use-case inventory 10
Classifying each item by the scrutiny it warrants 11
Naming the governance evidence a reviewer could ask to see 12
Recording the standard and date 12
Context: NIST AI RMF and the EU AI Act 13
Case study: Mosel Health Network 14
Consulting templates 16
Best practices and pitfalls 19
Module 1 cheat sheet 20
Knowledge check 21
Knowledge check: answers and explanations 23

MODULE 2 **Data and Technology Readiness** 25
Data readiness is fitness for a specific purpose 26
Data quality 26
Data governance 27
Data lineage and provenance 27
Access control and permissions 28
Single-source-of-truth data architecture 28
Training data and labelled outcome records 29
Data fidelity: does the record reflect reality? 29
Can access, should access, permitted to access 30
Tiered controls for higher-stakes and regulated data 30
Technology readiness 31
Ownership of data assets and of the deployed system 31
Writing a data readiness finding a sponsor can act on 32
Practice scenario: Orbis Facilities 33
Consulting templates 35
Best practices and pitfalls 37
Module 2 cheat sheet 38
Knowledge check 39
Knowledge check: answers and explanations 41

MODULE 3 **Organizational Readiness** 43
Why organizational readiness decides value 44
Establishing what readiness is measured against 44
Mapping the roles and control points a change touches 45
Operating conditions vs assurances 45
Workforce AI proficiency 46
Human oversight and human-in-the-loop review design 46

Accountability and ownership 47
AI policy clarity 47
Reinforcement and frontline managers 47
Change and adoption readiness, resistance and shadow AI 48
Rework and value capture 48
The readiness finding: proceed, phase or not yet 49
Case study: Lindqvist Bank 50
Consulting templates 52
Best practices and pitfalls 54
Module 3 cheat sheet 55
Knowledge check 56
Knowledge check: answers and explanations 58

MODULE 4 **Reporting Readiness** 60
From assessment to a decision-ready report 61
Evidence grading: verified vs asserted 61
Documenting how findings were learned 62
Connecting each condition to a business consequence 63
Maturity gaps against the required state 63
Evidence-based, sequenced recommendations 63
The professional use of “not yet” 64
Communicating to different stakeholders 64
The readiness report structure 65
Guided activity: from raw observations to findings 66
Consulting templates 68
Best practices and pitfalls 70
Module 4 cheat sheet 71
Knowledge check 72
Knowledge check: answers and explanations 74

BACK **Course at a glance: the master cheat sheet** 76
Template index 76
Glossary 78
Mock exam 81
Mock exam: answer key and explanations 87
Further reading and next steps 95

MODULE 1

The Dimensions of Readiness

The diagnostic lens behind an AI readiness assessment: seven dimensions along which readiness can fail, the line between model-level and organization-level risk, trustworthiness in context — and the opening moves that make later scoring defensible.

YOU WILL BE ABLE TO

- Describe the seven dimensions along which organizational readiness for AI can fail
- Explain why a gap in one dimension stalls a capability that performs well in a demonstration
- Separate model-level risk (established by testing) from organization-level risk (established by inventory and governance review)
- Treat trustworthiness as a property of a system in context, not of a model in isolation
- Bound an AI use-case inventory that reaches embedded and shadow AI
- Classify each item by the scrutiny it warrants and name the evidence a reviewer could ask to see
- Record the standard and date that later scoring is measured against
- Describe NIST AI RMF and EU AI Act readiness expectations as context

IN THIS MODULE

1. What a readiness assessment is
2. The seven dimensions
3. Why one gap stalls a good demo
4. Model-level vs organization-level risk
5. Trustworthiness in context
6. Bounding the use-case inventory
7. Classifying by scrutiny
8. Naming the evidence
9. Recording the standard and date
10. Regulatory and framework context
11. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4–5 hours

An AI readiness assessment answers one question: **can this organization make this AI capability work, safely and with value, in its real operating context?** This module builds the diagnostic lens for answering it — seven dimensions along which readiness fails — and teaches the opening moves that make every later score defensible: bounding the inventory, classifying scrutiny, naming the evidence and fixing the standard and date.

KEY TAKEAWAYS

1. Readiness is assessed **for a specific capability in a specific context** — not as a generic maturity level.
2. Seven dimensions: **data, technology, ownership, adoption, governance, measurement, trust**. Readiness is a chain: the weakest link decides.
3. A demo exercises only the model and curated data; production exercises all seven dimensions — so **one gap stalls a capability that demonstrated well**.
4. **Model-level risk** is established by testing; **organization-level risk** by structural inventory and governance review. One does not substitute for the other.
5. **Trustworthiness is a property of a system in context** — model, data, people, process, oversight and purpose together.
6. Open every assessment by **bounding the inventory** (including embedded and shadow AI), **classifying scrutiny**, **naming the evidence** a reviewer could ask for, and **recording the standard and date**.

1.1 What an AI readiness assessment is — and is not

	AI maturity assessment	AI readiness assessment
Question	How capable is the organization at AI in general?	Can the organization make *this* capability work, in *this* context, now?
Unit	Organization or unit	A use case or capability in its operating context
Benchmark	Generic staged maturity model	A defined target state and a recorded standard
Output	Maturity level, development roadmap	Finding: proceed, phase or not yet — with named remediation
Timing	Periodic	Before deployment or scaling decisions; repeated at gates

KEY IDEA

An organization can be "mature" overall and still not ready for a specific use case — and vice versa. Readiness is always **readiness for something**.

NOTE · Where this sits in the series

Course 2 selected and prioritized opportunities. Readiness assessment typically happens **after selection and before a deployment or scaling commitment** — the pre-deployment assessment — and again at later gates.

1.2 The seven dimensions of readiness

Dimension	Ready means...	Typical failure	Opening test question
Data	The data the capability needs is available, permitted, fit, representative and owned	Pilot data was curated; production data is incomplete or not permitted	"Show me the production data this will run on."
Technology	Infrastructure, integration, security and run tooling support the capability at production scale	Works standalone; cannot integrate with systems of record, identity or logging	"How does an output reach the system where work happens?"
Ownership	Named owners for the business outcome, the system and the data, with authority	Innovation team owns pilot; nobody owns production	"Who answers when it goes wrong at 3 a.m.?"
Adoption	People will use it as intended; process, skills and incentives support use	Users bypass or rubber-stamp it; managers don't reinforce	"What will a frontline user do differently on Monday?"
Governance	Rules of use, risk classification, oversight, decision rights and documentation exist and operate	Legal/risk review starts at go-live; rules unclear	"Who approved this use, against which policy?"
Measurement	Baseline, success metrics, monitoring and feedback loops are defined and running	No baseline; value cannot be shown; drift undetected	"What number will tell us in 90 days whether it works?"
Trust	Affected people and users have justified, calibrated trust — neither blind reliance nor rejection	Over-reliance (automation bias) or distrust; complaints from affected people	"When should a user *not* rely on this output — and do they know?"

Readiness is a chain, not an average

A capability is as ready as its weakest dimension. Six strong links do not carry load across one broken link — which is why averaged "readiness scores" mislead.

Figure 1.1 — Readiness as a chain

> **EXAM TRAP**
> An option that averages dimension scores ("overall readiness 3.9/5, so proceed") is almost always wrong when one dimension is critically weak. Readiness behaves like a chain: report each dimension and decide on the **weakest material link**.

1.3 Why one gap stalls a capability that demonstrates well

A demonstration tests a narrow slice of the system: a model on curated inputs, operated by its builders. Production exercises the full chain — live data, integrations, owners, users, rules, metrics and the trust of affected people. A gap in any of them does not usually make the system fail visibly; it makes it **stall**: go-live is postponed, usage stays low, or value cannot be shown.

Demo looked like…	Production revealed…	Dimension
95% correct answers on sample questions	Knowledge base 40% out of date	Data
Instant responses in a sandbox	No route into the case-management system; manual copy-paste	Technology
Enthusiastic innovation team	No business unit will run or fund it	Ownership
Users loved the pilot	Under time pressure, users skip it; managers measure old KPIs	Adoption
“Legal is fine with it”	No documented approval, risk class or oversight design	Governance
Positive feedback forms	No baseline; cannot prove any saving	Measurement
Accurate on average	Customer complaints after confident wrong answers; staff stop trusting it	Trust

1.4 Model-level risk vs organization-level risk

Two kinds of risk need two kinds of evidence. Confusing them is one of the most common assessment errors — typically, an excellent model evaluation is accepted as evidence that the organization is ready.

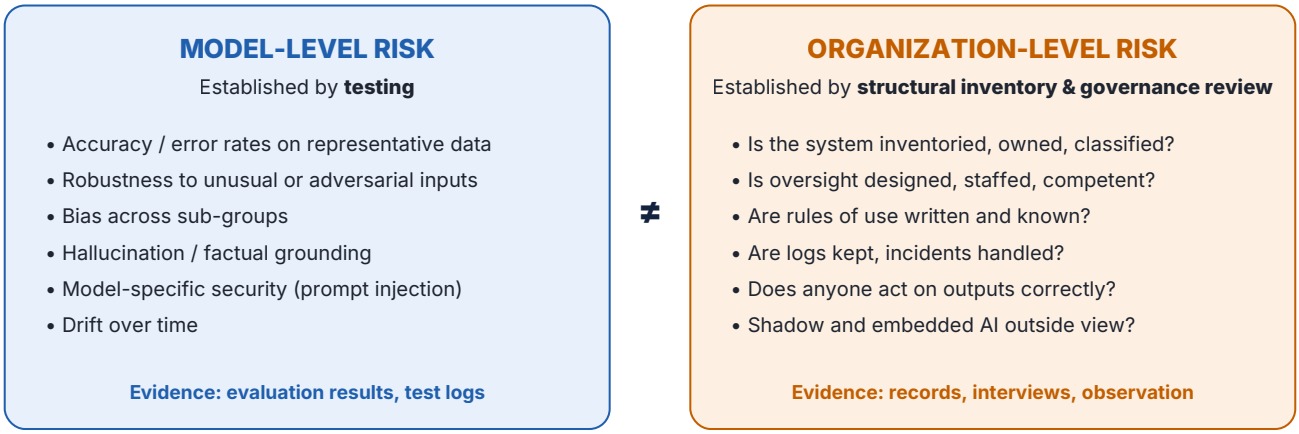


Figure 1.2 — Two kinds of risk, two kinds of evidence

Statement	Which risk?	What would establish it
“The model’s error rate on German invoices is 4%.”	Model-level	Evaluation on a representative test set
“Every output is reviewed by a qualified clerk.”	Organization-level	Oversight specification, staffing records, observation of review
“The model treats older applicants fairly.”	Model-level	Sub-group performance testing

Statement	Which risk?	What would establish it
"Nobody uses unapproved AI tools for customer data."	Organization-level	Inventory discovery: logs, survey, interviews
"The system resists prompt injection."	Model-/system-level	Security testing (red-teaming)
"The system can be stopped if it misbehaves."	Organization-level	Documented stop procedure, owner, tested drill

PITFALL · Testing your way to organizational readiness

No amount of model testing shows that someone owns the system, that reviewers have time and authority to override it, or that shadow tools aren't processing the same data elsewhere. Those facts come from records, interviews, observation and inventory.

1.5 Trustworthiness as a property of a system in context

"Is this model trustworthy?" is the wrong question. The same model can be trustworthy summarising internal meeting notes and untrustworthy drafting medical discharge letters without review. Trustworthiness emerges from the combination of **model + data + people + process + oversight + purpose**.

Trustworthiness characteristic (NIST AI RMF)	What it depends on in context
Valid and reliable	Performance on *this* population and data, over time
Safe	Consequences of failure in *this* process; fallback and stop mechanisms
Secure and resilient	Integrations, access, tool permissions, threat environment
Accountable and transparent	Owners, documentation, disclosure to affected people
Explainable and interpretable	Whether users and affected people can understand outputs enough for *this* decision
Privacy-enhanced	Data used, purpose, legal basis, controls
Fair, with harmful bias managed	Who is affected and how outcomes differ across groups in *this* use

EXAMPLE

A vendor states its model is "trustworthy AI — certified". The assessor's response: a certificate may evidence parts of model-level and provider-level practice, but whether the **system** is trustworthy depends on how this organization deploys it — data fed in, users, oversight, purpose. The certificate is input evidence, not the conclusion.

1.6 The opening move: bounding the AI use-case inventory

Every assessment starts by deciding what is in scope. If the inventory only lists the systems the sponsor mentioned, the assessment will miss the AI most likely to cause trouble: embedded features and shadow use.

Bounding decision	Questions	Example
Organizational scope	Which units, entities, countries?	Claims operations, DE and AT

Bounding decision	Questions	Example
Use-case scope	The target capability only, or everything touching the same process and data?	The triage assistant plus all AI touching claim data
System types	Built, bought, embedded, shadow?	All four
Lifecycle	Production, pilot, planned?	Production and pilot; planned listed separately
Time	As of which date?	Inventory as of 15 March 2026
Counting rules	What is one item? (Course 1, Module 2)	One item = one use of AI for one purpose

BEST PRACTICE · Reach embedded and shadow AI

- **Embedded:** ask each platform owner which AI features are enabled; review release notes of core SaaS tools.
- **Shadow:** aggregated proxy/SSO logs (with privacy and works-council agreement), expense data, anonymous survey, interviews.
- **Adjacent use:** other AI processing the same data or feeding the same decision — it shares the risk.

1.7 Classifying each item by the scrutiny it warrants

Not every inventory item deserves the same depth of review. Classify each by **scrutiny tier**, using use-based risk factors from Course 2 (who is affected, consequence, reversibility, autonomy, scale, regulated domain), so assessment effort is proportionate.

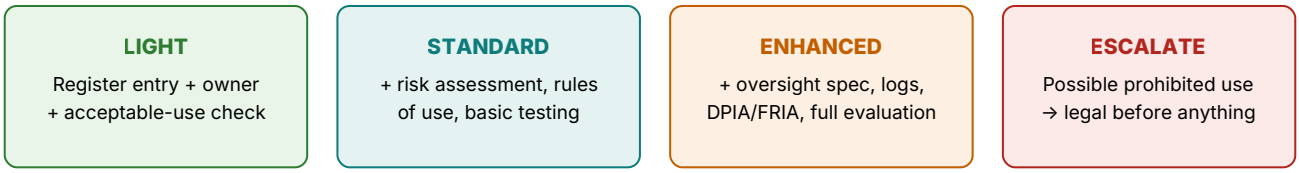


Figure 1.3 — Scrutiny tiers

Item	Key factors	Scrutiny tier
Embedded meeting-transcription feature	Internal; personal data of staff; employee-monitoring sensitivity	Standard (works-council question)
Claims triage assistant (target)	Customer impact; routing affects payment speed; human review	Enhanced
Staff using a public chatbot for e-mails (shadow)	Customer personal data leaving the company	Standard → urgent governance action
Spelling correction in office suite	Negligible	Light
Proposed emotion analysis of call-centre agents	Possible prohibited practice in workplace context	Escalate

1.8 Naming the governance evidence a reviewer could ask to see

For each item, list in advance the evidence an internal auditor, regulator or works council could reasonably ask for. Naming evidence up front turns the assessment from conversation into verification — and shows the client what "ready" looks like in documents.

Evidence	What it shows	Tier
Inventory entry	The system is known, described and classified	All
Ownership record	Named business, system and data owners	All
Intended-purpose statement and rules of use	What it may and may not be used for	Standard+
Risk assessment / classification record	Risk was assessed against a stated framework	Standard+
DPIA (and, where required, fundamental-rights impact assessment)	Privacy and rights impacts assessed	Enhanced, where triggered
System / model documentation (e.g., provider instructions, model card)	Capabilities, limits, intended conditions	Standard+
Evaluation and test results	Model-level performance on representative data	Standard+
Human oversight specification	Who reviews what, with what authority, competence and time; how to override or stop	Enhanced
Training / AI literacy records	Users and overseers are competent	Standard+
Logs and log-retention arrangements	Events can be reconstructed	Enhanced
Monitoring reports and incident records	Performance and problems are tracked and handled	Enhanced
Vendor agreements (data use, change notice, exit)	Third-party terms support the use	Where bought
Change log	Changes to model, prompts, data, configuration are controlled	Enhanced

> **CONSULTANT'S TIP**
> Ask for evidence **by name** ("please share the oversight specification and last month's review sample") rather than by topic ("how do you do oversight?"). Named requests reveal quickly whether a document exists — and absence is itself a finding.

1.9 Recording the standard and date

Before scoring anything, record **what readiness is measured against** and **when**. Standards, regulations, internal policies and systems all change; a score without a recorded standard and date cannot be reproduced, compared at re-assessment or defended later.

Baseline element	Example
Standards and frameworks	Internal AI Policy v2.1 (approved 10 Jan 2026); NIST AI RMF 1.0; ISO/IEC 42001:2023 (selected controls)

Baseline element	Example
Regulation as applicable on the date	EU AI Act obligations applicable to deployers as of the assessment date; GDPR
Target state	Triage assistant at autonomy level 2, 30,000 claims/month, human review of all routings to fraud team
Assessment date and evidence cut-off	Fieldwork 3–14 March 2026; evidence accepted until 14 March
Scope reference	Inventory bounding sheet v1.0
Assessor(s)	Names and roles; independence statement

KEY IDEA

"Ready" is not a timeless property. It means: **ready against standard S, for target state T, on date D**. Without S, T and D, the finding cannot be checked by anyone else.

1.10 Context: NIST AI RMF and the EU AI Act

Two frameworks shape readiness expectations for many enterprises. They are described here as **context**; formal legal applicability must be determined by legal specialists.

Framework	Readiness-relevant expectations
NIST AI RMF 1.0 (voluntary)	**Govern**: policies, accountability structures, roles, workforce competence and culture are in place. **Map**: context, intended purpose and impacts are understood and AI systems are inventoried. **Measure**: risks and trustworthiness characteristics are assessed and tracked. **Manage**: risks are prioritized, responded to and monitored, with incident response.
EU AI Act (Regulation (EU) 2024/1689)	Risk-based obligations by intended purpose. Prohibited practices and an AI-literacy duty have applied since February 2025; since the AI Omnibus (Regulation (EU) 2026/1744, in force 27 July 2026) the literacy duty requires providers and deployers to take measures to support staff AI literacy, without guaranteeing a specific individual level. Deployers of high-risk systems must, among other things, use them according to instructions, assign human oversight to competent people with the necessary authority, ensure input data is relevant and sufficiently representative where they control it, monitor operation, keep logs under their control for an appropriate period, and inform affected workers; certain deployers must carry out a fundamental-rights impact assessment. After the Omnibus, high-risk obligations apply from 2 December 2027 for stand-alone (Annex III) systems and 2 August 2028 for AI in regulated products (Annex I); confirm the consolidated text before relying on dates.

NOTE

These expectations map naturally onto the seven dimensions: governance and ownership (Govern; oversight, roles), data (input data), adoption and trust (AI literacy, informing workers), measurement (monitoring, logs). Use them to enrich the evidence list — not to replace a use-case-specific assessment.

1.11 Case study: Mosel Health Network

CASE STUDY **The documentation assistant that demonstrated perfectly**

Mosel Health Network (fictional) · five hospitals · 9,000 staff

Situation

A vendor demonstration of an AI assistant that drafts discharge letters from ward notes impressed the clinical board: letters in seconds, accurate on all 30 demo cases. The board asks for a readiness assessment before a group-wide rollout.

Opening moves

- **Bounding:** scope widened from "the assistant" to all AI touching patient documentation. Discovery found an embedded speech-to-text feature switched on in the electronic record at two hospitals, and junior doctors using a public chatbot to polish letters (shadow AI with patient data).
- **Scrutiny:** discharge-letter drafting — enhanced (patient safety; clinical decisions downstream). Embedded speech-to-text — standard. Shadow chatbot — standard, with urgent governance action. Spell-check — light.
- **Evidence named:** oversight specification for letter sign-off, clinician training records, evaluation on the network's own notes, DPIA, vendor data-processing terms, log arrangements, incident process.
- **Baseline recorded:** Network AI Policy v1.3, NIST AI RMF 1.0, applicable EU AI Act and medical-device considerations as of 2 February 2026 (legal opinion requested), target state: drafts reviewed and signed by the responsible physician.

Dimension scan (summary)

Dimension	Finding	Status
Data	Ward notes vary by hospital; two use free text only; demo used structured notes	Weak
Technology	Integration with the record system available at three of five hospitals	Partial
Ownership	Chief Medical Officer sponsors; no named system owner	Weak
Adoption	Doctors keen; nursing staff not consulted about changed hand-offs	Partial
Governance	No oversight specification; sign-off assumed, not designed	Weak
Measurement	No baseline for letter turnaround or error rates	Weak
Trust	Model-level testing on own notes not done; demo results only	Unknown

Discussion questions

1. Why did the demo say little about readiness?
2. Which statements in the scan concern model-level risk and which organization-level risk?
3. Why is "trust" marked unknown rather than strong?
4. Why must the baseline be recorded before scoring begins here?

Model answers (brief)

1. It tested a model on 30 curated, structured cases run by the vendor — one link of the chain. Production would exercise all seven dimensions.
2. Model-level: performance on the network's own (free-text) notes. Organization-level: ownership, oversight design, integration coverage, consultation of nurses, baseline, shadow use.
3. Trustworthiness depends on the system in context; without testing on own data and an operating oversight design, there is no evidence either way.
4. Regulatory interpretation (including possible medical-device implications) and internal policy may change; the finding must state which standard and date it was measured against to be defensible and comparable at re-assessment.

1.12 Consulting templates

T1.1 Readiness Dimension Scan

Purpose: First-pass view of all seven dimensions for one capability. **Use when:** first week of any readiness assessment

Capability / target state: .....

Standard & date reference (T1.4): .....

Dimension	Ready means (for this use)	Initial observation	Status (Strong / Partial / Weak / Unknown)	Evidence to request
Data				
Technology				
Ownership				
Adoption				
Governance				
Measurement				
Trust				

Weakest material link: .....

Model-level questions to test: .....

Organization-level questions to review: .....

T1.2 Inventory Bounding & Scrutiny Register

Purpose: Bound the inventory, reach embedded and shadow AI, and assign scrutiny.

BOUNDING

Organizational scope: .....

Use-case scope (target + adjacent uses of same data/decision): .....

System types included (built / bought / embedded / shadow): .....

Lifecycle stages: .....

Inventory as-of date and counting rule: .....

REGISTER

ID	Item	Type	Use / decision affected	Key risk factors	Scrutiny tier	Owner
R-03	*Public chatbot for letters*	*Shadow*	*Patient letters*	*Patient data leaves org*	*Standard + urgent*	*—*

T1.3 Governance Evidence Request List

Purpose: Name, in advance, the evidence a reviewer could ask to see.

Item / scrutiny tier: .....

Evidence	Requested from	Date requested	Received? (Y/N/partial)	Notes
Inventory entry & ownership record				
Intended purpose & rules of use				
Risk classification record				
DPIA / FRIA (if triggered)				
System / model documentation				
Evaluation results on own data				
Human oversight specification				
Training / AI literacy records				
Logs & retention arrangements				
Monitoring & incident records				
Vendor terms (data, change, exit)				
Change log				

T1.4 Assessment Baseline Record

Purpose: Record what readiness is measured against, and when. **Use when:** before any scoring

Standards / frameworks and versions: .....

Internal policies and versions (approval dates): .....

Regulation considered — as applicable on date: .....

Legal determinations requested / received: .....

Target state (autonomy, volume, oversight design, users): .....

Fieldwork dates and evidence cut-off date: .....

Scope reference (T1.2 version): .....

Assessor(s) and independence statement: .....

Approved by sponsor / date: .....

1.13 Best practices and pitfalls

✓ DO THIS

- Define readiness for a specific capability and target state.
- Report each dimension; decide on the weakest material link.
- Test model-level risk; review organization-level risk.
- Judge trustworthiness of the system in its context.
- Bound the inventory to reach embedded, shadow and adjacent AI.
- Make scrutiny proportionate to use-based risk.
- Name evidence in advance; record standard and date first.

× AVOID THIS

- Using a generic maturity level as a readiness verdict.
- Averaging dimension scores into one number.
- Treating a good demo or evaluation as readiness.
- Accepting “trustworthy model” certificates as the conclusion.
- Assessing only the system the sponsor named.
- Reviewing everything at the same depth.
- Scoring before the baseline is recorded.

1.14 Module 1 cheat sheet

MODULE 1 · ONE-PAGE SUMMARY

The Dimensions of Readiness

READINESS ≠ MATURITY

Readiness = can **this** capability work in **this** context now? Output: proceed / phase / not yet.

SEVEN DIMENSIONS

Data · technology · ownership · adoption · governance · measurement · trust. A chain: the weakest material link decides. Never average.

WHY DEMOS MISLEAD

Demo = model + curated data + builders. Production = all seven links. One gap → stall, not visible failure.

TWO KINDS OF RISK

Model-level (accuracy, robustness, bias, hallucination, drift) → **testing**. **Organization-level** (inventory, owners, oversight, rules, logs, shadow AI) → **structural inventory & governance review**.

TRUSTWORTHINESS IN CONTEXT

System = model + data + people + process + oversight + purpose. NIST characteristics: valid & reliable · safe · secure & resilient · accountable & transparent · explainable & interpretable · privacy-enhanced · fair with harmful bias managed.

OPENING MOVES

1. Bound inventory (incl. embedded, shadow, adjacent)
2. Classify scrutiny: light · standard · enhanced · escalate
3. Name evidence a reviewer could ask for
4. Record standard & date

EVIDENCE LIST

Inventory entry · owners · purpose & rules · risk record · DPIA/FRIA · documentation · evaluation · oversight spec · training · logs · monitoring & incidents · vendor terms · change log.

CONTEXT FRAMEWORKS

NIST AI RMF: Govern · Map · Measure · Manage. **EU AI Act**: prohibited practices & AI literacy (since Feb 2025); high-risk deployer duties — use per instructions, competent human oversight, relevant input data, monitoring, logs, inform workers, FRIA for certain deployers; high-risk duties apply from 2 Dec 2027 (Annex III) / 2 Aug 2028 (Annex I) after the 2026 Omnibus. “Ready” = against standard S, for target T, on date D.

1.15 Knowledge check

Answer all questions before turning the page. Explanations for every option follow.

K1 **An assessment rates six dimensions "strong" and governance "weak", then reports "overall readiness 4.1/5 — proceed". What is wrong?**

(A) The scale should be 1–10.
(B) Nothing; 4.1 is a high score.
(C) Governance should have been weighted more heavily.
(D) Readiness behaves like a chain; a weak material dimension should drive the finding, not an average.

K2 **A vendor demo on 30 curated cases was flawless. Why does this say little about readiness?**

(A) A demo exercises only the model on curated data; production exercises all seven dimensions.
(B) Vendors always cheat in demos.
(C) Thirty cases are too few to compute accuracy.
(D) Demos measure trust, not performance.

K3 **Which statement concerns organization-level risk?**

(A) "The model resists prompt injection."
(B) "A named, trained reviewer with authority to override checks every routing to the fraud team."
(C) "Outputs are fair across age groups."
(D) "Error rate on our invoices is 4%."

K4 **A provider shows a certificate stating its model is "trustworthy AI". How should the assessor treat it?**

(A) As a reason to skip testing on own data
(B) As proof that the deployment is trustworthy
(C) As input evidence about the model or provider, not as the conclusion about the deployed system
(D) As irrelevant

K5 **A sponsor asks for a readiness assessment of "our new triage assistant only". What is the best bounding advice?**

(A) Assess every AI system in the company.
(B) Agree; scope is the sponsor's choice.
(C) Include the target plus embedded, shadow and adjacent AI that touch the same data or decision, and record the boundary.
(D) Postpone until a full maturity assessment is done.

K6 **Why classify inventory items by scrutiny tier?**

(A) To make assessment effort proportionate to use-based risk, so it finishes in time to matter
(B) To reduce the number of items listed
(C) Because regulators require four tiers
(D) To avoid reviewing shadow AI

K7 **Why record the standard and date before any scoring?**

(A) To make the report longer

(B) Because standards, regulation, policies and systems change; without them the finding cannot be reproduced, compared or defended

(C) Because the EU AI Act requires a date on every document

(D) To limit the assessor's liability only

K8 **Which statement about the EU AI Act is best used as readiness context?**

(A) It replaces internal AI policies.

(B) It rates organizations on a maturity scale.

(C) It only applies to AI providers, never to deployers.

(D) Deployers of high-risk systems must, among other things, assign human oversight to competent people and keep logs under their control; an AI-literacy duty already applies.

1.16 Knowledge check: answers and explanations

K1 **Answer: D** Review: 1.2

An assessment rates six dimensions "strong" and governance "weak", then reports "overall readiness 4.1/5 — proceed". What is wrong?

- ✗ A The scale should be 1–10. — **Wrong.** Scale granularity is not the problem.
- ✗ B Nothing; 4.1 is a high score. — **Wrong.** Averaging lets six strong links hide one broken one.
- ✗ C Governance should have been weighted more heavily. — **Wrong.** Reweighting still produces an average that can hide a critical gap.
- ✓ D Readiness behaves like a chain; a weak material dimension should drive the finding, not an average. — **Correct.** The weakest material link decides whether the capability can work.

K2 **Answer: A** Review: 1.3

A vendor demo on 30 curated cases was flawless. Why does this say little about readiness?

- ✓ A A demo exercises only the model on curated data; production exercises all seven dimensions. — **Correct.** Readiness gaps in data, ownership, governance and the others do not appear in a demo.
- ✗ B Vendors always cheat in demos. — **Wrong.** The issue is coverage, not honesty.
- ✗ C Thirty cases are too few to compute accuracy. — **Wrong.** Sample size matters for model evidence but not for the organizational dimensions.
- ✗ D Demos measure trust, not performance. — **Wrong.** Incorrect.

K3 **Answer: B** Review: 1.4

Which statement concerns organization-level risk?

- ✗ A "The model resists prompt injection." — **Wrong.** Model/system-level, established by security testing.
- ✓ B "A named, trained reviewer with authority to override checks every routing to the fraud team." — **Correct.** This concerns oversight design, staffing and authority — established by governance review.
- ✗ C "Outputs are fair across age groups." — **Wrong.** Model-level, established by sub-group testing.
- ✗ D "Error rate on our invoices is 4%." — **Wrong.** Model-level, established by testing.

K4 **Answer: C** Review: 1.5

A provider shows a certificate stating its model is "trustworthy AI". How should the assessor treat it?

- ✗ A As a reason to skip testing on own data — **Wrong.** Own-context testing is still needed.
- ✗ B As proof that the deployment is trustworthy — **Wrong.** Trustworthiness is a property of the system in context, which the certificate does not cover.
- ✓ C As input evidence about the model or provider, not as the conclusion about the deployed system — **Correct.** Context — data, people, process, oversight, purpose — still has to be assessed.
- ✗ D As irrelevant — **Wrong.** It may evidence useful model-level practice.

K5 **Answer: C** Review: 1.6

A sponsor asks for a readiness assessment of "our new triage assistant only". What is the best bounding advice?

- ✗ A Assess every AI system in the company. — **Wrong.** Unbounded scope wastes effort; it should be proportionate.
- ✗ B Agree; scope is the sponsor's choice. — **Wrong.** A narrow scope would miss embedded, shadow and adjacent AI that share data and risk.
- ✓ C Include the target plus embedded, shadow and adjacent AI that touch the same data or decision, and record the boundary. — **Correct.** This reaches the AI most likely to cause trouble and makes the scope explicit.
- ✗ D Postpone until a full maturity assessment is done. — **Wrong.** Maturity is a different instrument.

K6 **Answer: A** Review: 1.7

Why classify inventory items by scrutiny tier?

✓ **A** To make assessment effort proportionate to use-based risk, so it finishes in time to matter — **Correct.** Proportionate scrutiny focuses depth where risk is highest.

✗ **B** To reduce the number of items listed — **Wrong.** All items stay listed; only review depth differs.

✗ **C** Because regulators require four tiers — **Wrong.** No regulation prescribes these tiers.

✗ **D** To avoid reviewing shadow AI — **Wrong.** Shadow AI is classified like any other item.

K7 **Answer: B** Review: 1.9

Why record the standard and date before any scoring?

✗ **A** To make the report longer — **Wrong.** Irrelevant.

✓ **B** Because standards, regulation, policies and systems change; without them the finding cannot be reproduced, compared or defended — **Correct.** Ready means ready against standard S, for target T, on date D.

✗ **C** Because the EU AI Act requires a date on every document — **Wrong.** That is not the reason.

✗ **D** To limit the assessor's liability only — **Wrong.** Defensibility helps, but the main purpose is reproducibility and comparability.

K8 **Answer: D** Review: 1.10

Which statement about the EU AI Act is best used as readiness context?

✗ **A** It replaces internal AI policies. — **Wrong.** Internal policies remain necessary.

✗ **B** It rates organizations on a maturity scale. — **Wrong.** It sets obligations by use-based risk, not maturity levels.

✗ **C** It only applies to AI providers, never to deployers. — **Wrong.** Deployers have obligations, especially for high-risk systems.

✓ **D** Deployers of high-risk systems must, among other things, assign human oversight to competent people and keep logs under their control; an AI-literacy duty already applies. — **Correct.** These are readiness-relevant deployer expectations; dates should be confirmed.

MODULE 2

Data and Technology Readiness

Where enterprise AI most often stalls. Judge whether data and technical foundations can support a specific use case — quality, governance, lineage, access, a single source of truth, labelled outcomes, fidelity and tiered controls — and write a finding a sponsor can act on.

YOU WILL BE ABLE TO

- Judge data readiness for a specific use case, not in the abstract
- Assess data quality, governance, lineage and provenance
- Assess access control, permissions and single-source-of-truth architecture
- Check training data and labelled outcome records, and data fidelity
- Distinguish what an organization can, should and is permitted to access
- Apply tiered controls for higher-stakes and regulated data
- Assess technology readiness and ownership of data assets and the deployed system
- Write a data readiness finding a sponsor can act on

IN THIS MODULE

1. Fit for a specific purpose
2. Data quality
3. Data governance
4. Lineage and provenance
5. Access control and permissions
6. Single source of truth
7. Training data and labelled outcomes
8. Data fidelity
9. Can, should, permitted
10. Tiered controls
11. Technology readiness
12. Ownership
13. Writing the finding
14. Practice scenario · templates · cheat sheet · knowledge check

Suggested study time · 4–5 hours

Data readiness is where enterprise AI initiatives most often stall. The question is never "is our data good?" in general, but "can **this** data, through **this** technology, support **this** use case — lawfully, faithfully and at production scale?" This module gives you the tests, the vocabulary and a finding format that turns technical observations into something a sponsor can act on.

KEY TAKEAWAYS

1. Data readiness is **fitness for a specific purpose**; the same data can be ready for one use case and not for another.
2. Assess **quality**, **governance**, **lineage and provenance**, **access and permissions**, and a **single source of truth**.
3. ML learns from **labelled outcome records**; GenAI depends on the quality of its **knowledge content**. Check both.
4. **Data fidelity**: does the recorded data faithfully represent what actually happened?
5. Separate what the organization **can**, **should** and is **permitted** to access.
6. Apply **tiered controls** to higher-stakes and regulated data; assess **technology readiness** for production, not demo.
7. Name **owners** of the data assets and of the deployed system — and write **actionable findings**: condition, criterion, evidence, consequence, recommendation.

2.1 Data readiness is fitness for a specific purpose

Customer data that is perfectly adequate for monthly reporting may be useless for real-time next-best-action: too late, too aggregated, missing the outcome. Always assess data against the **target state** recorded in the baseline (Module 1): the task, the decision, the latency, the volume, the population.

EXAM TRAP

"Our data is high quality — we passed the data-governance audit" is not evidence of readiness for a specific AI use case. The audit measured data against *its* criteria, which may not include the fields, history, latency or labels this use case needs.

2.2 Data quality

Quality dimension	Question for the use case	How to test
Accuracy	Do values match reality?	Compare a sample against source documents or physical checks
Completeness	Are the needed fields filled for the needed population?	Profile null rates by field, segment and period
Consistency	Are values defined and coded the same way across systems, units and time?	Compare code lists; look for definition changes
Timeliness	Is data available when the decision is made?	Measure lag from event to availability
Validity	Do values conform to formats and rules?	Rule checks (dates, ranges, reference values)

Quality dimension	Question for the use case	How to test
Uniqueness	Are entities duplicated?	Duplicate detection on keys and fuzzy matching

CONSULTANT'S TIP · Profile, don't ask

Asking "is the data complete?" produces assurances. Running a profile on a recent extract produces evidence: "field *failure_code* is empty in 37% of 2025 work orders; 21% of the rest are 'OTHER'." Ten minutes of profiling beats an hour of interviews.

2.3 Data governance

Data governance is the set of roles, policies and processes that keep data fit for use over time. For AI readiness, test whether it **operates** for the specific data in scope.

Element	Ready when…
Ownership & stewardship	A named data owner is accountable; stewards maintain definitions and quality
Definitions	Key terms (e.g., "active customer", "resolved") are defined in a glossary and used consistently
Quality management	Quality rules are monitored with thresholds and an issue process
Policy for AI use	Policy states whether and how this data may be used for AI training, fine-tuning or retrieval
Retention & deletion	Rules exist and are applied — including to copies in AI pipelines and vector stores
Change management	Schema and code changes are communicated to downstream AI users

2.4 Data lineage and provenance

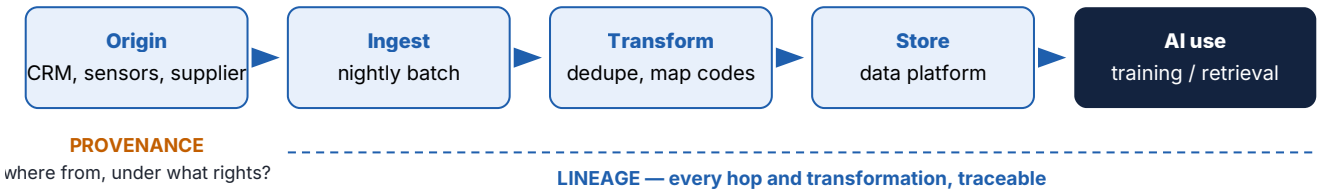


Figure 2.1 — Provenance is the origin and rights; lineage is the path

	Provenance	Lineage
Question	Where did the data originally come from, and under what rights and conditions?	Which systems and transformations did it pass through to reach the AI?
Why it matters	Rights to use (consent, licence, contract); trust in the source; third-party and scraped content	Debugging errors; impact of upstream changes; audit and reproducibility
Evidence	Source records, licences, consent records, supplier contracts	Lineage diagrams/metadata, pipeline code, data catalogue

> **NOTE**
> For providers of high-risk systems, the EU AI Act sets data-governance requirements for training, validation and testing data (e.g., relevance, representativeness, examination for possible biases). Deployers must ensure input data under their control is relevant and sufficiently representative for the intended purpose. Lineage and provenance evidence is how you show it.

2.5 Access control and permissions

AI changes how access control fails. A generative AI assistant connected to document stores retrieves whatever the connecting identity can reach. If file permissions were sloppy — "everyone in the company" on HR folders, forgotten shared links — the assistant will surface those documents to anyone who asks. The permissions were always wrong; AI makes them **discoverable**.

Check	Ready when...
Role-based access	Access is granted by role and least privilege, reviewed periodically
Permission-aware retrieval	The AI enforces the end user's permissions, not a broad service account's
Oversharing review	Sensitive repositories have been reviewed for over-broad sharing before connection
Service accounts and tool permissions	Agents and pipelines have scoped credentials; actions are logged
Vendor access	Third parties' access to data is contractually limited and technically enforced

> **PITFALL · The oversharing surprise**
> A company connects an AI assistant to its document platform; within a week a junior employee asks it about salary bands and receives the board's compensation file, which had been shared "company-wide" years earlier. Readiness assessment must include a permissions review of every repository the assistant will reach.

2.6 Single-source-of-truth data architecture

When the same fact lives in several systems with different values — three customer addresses, two asset registers — an AI system will learn or retrieve whichever version it is pointed at, and different AI systems will contradict one another. Readiness requires a designated **system of record** for each critical data entity, or a mastered view that reconciles sources.

Symptom	Readiness implication
Teams argue about which number is right	Define the system of record before training or retrieval
Knowledge base and policy portal give different answers	Retire or reconcile duplicates before connecting a GenAI assistant
Manual spreadsheets hold the "real" data	The AI would learn from the official, not the real, version (fidelity issue)

2.7 Training data and labelled outcome records

A machine-learning system learns from historical examples paired with their outcomes. Recall the label test from Course 2: the outcome must be recorded, with sufficient coverage and consistency, without leakage or selection bias. At readiness stage, verify it on real extracts.

Use type	What it learns from / relies on	Readiness checks
Predictive ML	Historical inputs + recorded outcomes (labels)	Label coverage, consistency, timing, selection bias, volume, recency, representativeness
Classification	Labelled examples per class	Class balance; labelling guidelines; inter-labeller agreement
Generative AI with retrieval	Enterprise knowledge content	Currency, authority (approved sources), duplication, permissions, coverage of the questions users ask
Fine-tuned GenAI	Curated example inputs and approved outputs	Quality and rights of examples; representativeness
Evaluation (all types)	A held-out, representative test set with known-good answers	Exists? Independent of training? Reflects production?

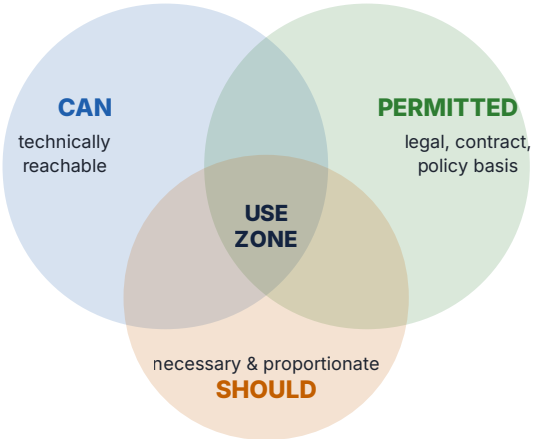
2.8 Data fidelity: does the record reflect reality?

Data fidelity is how faithfully recorded data represents what actually happened. Data can be complete, valid and consistent — and still unfaithful, because it was recorded for a different purpose, under different incentives, or through workarounds.

Fidelity problem	Example	Consequence for AI
Proxy records	“Ticket closed” recorded when the customer stopped replying, not when the problem was solved	A model learns to predict closure, not resolution
Incentive distortion	Agents choose the code that avoids escalation targets	Labels reflect behaviour under targets, not reality
Default and catch-all values	“Other” or the first drop-down option chosen to save time	Patterns are noise
Off-system reality	Real decisions made in e-mail or spreadsheets	System history misses the actual logic
Process change	Coding rules changed mid-history	Old and new labels mean different things

> **CONSULTANT'S TIP · Testing fidelity**
> Walk a sample of 20–30 records back to reality: read the underlying e-mails, inspect the asset, ask the person who handled the case. Record the share where the data told the true story.

2.9 Can access, should access, permitted to access



Outside the use zone:

- Can, not permitted → breach
- Can + permitted, not should → over-collection, trust loss
- Should + permitted, cannot → a technology gap to fix
- GenAI assistants inherit "can" — oversharing risk

Figure 2.2 — Three different questions about data access

Question	Who answers	Example
Can we access it?	IT, data engineering	The call recordings are technically reachable in the telephony archive
Are we permitted to?	Legal, DPO, contracts, policy owners	Customers were told recordings are for quality and training of staff; is AI analysis compatible with that purpose? Works agreement on agent recordings?
Should we?	Business owner with ethics / risk input	Even if permitted, is analysing agents' emotional tone necessary and proportionate for the use case?

KEY IDEA

The use zone is where all three answers are yes. Technical reach is the weakest of the three — and the one most often mistaken for permission.

2.10 Tiered controls for higher-stakes and regulated data

Data tier	Examples	Controls for AI use (illustrative)
Public	Published product information	Standard quality checks
Internal	Internal procedures, non-personal operational data	Access by role; approved tools only
Confidential	Customer personal data, contracts, financials	Legal basis confirmed; pseudonymisation where possible; EU/approved hosting; no vendor training on data; logging
Restricted / special category	Health, biometric, union membership, criminal data; trade secrets; regulated data	DPIA; explicit approval; isolated environment; minimisation; strict access and logging; enhanced vendor terms; human review of outputs

NOTE

Use the organization's existing classification scheme where one exists; the assessment checks that AI use is mapped onto it and that controls **operate** (e.g., pseudonymisation actually applied in the pipeline), not just that a policy says so.

2.11 Technology readiness

Area	Ready when…	Typical gap
Integration	Outputs reach the systems where work happens; inputs flow at required latency	Standalone tool; copy-paste
Infrastructure & hosting	Capacity, resilience and data residency meet requirements	Pilot ran on a developer's cloud account
Identity & security	SSO, role-based access, secrets management, AI-specific threats considered	Shared credentials; no prompt-injection testing
Run tooling (MLOps / LLMOps)	Versioning of models, prompts and data; evaluation; monitoring; rollback	No way to know what changed when performance dropped
Logging	Inputs, outputs, overrides and changes logged and retained per policy	Logs only in the vendor's console, short retention
Scalability & cost	Performance and unit cost at production volume are known	Cost per transaction unknown at scale
Resilience & fallback	Process continues if the AI is unavailable or switched off	No manual fallback

2.12 Ownership of data assets and of the deployed system

Owner	Accountable for	Assessment question
Data owner	Fitness, permitted use and quality of a data asset	Has the owner approved this AI use of their data?
System owner (business)	The deployed AI system's purpose, performance and use in the process	Who decides to change, pause or retire it?
Technical owner	Operation, security, monitoring, changes	Who is on call; who approves model/prompt changes?
Provider (internal or vendor)	The model/product as supplied	What has the provider committed to (documentation, change notice, support)?

PITFALL

"IT owns it" usually means IT runs the infrastructure but nobody owns the *use*. Ownership of the deployed system must sit with a business role that can be held accountable for outcomes.

2.13 Writing a data readiness finding a sponsor can act on

A finding is not an observation. It connects what you found to the standard, the evidence, the business consequence and a specific action.

Element	Content
Condition	What is true now — specific and measured
Criterion	What it should be, by the recorded standard and target state
Evidence (and grade)	How you know: profile, sample, document, interview — graded (Module 4)
Consequence	What happens to the business if unaddressed: cost, risk, delay, lost value
Recommendation	Specific remediation, owner, effort, timing — and whether it blocks go-live

EXAMPLE · Weak vs actionable

Weak: "Data quality issues exist in maintenance records."

Actionable: "**Condition:** 37% of 2025 work orders have no failure code and 21% of the rest use 'OTHER' (profile of 48,210 records, 12 March). **Criterion:** the failure-prediction target state needs ≥ 85% usable failure codes for the last 24 months. **Consequence:** the model cannot learn failure modes for 4 of 6 asset classes; the projected €1.2m downtime saving is not achievable on current data. **Recommendation:** mandatory coded failure field in the mobile work-order app (owner: Head of Maintenance Ops; 6 weeks) and 6 months of clean capture before model training; meanwhile pilot on the two asset classes with adequate history. **Blocks go-live** for the full scope."

2.14 Practice scenario: Orbis Facilities

CASE STUDY **Readiness assessment for predictive maintenance and a technician assistant**

Orbis Facilities (fictional) · facility management for 1,400 buildings · 3,000 technicians

Target state (from baseline)

(a) Predict HVAC failures 7 days ahead for 1,400 buildings; (b) a GenAI assistant answering technicians' questions from manuals and past work orders on mobile devices.

Fieldwork results

- **Quality:** sensor data complete for 60% of buildings (newer contracts); timestamps in two time zones unreconciled.
- **Labels:** failures recorded in work orders; 37% lack failure codes (see example finding).
- **Fidelity:** technicians often close orders as "resolved" at shift end even when a return visit is booked.
- **Single source of truth:** asset register in ERP and a separate spreadsheet per region; 14% of assets mismatch.
- **Provenance:** sensor data in 30% of buildings belongs to building owners; contracts silent on secondary use for AI.
- **Access:** manuals repository shared "all employees", including 2,000 client contract PDFs with pricing.
- **Technology:** mobile app can call an API; no logging design; vendor model hosted outside the EU by default.
- **Ownership:** no named owner for the assistant; data owner for sensor data unclear.

Your task

1. Classify each fieldwork result under the module's headings and decide whether it blocks go-live for (a), (b) or both.
2. Apply can / should / permitted to the building-owner sensor data.
3. Write one actionable finding for the manuals repository.

Model answers (brief)

1. (a) blocked by labels, fidelity ("resolved"), asset-register mismatch and sensor provenance for 30% of buildings — pilot on owned-sensor, coded-history buildings is possible. (b) blocked by the oversharing of client contracts, hosting location for confidential data, missing logging and no owner; content quality otherwise adequate.
2. Can: yes, data is reachable. Permitted: unclear — contracts silent; legal must determine. Should: proportionate if limited to equipment telemetry; exclude occupancy data not needed for HVAC failure prediction.
3. "**Condition:** the manuals repository the assistant will index is shared with all employees and contains ~2,000 client contracts with pricing. **Criterion:** Information Classification Policy v3 — client contracts are Confidential, need-to-know. **Consequence:** any technician could retrieve client pricing through the assistant — contractual breach and commercial exposure. **Recommendation:** remove

contracts from the indexed scope and fix permissions before connection (owner: Head of IT Ops, 2 weeks); permission-aware retrieval as a technical requirement. **Blocks go-live of (b).**”

2.15 Consulting templates

T2.1 Data Quality & Fidelity Profile

Purpose: Turn assurances into measured evidence for the fields the use case needs.

Use case / dataset / extract date / record count: .....

Field / entity	Needed for	Completeness %	Validity / consistency issues	Timeliness (lag)	Fidelity sample (n, % faithful)	Status

System of record for each critical entity: .....

Conflicting sources found: .....

T2.2 Lineage, Provenance & Labels Record

Purpose: Trace where data comes from, the rights attached, and what the model learns from.

PROVENANCE & LINEAGE

Data asset	Origin	Rights / basis (consent, licence, contract)	Hops & transformations	Lineage evidence

LABELS / REFERENCE CONTENT

- ☐ Target or reference content defined precisely
- ☐ Coverage adequate for the population
- ☐ Consistent definitions over time and units
- ☐ No leakage (inputs available at decision time)
- ☐ No material selection bias
- ☐ Held-out, representative evaluation set exists

T2.3 Access Triad & Tier Controls Check

Purpose: Test can / permitted / should, and whether tier controls operate.

Data asset	Tier	CAN (who confirmed)	PERMITTED (basis, who confirmed)	SHOULD (necessary & proportionate?)	Controls required for tier	Controls operating? (evidence)

RETRIEVAL-SPECIFIC

- ☐ Permission-aware retrieval (end-user permissions enforced)
- ☐ Oversharing review of connected repositories completed
- ☐ Agent / service-account credentials scoped and logged
- ☐ Vendor training on our data excluded contractually

T2.4 Technology Readiness & Ownership Checklist

Purpose: Assess production technology and name owners.

TECHNOLOGY

Area	Ready? (Y / Partial / N)	Evidence	Gap / action
Integration (in & out)			
Hosting & data residency			
Identity & security (incl. AI threats)			
Versioning & evaluation (MLOps/LLMOps)			
Monitoring & logging			
Scale & unit cost			
Fallback if switched off			

OWNERSHIP

Data owner(s) — approval of this AI use: .....

System owner (business): .....

Technical owner: .....

Provider commitments (docs, change notice, support): .....

T2.5 Data Readiness Finding

Purpose: Write a finding a sponsor can act on.

Finding title: .....

Condition (specific, measured): .....

Criterion (standard / target state reference): .....

Evidence and grade: .....

How learned (method, date, sample): .....

Business consequence: .....

Recommendation (action, owner, effort, date): .....

Blocks go-live? ☐ yes ☐ partial scope ☐ no .....

Confidence (H / M / L): .....

2.16 Best practices and pitfalls

✓ DO THIS

- Assess data against the recorded target state.
- Profile real extracts rather than collecting assurances.
- Walk a sample of records back to reality (fidelity).
- Separate can, should and permitted — and get each from the right people.
- Review permissions before connecting GenAI assistants.
- Check that tier controls operate, not just exist.
- Write findings with condition, criterion, evidence, consequence, recommendation.

× AVOID THIS

- "Our data is good" as a general verdict.
- Treating a passed governance audit as AI readiness.
- Confusing provenance with lineage.
- Assuming technical access implies permission.
- Letting a service account's rights define what users see.
- Assessing the demo stack instead of the production stack.
- Findings that describe problems without consequences or owners.

2.17 Module 2 cheat sheet

MODULE 2 · ONE-PAGE SUMMARY

Data and Technology Readiness

FIT FOR PURPOSE

Same data, different readiness per use case. Assess against the target state (task, decision, latency, volume, population).

QUALITY

Accuracy · completeness · consistency · timeliness · validity · uniqueness. **Profile, don't ask.**

GOVERNANCE

Owner & stewards · definitions · quality monitoring · AI-use policy · retention incl. AI copies · change management.

PROVENANCE VS LINEAGE

Provenance: origin + rights. **Lineage**: path + transformations. Both needed for trust, rights, debugging, audit.

ACCESS

Least privilege · permission-aware retrieval · oversharing review · scoped agent credentials · vendor limits. AI makes bad permissions discoverable.

SINGLE SOURCE OF TRUTH

Designated system of record per critical entity, or a reconciled master view — before training or retrieval.

LABELS & CONTENT

ML: recorded outcomes (coverage, consistency, timing, selection bias). GenAI: current, authoritative, permissioned content. Held-out evaluation set.

FIDELITY

Does the record reflect reality? Proxies, incentives, defaults, off-system decisions, rule changes. Walk records back to reality.

CAN · PERMITTED · SHOULD

Technical · legal/contract/policy · necessary & proportionate. Use only where all three are yes.

TIERS & TECHNOLOGY

Public · internal · confidential · restricted — controls must operate. Tech: integration, hosting, identity/security, versioning, monitoring/logging, scale/cost, fallback.

OWNERSHIP & FINDINGS

Data owner · system owner (business) · technical owner · provider. Finding = **condition · criterion · evidence (graded) · consequence · recommendation** (owner, date, blocks go-live?).

2.18 Knowledge check

Answer all questions before turning the page.

K1 **A client says its customer data "passed last year's data-governance audit". What is the right assessment response?**

- (A) Ask for a second audit.
- (B) Accept it as evidence of data readiness.
- (C) Test the specific fields, history, latency and labels the use case needs against its target state.
- (D) Ignore the audit entirely.

K2 **What is the difference between data provenance and data lineage?**

- (A) Provenance is the origin and rights attached to data; lineage is its path and transformations through systems.
- (B) Lineage is only relevant for GenAI.
- (C) They are synonyms.
- (D) Provenance concerns quality; lineage concerns security.

K3 **A GenAI assistant shows a junior employee a board compensation file. What is the root readiness issue?**

- (A) The model hallucinated.
- (B) Over-broad permissions on connected repositories that the assistant made discoverable
- (C) Poor data quality
- (D) Lack of AI literacy training

K4 **Service tickets are marked "closed" when customers stop replying, not when problems are solved. A model is to predict "resolution". Which issue is this?**

- (A) Timeliness
- (B) Uniqueness
- (C) Completeness
- (D) Data fidelity: the record is a proxy that does not reflect reality

K5 **Call recordings are technically reachable and the legal team confirms AI analysis is compatible with the stated purpose. Is the access question settled?**

- (A) Yes — can and permitted are both satisfied.
- (B) No — whether the analysis is necessary and proportionate for the use case (should) still has to be answered.
- (C) Yes — legal approval covers ethics too.
- (D) No — technical access must be removed.

K6 **Which evidence best shows that tier controls for restricted health data are operating?**

- (A) A pipeline sample showing pseudonymised fields, access logs and the approved DPIA
- (B) A policy document requiring pseudonymisation
- (C) The vendor's marketing statement about security
- (D) A training slide on data protection

K7 **Which is a technology-readiness gap rather than a data gap?**

(A) Two systems hold different asset lists
(B) 37% of records lack failure codes
(C) Labels exist only for investigated cases
(D) No way to roll back a prompt change after performance drops

K8 **Which statement is the most actionable data readiness finding?**

(A) “Data is not AI-ready.”
(B) “Buy a data quality tool.”
(C) “37% of 2025 work orders lack failure codes (profile, n=48,210) against a ≥85% requirement; the model cannot learn 4 of 6 failure modes, so the €1.2m saving is unreachable; mandate coded capture (Maintenance Ops, 6 weeks); blocks full-scope go-live.”
(D) “Data quality could be improved.”

2.19 Knowledge check: answers and explanations

K1 **Answer: C** Review: 2.1

A client says its customer data "passed last year's data-governance audit". What is the right assessment response?

- ✗ (A) Ask for a second audit. — **Wrong.** Another generic audit would have the same limitation.
- ✗ (B) Accept it as evidence of data readiness. — **Wrong.** The audit used its own criteria, not this use case's fields, latency, history and labels.
- ✓ (C) Test the specific fields, history, latency and labels the use case needs against its target state. — **Correct.** Data readiness is fitness for a specific purpose.
- ✗ (D) Ignore the audit entirely. — **Wrong.** It is useful context, just not sufficient.

K2 **Answer: A** Review: 2.4

What is the difference between data provenance and data lineage?

- ✓ (A) Provenance is the origin and rights attached to data; lineage is its path and transformations through systems. — **Correct.** This is the distinction.
- ✗ (B) Lineage is only relevant for GenAI. — **Wrong.** Lineage matters for all AI.
- ✗ (C) They are synonyms. — **Wrong.** They answer different questions.
- ✗ (D) Provenance concerns quality; lineage concerns security. — **Wrong.** Neither definition is correct.

K3 **Answer: B** Review: 2.5

A GenAI assistant shows a junior employee a board compensation file. What is the root readiness issue?

- ✗ (A) The model hallucinated. — **Wrong.** The file is real; the problem is access.
- ✓ (B) Over-broad permissions on connected repositories that the assistant made discoverable — **Correct.** Assistants retrieve what permissions allow; oversharing must be fixed before connection.
- ✗ (C) Poor data quality — **Wrong.** The data was accurate.
- ✗ (D) Lack of AI literacy training — **Wrong.** Training does not fix permissions.

K4 **Answer: D** Review: 2.8

Service tickets are marked "closed" when customers stop replying, not when problems are solved. A model is to predict "resolution". Which issue is this?

- ✗ (A) Timeliness — **Wrong.** The lag is not the issue.
- ✗ (B) Uniqueness — **Wrong.** Duplicates are not the issue.
- ✗ (C) Completeness — **Wrong.** The field is filled.
- ✓ (D) Data fidelity: the record is a proxy that does not reflect reality — **Correct.** The model would learn to predict closure, not resolution.

K5 **Answer: B** Review: 2.9

Call recordings are technically reachable and the legal team confirms AI analysis is compatible with the stated purpose. Is the access question settled?

- ✗ (A) Yes — can and permitted are both satisfied. — **Wrong.** The "should" question remains.
- ✓ (B) No — whether the analysis is necessary and proportionate for the use case (should) still has to be answered. — **Correct.** All three questions must be yes.
- ✗ (C) Yes — legal approval covers ethics too. — **Wrong.** Permission and proportionality are different questions.
- ✗ (D) No — technical access must be removed. — **Wrong.** Technical access is fine.

K6 **Answer: A** Review: 2.10

Which evidence best shows that tier controls for restricted health data are operating?

✓ **A** A pipeline sample showing pseudonymised fields, access logs and the approved DPIA — **Correct.** These show controls applied in practice.

✗ B A policy document requiring pseudonymisation — **Wrong.** This shows intent, not operation.

✗ C The vendor's marketing statement about security — **Wrong.** This is an assertion, not evidence of operation.

✗ D A training slide on data protection — **Wrong.** Training does not prove controls operate.

K7 **Answer: D** Review: 2.11

Which is a technology-readiness gap rather than a data gap?

✗ A Two systems hold different asset lists — **Wrong.** A single-source-of-truth data gap.

✗ B 37% of records lack failure codes — **Wrong.** A data quality/label gap.

✗ C Labels exist only for investigated cases — **Wrong.** A label / selection-bias gap.

✓ **D** No way to roll back a prompt change after performance drops — **Correct.** A run-tooling (versioning/LLMOps) gap in technology readiness.

K8 **Answer: C** Review: 2.13

Which statement is the most actionable data readiness finding?

✗ A "Data is not AI-ready." — **Wrong.** A verdict without evidence or action.

✗ B "Buy a data quality tool." — **Wrong.** A solution without diagnosis.

✓ **C** "37% of 2025 work orders lack failure codes (profile, n=48,210) against a ≥85% requirement; the model cannot learn 4 of 6 failure modes, so the €1.2m saving is unreachable; mandate coded capture (Maintenance Ops, 6 weeks); blocks full-scope go-live." — **Correct.** It has condition, evidence, criterion, consequence, owner, timing and go-live impact.

✗ D "Data quality could be improved." — **Wrong.** No condition, criterion, consequence or owner.

MODULE 3

Organizational Readiness

Organizational readiness — not technology — usually decides whether AI delivers value. Establish what readiness is measured against, map the roles and control points a change touches, test operating conditions rather than assurances, and close with a finding: proceed, phase or not yet.

YOU WILL BE ABLE TO

- Establish what organizational readiness is being measured against
- Map the roles and control points an AI change touches
- Test workforce proficiency, oversight, accountability, policy clarity and reinforcement as operating conditions, not assurances
- Assess human-in-the-loop review design and AI governance roles
- Assess change and adoption readiness, including shadow AI, frontline managers, resistance and rework
- Check that value capture is designed, not assumed
- Write a readiness finding — proceed, phase or not yet — with named remediation and a sequence

IN THIS MODULE

1. Why organizational readiness decides value
2. Readiness for what?
3. Roles and control points
4. Operating conditions vs assurances
5. Workforce AI proficiency
6. Human oversight and review design
7. Accountability and ownership
8. AI policy clarity
9. Reinforcement and frontline managers
10. Change, adoption, resistance and shadow AI
11. Rework and value capture
12. The readiness finding
13. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4–5 hours

Technology assessments routinely miss what usually decides AI value: whether people, roles, controls and incentives will actually change. This module teaches you to assess organizational readiness as **operating conditions you can observe and test**, not as assurances from leaders — and to close with a finding that is honest, sequenced and actionable.

KEY TAKEAWAYS

1. Start by fixing **what readiness is measured against**: the target operating state for this capability.
2. Map the **roles and control points** the change touches — especially checks that AI could quietly weaken.
3. Test **operating conditions** (observed, sampled, recorded) rather than accepting **assurances** ("we will train everyone").
4. Assess **workforce AI proficiency**, **human oversight design**, **accountability**, **policy clarity** and **reinforcement**.
5. **Frontline managers** decide adoption; **shadow AI**, **resistance** and **rework** are readiness signals.
6. Value must be **captured by design** — someone must bank it.
7. Conclude with **proceed, phase or not yet** — each with named remediation, owners and a sequence.

3.1 Why organizational readiness decides value

An AI capability creates value only when people change what they do: they use the output, check it properly, act on it, and stop doing the work it replaces. Technology assessments look at whether the system *can* work; organizational assessments look at whether the organization *will* work differently. The second is harder to see and is usually where value is lost.

KEY IDEA

Most AI value losses are **organizational**: outputs ignored, reviews rubber-stamped, work duplicated, saved time absorbed, owners missing. None of them shows up in a model evaluation.

3.2 Establishing what readiness is measured against

"Is the organization ready?" has no answer until you state **ready for what**. Write the organizational target state alongside the baseline from Module 1.

Target-state element	Example (credit-memo drafting assistant)
Process change	Analysts review and edit AI-drafted memos instead of writing from scratch
Autonomy level	Level 2 — draft for review; no memo reaches committee without analyst sign-off
Scale and timing	120 analysts in 4 regions; 2,400 memos/quarter; from Q3
Oversight design	Analyst verifies every figure against source documents; quality team samples 10%
Roles affected	Relationship managers, analysts, team leads, credit committee, model risk
Value mechanism	Memo time −40%; capacity redeployed to portfolio reviews (backlog of 600)
Standard	Bank AI Policy v2; Model Risk Policy; applicable regulation as of the assessment date

3.3 Mapping the roles and control points a change touches

A **control point** is any place in a process where something is checked, approved or decided: a four-eyes review, a sign-off, a quality sample, an escalation threshold. AI changes who does what at these points — and can quietly hollow them out. A reviewer who used to *write* now *reviews*; a second reviewer may now see the same AI draft as the first, so the four-eyes check is no longer independent.

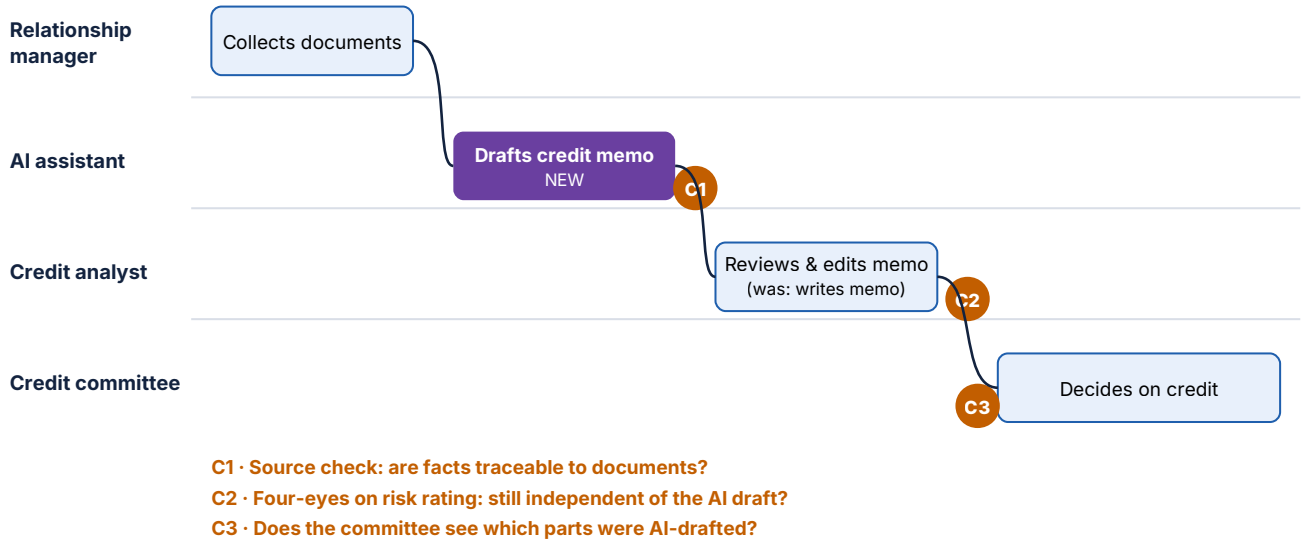


Figure 3.1 — Role and control-point map for a credit-memo assistant

Question for each control point	Why
Does it still exist after the change?	Steps silently disappear when AI "handles" them
Is it still independent?	Two reviewers looking at one AI draft may share its errors
Does the person have the information to check?	Without sources, review becomes reading for fluency
Do they have time and authority to reject?	Time pressure produces rubber-stamping
Is the decision-maker told what was AI-generated?	Transparency for accountable decisions

3.4 Operating conditions vs assurances

Organizational readiness is full of assurances: "we will train everyone", "humans always review", "managers are on board". An assurance describes an intention. An **operating condition** describes something true now that you can observe or verify.

Assurance (not evidence)	Operating condition (evidence)	How to test
"Everyone will be trained."	84% of pilot users completed role-based training; 76% passed a scenario check	Training records; scenario test
"Humans review every output."	Review logs show median 40 s per memo; 3% edited	System logs; observation
"The policy is clear."	8 of 10 sampled users correctly answered 5 policy scenarios	Scenario interviews

Assurance (not evidence)	Operating condition (evidence)	How to test
“Managers support it.”	Team KPIs updated; 5 of 6 team leads hold weekly usage reviews	KPI documents; meeting notes
“The business owns it.”	Owner named in system register; owner chaired last incident review	Register; minutes

EXAM TRAP

When a question offers “the COO confirmed staff will be trained before go-live” as evidence of readiness, it is an **assurance**. The better answer tests a current condition: training records, scenario results, observed use.

3.5 Workforce AI proficiency

Proficiency level	Can do	Needed by
Aware	Knows what the system does, its limits and the rules of use	Everyone affected
User	Uses it effectively in workflow; recognises poor outputs	Daily users
Overseer	Verifies outputs critically, knows failure modes, overrides and escalates confidently	Reviewers, approvers
Owner / steward	Monitors performance, manages changes and incidents	System owners, team leads
Builder	Configures, prompts, evaluates and improves	Technical teams

Test proficiency with **scenarios**, not attendance: give users three outputs, one with a subtle error, and see who catches it. The EU AI Act’s AI-literacy duty (Module 1) makes role-appropriate competence an expectation, not a nice-to-have.

3.6 Human oversight and human-in-the-loop review design

“Human in the loop” is only meaningful if the human can actually catch and correct errors. Oversight fails in predictable ways — above all **automation bias**: the tendency to accept a fluent, confident output, especially under time pressure.

Condition for effective oversight	Test
Competence	Reviewers have domain skill and know the system’s failure modes
Authority	They can reject, override or stop — without penalty
Time	Review time is budgeted realistically in the process and targets
Information	Sources, confidence indicators or highlights let them verify, not just read
Independence	Second-line checks are not reviewing the same unverified draft
Feedback	Corrections feed back into improvement; error rates are monitored

Review design	When it fits
Pre-release review of every output	High-consequence outputs (decisions about people, money, safety)

Review design	When it fits
Risk-based / exception review	High volume; system flags uncertain or high-risk cases
Sampling review	Low-consequence outputs; quality monitoring
Escalation thresholds	Agentic actions above value or risk limits require approval

PITFALL · Signs of rubber-stamping

Very short review times, near-zero edit or rejection rates, identical approval patterns across reviewers, and reviewers unable to explain what they checked. Each is a readiness finding against the oversight dimension.

NOTE · AI governance roles

Typical roles: an accountable executive; an AI governance or risk committee; system owners; data owners; model-risk or validation function; compliance and DPO; and named overseers in operations. Readiness requires these roles to be **staffed and active** for the system in scope — not just drawn on an organisation chart.

3.7 Accountability and ownership

Test accountability with the question: **"When this system produces a harmful output next Tuesday, who answers for it — and what will they do?"** If the answer is a committee, a vendor or "IT", accountability is not ready.

Role	Ready when…
Accountable business owner	Named, accepts outcomes, has budget and authority to change or stop the system
Operational overseers	Named per team, with defined duties and time
Escalation path	Documented and known; tested at least once (table-top exercise)
Incident handling	Who decides to pause the system; who informs affected people and authorities if needed

3.8 AI policy clarity

An AI policy is ready when people **know** it, **understand** it and can **apply** it to real situations. Test with scenarios: "May you paste a client's balance sheet into the assistant?" "May you send an AI-drafted memo without checking the figures?" Count correct answers. Ambiguous policies produce two failure modes: over-caution (no adoption) and over-use (shadow risk).

CONSULTANT'S TIP

Policies written as principles ("use AI responsibly") score badly in scenario tests. Policies with concrete rules, examples and a named contact for questions score well.

3.9 Reinforcement and frontline managers

Behaviour change sticks when it is **reinforced**: targets, incentives, routines and leadership attention that reward the new way of working. Frontline managers are the decisive group — they set daily priorities, interpret targets and model behaviour.

Reinforcement mechanism	Ready when...
Targets and KPIs	Updated so they don't reward the old process (e.g., memo count without quality)
Manager routines	Team leads review usage, quality and issues regularly
Recognition	Good use and good catches (errors found) are recognised
Process embedding	The new step is in procedures, systems and onboarding
Leadership signals	Senior leaders use, discuss and ask about it

3.10 Change and adoption readiness, resistance and shadow AI

Signal	What it may indicate	Assessment move
Resistance: "it makes more work"	Real rework or poor fit	Measure end-to-end time, not just the AI step
Resistance: "it threatens my expertise / job"	Identity or job-security concern	Check communication on role change; involve works council
Low use after launch	Poor fit, low trust or no reinforcement	Observe users; interview frontline managers
High use, no edits	Over-reliance / rubber-stamping	Sample outputs for errors
Shadow AI in the same teams	Unmet need, or sanctioned tool inferior	Treat as demand evidence and governance risk

> **KEY IDEA**
> Shadow AI is a **readiness signal in both directions**: it shows appetite and skills (adoption potential) and weak governance (control gap). Read it, don't only police it.

Training and enablement

Ready enablement is **role-based** (users vs overseers vs owners), **scenario-led** (including spotting errors), **timed** close to go-live, supported by **champions** and quick help, and **measured** by proficiency, not attendance.

3.11 Rework and value capture

Rework is the hidden tax on AI: drafts rewritten, outputs double-checked, errors corrected downstream. It is the ledger from Course 2, observed in operation. Measure it in pilots (edit rate, time spent correcting, downstream error rates).

Value capture asks who will bank the benefit. Saved time is absorbed silently unless someone decides in advance what happens to it: reduced overtime or contractors, a backlog cleared, a target raised, a hiring plan changed.

Value-capture check	Ready when...
Mechanism defined	The business case states how benefit becomes cost, revenue or capacity redeployed
Owner named	A manager is accountable for realising it

Value-capture check	Ready when...
Plans adjusted	Budgets, targets or staffing plans reflect the expected change
Tracked	Baseline and benefit tracking are in place (measurement dimension)

3.12 The readiness finding: proceed, phase or not yet

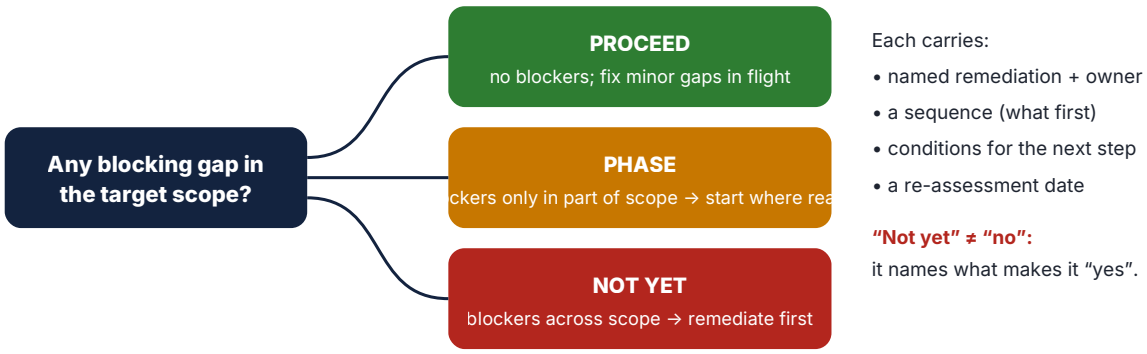


Figure 3.2 — From conditions to a readiness finding

Finding	Use when	Must include
Proceed	No blocking gaps in the target scope; remaining gaps can be closed in flight	In-flight remediation with owners and dates; monitoring conditions
Phase	Blocking gaps affect part of the scope (units, use types, autonomy level)	The ready phase defined; conditions to expand; sequence of remediation
Not yet	Blocking gaps across the scope	Named remediation, owners, sequence, criteria for "yes", re-assessment date

Sequencing remediation

1. **Ownership and decision rights first** — without an owner, nothing else gets fixed.
2. **Oversight design and policy clarity** before scaling exposure.
3. **Training and enablement** timed to go-live of each phase.
4. **Reinforcement and value capture** in place before the first phase ends.
5. **Re-assess** at the defined date or trigger.

EXAMPLE · A phased finding

Finding: PHASE. Proceed in Region North (owner named, team leads active, oversight design tested: median review 6 min, 22% edits). Not yet for Regions South and East: no named team-level overseers, KPIs still reward memo volume. **Remediation:** (1) appoint regional owners — Head of Credit, by 30 April; (2) revise analyst KPIs to include quality — HR and Credit, by 31 May; (3) role-based overseer training — L&D, June. **Expansion criterion:** review time ≥ 4 min median and ≥ 10% edit rate in sample. **Re-assess:** 1 July.

3.13 Case study: Lindqvist Bank

CASE STUDY **The memo assistant everyone said was ready**

Lindqvist Bank (fictional) · corporate banking · 120 credit analysts

Situation

A three-month pilot of a credit-memo drafting assistant in one region reported “40% time saving” and high satisfaction. The Chief Credit Officer wants a group-wide launch in Q3 and asks for a readiness assessment “to tick the box”.

What the assessor found

- **Control points:** the four-eyes check on risk ratings now saw the same AI draft as the analyst; the second reviewer’s independent view was lost.
- **Oversight:** review logs showed median 38 seconds per memo and 2% edit rate in the last month of the pilot — down from 9 minutes and 30% in month one.
- **Sample:** in 50 sampled memos, 6 contained figures not found in source documents; none had been corrected.
- **Policy:** 5 of 12 analysts thought client financials could be pasted into a public chatbot “for formatting”.
- **Reinforcement:** analyst KPIs still counted memos produced; two team leads said they had “no time” for quality reviews.
- **Value capture:** no decision on the freed time; the portfolio-review backlog had not changed.
- **Ownership:** the innovation lab owned the pilot; no credit executive had accepted system ownership.

Finding: NOT YET (with a defined route to “yes”)

Remediation sequence: (1) Credit executive named as system owner (2 weeks); (2) redesign four-eyes control so the second reviewer assesses risk independently of the draft (Credit Risk, 4 weeks); (3) source-linked figures in the assistant (vendor and IT, 6 weeks); (4) KPI change to quality-weighted measures and team-lead review routine (HR and Credit, 6 weeks); (5) overseer training with error-spotting scenarios (L&D); (6) value-capture decision: freed time allocated to the portfolio-review backlog. **Criteria for yes:** sampled unsupported-figure rate below 1%; median review time ≥ 4 minutes. **Re-assessment:** 10 weeks.

Discussion questions

1. Which pilot results were assurances and which were operating conditions?
2. What does the drop in review time and edit rate indicate?
3. Why is ownership first in the sequence?
4. Why “not yet” rather than “phase”?

Model answers (brief)

1. “40% saving” and satisfaction were self-reported assurances; review logs, the memo sample, the policy scenario test, KPI documents and the backlog figure were operating conditions.
2. Automation bias: reviewers increasingly trust fluent drafts, and oversight has degraded into rubber-stamping.

3. Without an accountable owner, no one has the authority to change controls, KPIs or the vendor configuration.
4. The blocking gaps (oversight, control independence, ownership) applied across all regions, including the pilot region — there was no ready subset to start with.

3.14 Consulting templates

T3.1 Organizational Target State & Control-Point Map

Purpose: Fix what readiness is measured against and where controls sit.

TARGET STATE

Process change: .....

Autonomy level and oversight design: .....

Scale, users, timing: .....

Value mechanism: .....

Standard reference (baseline record): .....

CONTROL POINTS

Control point	Before AI	After AI	Still exists?	Independent?	Info, time, authority to reject?	Issue

T3.2 Operating Condition Test Sheet

Purpose: Convert assurances into tested conditions.

Area	Assurance heard	Condition tested	Method (records / logs / observation / scenario)	Result	Ready?
Workforce proficiency					
Human oversight					
Accountability & ownership					
AI policy clarity					
Reinforcement / KPIs					
Frontline manager routines					
Value capture					

T3.3 Human Oversight Design Check

Purpose: Test whether human review can actually catch and correct errors.

Output type and consequence: .....

Review design (every output / risk-based / sample / escalation threshold): .....

- ☐ Competence — reviewers know domain and failure modes
- ☐ Authority — can reject / override / stop without penalty
- ☐ Time — realistic review time budgeted in targets
- ☐ Information — sources or highlights available to verify
- ☐ Independence — second-line checks not reviewing the same unverified draft
- ☐ Feedback — corrections captured and error rates monitored

RUBBER-STAMPING INDICATORS

Median review time: .....

Edit / rejection rate: .....

Sample error rate (n=): .....

T3.4 Adoption & Resistance Scan

Purpose: Read adoption signals, resistance and shadow AI.

Team / role	Usage level	Edit rate	Resistance signals (what, why)	Shadow AI observed	Manager routine?	Action

Rework observed (where, how much): .....

Training: role-based? scenario-tested? timed to go-live? .....

T3.5 Organizational Readiness Finding

Purpose: Close with proceed, phase or not yet — sequenced.

Finding: ☐ Proceed ☐ Phase ☐ Not yet .....

Ready scope (if phased): .....

Blocking gaps (dimension · evidence): .....

REMEDIATION SEQUENCE

#	Action	Owner	By when	Unblocks

Criteria for next step / “yes”: .....

Re-assessment date or trigger: .....

3.15 Best practices and pitfalls

✓ DO THIS

- Write the organizational target state before assessing.
- Map control points and test their independence after the change.
- Convert every assurance into a tested condition.
- Test proficiency and policy with scenarios.
- Look for rubber-stamping in review logs and samples.
- Interview frontline managers; check KPIs.
- End with proceed / phase / not yet, sequenced and owned.

× AVOID THIS

- Accepting “we will train everyone” as readiness.
- Assuming human-in-the-loop means effective oversight.
- Leaving ownership with an innovation lab.
- Ignoring incentives that reward the old process.
- Treating shadow AI only as a compliance issue.
- Counting time saved without a value-capture decision.
- A “not yet” without a route to “yes”.

3.16 Module 3 cheat sheet

MODULE 3 · ONE-PAGE SUMMARY

Organizational Readiness

READINESS FOR WHAT?

Target state: process change · autonomy & oversight · scale & timing · roles · value mechanism · standard.

CONTROL POINTS

Still exists? Independent? Info, time and authority to reject? Decision-maker told what is AI-generated?

CONDITIONS, NOT ASSURANCES

Assurance = intention ("we will…"). Condition = verifiable now (records, logs, observation, scenario tests). Test: proficiency · oversight · accountability · policy clarity · reinforcement · managers · value capture.

PROFICIENCY LEVELS

Aware · user · overseer · owner/steward · builder. Test with error-spotting scenarios. AI literacy expected (EU AI Act).

EFFECTIVE OVERSIGHT

Competence · authority · time · information · independence · feedback. Rubber-stamp signs: short reviews, ~0 edits, sample errors uncorrected (automation bias).

ACCOUNTABILITY & POLICY

"Who answers next Tuesday?" — a named business owner. Policy clear = users answer scenarios correctly.

REINFORCEMENT

KPIs · manager routines · recognition · process embedding · leadership signals. Frontline managers decide adoption.

ADOPTION SIGNALS

Resistance (rework? identity?) · low use · high use no edits · shadow AI = demand + control gap.

REWORK & VALUE CAPTURE

Measure edits and correction time. Decide who banks freed time, how, and track it.

FINDING

Proceed (no blockers) · **Phase** (start where ready) · **Not yet** (blockers across scope). Always: named remediation, owners, sequence (ownership → oversight & policy → training → reinforcement & value capture), criteria for next step, re-assessment date.

3.17 Knowledge check

Answer all questions before turning the page.

K1 **What must be established before organizational readiness can be assessed?**

(A) The organization's overall AI maturity level
(B) Employee satisfaction scores
(C) The target operating state: process change, autonomy and oversight, scale, roles, value mechanism and standard
(D) The vendor's product roadmap

K2 **After an AI drafting tool is introduced, the second reviewer in a four-eyes check reviews the same AI draft as the first. What is the readiness concern?**

(A) There is no concern because two humans still check.
(B) The control point is no longer independent, so shared AI errors can pass both reviewers.
(C) The check takes longer.
(D) The second reviewer is unnecessary.

K3 **Which is an operating condition rather than an assurance?**

(A) "Review logs show median review time of 6 minutes and a 22% edit rate over 400 memos."
(B) "The policy is clear to everyone."
(C) "All staff will complete training before launch."
(D) "Managers are fully committed."

K4 **Review time fell from 9 minutes to 38 seconds over a pilot, and edits fell to 2%; a sample shows uncorrected errors. What does this indicate?**

(A) The model improved.
(B) Users became more efficient.
(C) Training was excellent.
(D) Automation bias: oversight has degraded into rubber-stamping.

K5 **How is AI policy clarity best tested?**

(A) Count policy page views.
(B) Confirm the policy was published on the intranet.
(C) Check that legal approved it.
(D) Ask sampled users to apply the policy to realistic scenarios and count correct answers.

K6 **Analyst KPIs still count memos produced, and team leads have no quality-review routine. Which readiness condition is weak?**

(A) Reinforcement — incentives and manager routines
(B) Model accuracy
(C) Data readiness
(D) Technology readiness

K7 **A pilot saved 2,000 analyst hours, but nothing changed in overtime, backlog or staffing. What is missing?**

(A) More training

(B) Value capture: a defined mechanism and owner to convert saved time into cost reduction or redeployed capacity

(C) A bigger pilot

(D) Model monitoring

K8 **Blocking oversight gaps exist in two of three regions; the third region meets all conditions. Which finding fits?**

(A) Proceed everywhere, fixing gaps in flight

(B) Not yet everywhere

(C) Phase: proceed in the ready region with named remediation and expansion criteria for the others

(D) Reject the use case

3.18 Knowledge check: answers and explanations

K1 **Answer: C** Review: 3.2

What must be established before organizational readiness can be assessed?

- ✗ (A) The organization's overall AI maturity level — **Wrong.** Maturity is generic; readiness is for a specific target.
- ✗ (B) Employee satisfaction scores — **Wrong.** Useful context but not the benchmark.
- ✓ (C) The target operating state: process change, autonomy and oversight, scale, roles, value mechanism and standard — **Correct.** Readiness is always measured against a defined target.
- ✗ (D) The vendor's product roadmap — **Wrong.** Not the benchmark for readiness.

K2 **Answer: B** Review: 3.3

After an AI drafting tool is introduced, the second reviewer in a four-eyes check reviews the same AI draft as the first. What is the readiness concern?

- ✗ (A) There is no concern because two humans still check. — **Wrong.** Two people checking one draft is not an independent check.
- ✓ (B) The control point is no longer independent, so shared AI errors can pass both reviewers. — **Correct.** Independence of control points must be tested after the change.
- ✗ (C) The check takes longer. — **Wrong.** Time is not the main concern.
- ✗ (D) The second reviewer is unnecessary. — **Wrong.** The control may be essential; it needs redesign.

K3 **Answer: A** Review: 3.4

Which is an operating condition rather than an assurance?

- ✓ (A) "Review logs show median review time of 6 minutes and a 22% edit rate over 400 memos." — **Correct.** Observed, verifiable current evidence.
- ✗ (B) "The policy is clear to everyone." — **Wrong.** An assertion that needs a scenario test.
- ✗ (C) "All staff will complete training before launch." — **Wrong.** A future intention.
- ✗ (D) "Managers are fully committed." — **Wrong.** An unverified assertion.

K4 **Answer: D** Review: 3.6

Review time fell from 9 minutes to 38 seconds over a pilot, and edits fell to 2%; a sample shows uncorrected errors. What does this indicate?

- ✗ (A) The model improved. — **Wrong.** The sample shows errors persist.
- ✗ (B) Users became more efficient. — **Wrong.** Uncorrected errors show the speed is not genuine efficiency.
- ✗ (C) Training was excellent. — **Wrong.** Effective training would produce detection, not acceptance.
- ✓ (D) Automation bias: oversight has degraded into rubber-stamping. — **Correct.** These are classic rubber-stamping indicators.

K5 **Answer: D** Review: 3.8

How is AI policy clarity best tested?

- ✗ (A) Count policy page views. — **Wrong.** Views do not show understanding.
- ✗ (B) Confirm the policy was published on the intranet. — **Wrong.** Publication does not show understanding.
- ✗ (C) Check that legal approved it. — **Wrong.** Approval does not show clarity to users.
- ✓ (D) Ask sampled users to apply the policy to realistic scenarios and count correct answers. — **Correct.** Scenario tests show whether people can apply it.

K6 **Answer: A** Review: 3.9

Analyst KPIs still count memos produced, and team leads have no quality-review routine. Which readiness condition is weak?

✓ **A** Reinforcement — incentives and manager routines — **Correct.** KPIs and manager behaviour reinforce the old process.

✗ **B** Model accuracy — **Wrong.** Not a model issue.

✗ **C** Data readiness — **Wrong.** Not a data issue.

✗ **D** Technology readiness — **Wrong.** Not a technology issue.

K7 **Answer: B** Review: 3.11

A pilot saved 2,000 analyst hours, but nothing changed in overtime, backlog or staffing. What is missing?

✗ **A** More training — **Wrong.** Training does not capture value.

✓ **B** Value capture: a defined mechanism and owner to convert saved time into cost reduction or redeployed capacity — **Correct.** Freed time is absorbed unless capture is designed.

✗ **C** A bigger pilot — **Wrong.** Scale without capture repeats the problem.

✗ **D** Model monitoring — **Wrong.** Monitoring does not bank value.

K8 **Answer: C** Review: 3.12

Blocking oversight gaps exist in two of three regions; the third region meets all conditions. Which finding fits?

✗ **A** Proceed everywhere, fixing gaps in flight — **Wrong.** Blocking gaps should not be carried into live operation.

✗ **B** Not yet everywhere — **Wrong.** A ready subset exists, so phasing is more appropriate.

✓ **C** Phase: proceed in the ready region with named remediation and expansion criteria for the others — **Correct.** Phasing starts where ready and sequences the rest.

✗ **D** Reject the use case — **Wrong.** The gaps are remediable.

MODULE 4

Reporting Readiness

Turning a completed assessment into a decision-ready report: graded evidence, verified findings separated from asserted claims, a documented method, every condition tied to a business consequence, sequenced recommendations — and a professional “not yet”.

YOU WILL BE ABLE TO

- Grade evidence and distinguish verified findings from asserted claims
- Document how each finding was learned
- Connect each readiness condition to a business consequence
- Describe maturity gaps against the required state for the use case
- Write evidence-based, sequenced recommendations
- Use a “not yet” recommendation professionally
- Tailor communication to sponsors, governance, compliance and frontline managers without changing conclusions
- Convert raw observations and AI-assisted checklist output into credible findings

IN THIS MODULE

1. From assessment to decision
2. Evidence grading
3. Documenting how findings were learned
4. From condition to consequence
5. Maturity gaps against the required state
6. Evidence-based, sequenced recommendations
7. The professional “not yet”
8. Communicating to stakeholders
9. The readiness report structure
10. Guided activity · templates · cheat sheet · knowledge check

Suggested study time · 4 hours

An assessment that cannot be reported credibly has not been completed. This module turns fieldwork into a **readiness report** that decision-makers can act on before they commit investment: every finding graded for evidence strength, traceable to how it was learned, connected to a business consequence, paired with a sequenced recommendation — and communicated to each audience without changing the conclusion.

KEY TAKEAWAYS

1. **Grade every piece of evidence**; keep **verified findings** separate from **asserted claims**.
2. **Document how each finding was learned** — method, source, date, sample — so others can check it.
3. Connect every condition to a **business consequence**: condition → effect → consequence → decision → action.
4. Describe **maturity gaps against the required state for this use case**, not against a generic scale.
5. Make recommendations **evidence-based and sequenced**: must-fix-before-go, fix-in-flight, improve-later — with owners.
6. A **"not yet"** is a professional recommendation with a route to "yes", not a refusal.
7. **Tailor emphasis by audience; never tailor conclusions.**
8. Treat **AI-assisted checklist output** as assertion until verified.

4.1 From assessment to a decision-ready report

The purpose of the report is a **decision**: proceed, phase or not yet, with conditions. Everything in it either supports that decision or explains how to act on it. Recall the artifact test from Course 1: what decision does it enable, who holds it, what happens if the answer is no?

KEY IDEA

A readiness report is **decision-ready** when the decision-holder can read the first page and know: the recommendation, the evidence strength behind it, the consequences of each major gap, what must happen first, and when the next look will be.

4.2 Evidence grading: verified vs asserted

Assessments mix evidence of very different strength — a profile of 48,000 records next to a manager's comment in an interview. Grading makes the difference visible, so a conclusion resting on assertions is recognised as such.

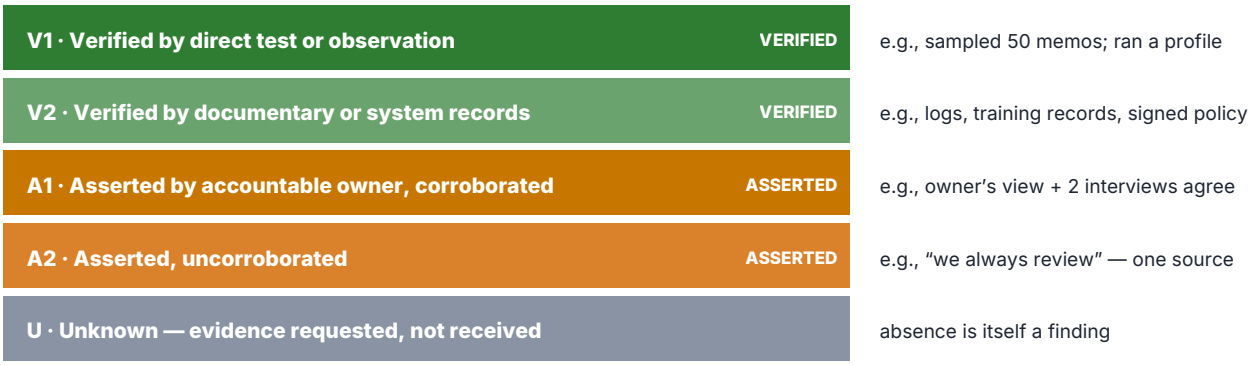


Figure 4.1 — An evidence grading scale

	Verified finding	Asserted claim
Definition	The assessor tested, observed or inspected records that establish the condition	Someone stated it; the assessor has not independently established it
Example	"In a sample of 50 memos, 6 contained unsupported figures."	"The team lead says every memo is checked."
How to report	As a finding, with method and grade	As a claim, attributed and graded; with what would verify it
Can it carry a 'proceed'?	Yes, if material	Not on its own for material conditions

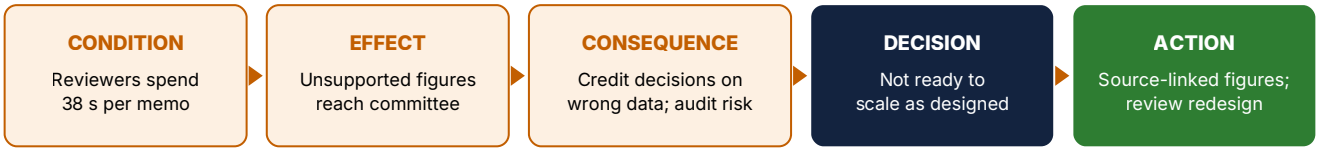
> **EXAM TRAP**
> A frequently tested error: reporting an asserted claim as a finding ("human review is in place") because it came from a senior person. Seniority does not change the evidence grade; only corroboration or verification does.

4.3 Documenting how findings were learned

Every finding needs a traceable **method record**. It protects the credibility of the report, lets reviewers check your work and allows the next assessment to repeat it.

Field	Example
Method	Data profile · document review · system log extract · sample re-performance · observation · interview · scenario test · AI-assisted analysis (verified?)
Source	Work-order table export from ERP; interview with Head of Credit Operations
Date / period	Extract taken 12 March; covers Jan 2024–Feb 2026
Sample and selection	50 memos, random from 2,412 in Q4
Performed by	Assessor name/role
Limitations	Region South logs unavailable; sample excludes restructured loans

4.4 Connecting each condition to a business consequence



If you cannot complete the chain to a business consequence, the finding is not yet worth an executive's attention.

Figure 4.2 — The consequence chain

Consequence type	Example phrasing
Value	"The projected €1.2m downtime saving cannot be realised on current data."
Cost / delay	"Integration rework adds an estimated 10–14 weeks and €180–250k."
Risk — regulatory / legal	"High-risk use without a designed oversight function exposes the bank to non-compliance with deployer obligations."
Risk — operational / customer	"Customers would receive confident but incorrect billing answers."
Risk — people	"Analysts' performance would be judged on volume, encouraging rubber-stamping."

4.5 Maturity gaps against the required state

Report gaps as the distance between the **current state** and the **required state for this use case** — not as a position on a generic five-level maturity scale. "Level 2 of 5 in data governance" tells a sponsor little; "failure codes 63% usable vs 85% required" tells them exactly what must change.

Dimension	Current state (evidence grade)	Required state	Gap	Blocking?
Data	63% usable failure codes (V1)	≥ 85% for 24 months	22 points; ~6 months clean capture	Yes, full scope
Ownership	No system owner (V2: register)	Named business owner	Appointment	Yes
Adoption	Team leads: no usage routine (A1)	Weekly review routine	Routine + KPI change	No — fix in flight

4.6 Evidence-based, sequenced recommendations

Recommendation class	Meaning	Example
Must fix before go	Blocking; the decision depends on it	Appoint system owner; fix repository permissions
Fix in flight	Material but can be closed during early operation with monitoring	Team-lead usage routines
Improve later	Desirable; not a condition	Automate lineage documentation

Each recommendation states: **the finding it answers**, **action**, **owner**, **effort or cost range**, **date**, **dependency** (what must come first) and **how completion will be verified**. Sequence by dependency: ownership and decision rights → blocking data/permission fixes → oversight and policy → training → reinforcement and value capture.

> **PITFALL · The 40-recommendation report**
> Long, unprioritized lists signal that the assessor has not done the judgment work. Group, sequence and mark blocking items; put the rest in an appendix.

4.7 The professional use of "not yet"

"Not yet" is one of the most valuable things an assessment can say — it prevents investment in a capability that would stall. Delivered badly, it sounds like obstruction. Delivered well, it is a plan.

Element of a professional "not yet"	Example
Acknowledge what works	"The model performs well on structured notes; clinicians want it."
Name the blocking conditions, graded	"No oversight design (V2); no system owner (V2); free-text notes untested (U)."
State the consequence of proceeding now	"Unreviewed errors in discharge letters; no one accountable to stop the system."
Give the route to yes	"Three actions, owners, 10 weeks."
Define the criteria for yes	"Owner named; oversight spec approved; error rate on own notes ≤ 2% in a 200-letter test."
Set the re-assessment date	"Re-assess 15 May."
Offer a ready subset if one exists	"Meanwhile, proceed with structured-note wards only." (→ phase)

> **CONSULTANT'S TIP · Holding the line under pressure**
> *"I understand the timeline pressure. If we go now, the most likely outcome is a stalled rollout and a harder conversation in six months. If we fix these three things, I expect to recommend proceeding in May — and I'll tell you sooner if the evidence allows."*

4.8 Communicating to different stakeholders

Audience	Needs	Emphasise	Format
Sponsor / executive	Decision, cost, timing, risk to objectives	Recommendation, consequences, route to yes	1-page summary + briefing
Governance / AI committee	Evidence quality, criteria, residual risk, accountability	Evidence grades, standard and date, oversight design	Full report + evidence register
Compliance / legal / DPO	Regulatory exposure, documentation, data protection	Classification, obligations, gaps, legal questions referred	Findings by obligation; open legal questions
Frontline managers	What changes for their teams, support, timing	Role changes, oversight duties, training, KPIs	Team briefing; FAQs
Technical teams	Precise remediation requirements	Conditions, criteria, verification tests	Remediation backlog

KEY IDEA

Same facts, same conclusion, different emphasis. If the sponsor version says “proceed” and the compliance version implies “not yet”, the report has lost its integrity.

4.9 The readiness report structure

Section	Content
1 · Decision summary	Recommendation (proceed / phase / not yet), top 3–5 findings with consequences, route to yes, decision required and by whom
2 · Scope and baseline	Capability, target state, inventory boundary, standard and date (Module 1)
3 · Readiness by dimension	Seven dimensions: status, weakest link, evidence grade (heat map)
4 · Findings	Per finding: condition, criterion, evidence and grade, how learned, consequence, recommendation
5 · Sequenced remediation	Must-fix / in-flight / later; owners, dates, dependencies, verification
6 · Limitations	What was not assessed, evidence not received, assumptions
Appendices	Evidence register; method log; inventory and scrutiny register; asserted claims awaiting verification

4.10 Guided activity: from raw observations to findings

CASE STUDY **Aldermoor Housing: AI triage of tenant repair requests**

Aldermoor Housing Association (fictional) · 28,000 homes · repair requests by phone, web and app

Target state

An AI system classifies incoming repair requests by urgency (emergency / urgent / routine) and trade, and routes them to contractors. Emergency classifications (gas, flooding, no heat for vulnerable tenants) are always reviewed by a duty officer within 15 minutes.

Raw observations from fieldwork

1. Head of Repairs: “Duty officers check every emergency — it’s in our procedure.”
2. Procedure document v4 (signed, Jan 2026) requires duty-officer review of emergencies within 15 minutes.
3. System log extract (Feb 2026): 312 emergency classifications; 241 had a duty-officer action recorded; median time 47 minutes.
4. Two duty officers interviewed: the night shift has one officer for the whole region; alerts arrive by e-mail.
5. Pilot test on 1,000 historic requests: 94% correct urgency. Of 40 true emergencies, 5 were classified as routine — 3 involved tenants flagged as vulnerable.
6. Vulnerability flags are held in the housing-management system and are not passed to the AI model.
7. IT manager: “The vendor confirmed data stays in the EU.” Contract not yet seen.
8. Contractor managers say routing is “much better”; no measurement.

The AI-assisted checklist

The project team used a general-purpose AI assistant to generate a readiness checklist and pre-fill it from meeting notes. Excerpt:

Checklist item (AI-generated)	AI pre-fill
Human oversight in place	✓ Yes — duty officers review all emergencies
Data residency compliant	✓ Yes — EU hosting confirmed
Model accuracy acceptable	✓ 94% accuracy exceeds industry benchmark of 90%
GDPR Article 22 assessment completed	✓ Not applicable — no automated decisions
Bias testing completed	— (blank)

Worked conversion (three findings)

F1 · Emergency oversight does not operate as designed. *Condition:* 71 of 312 emergency classifications (23%) had no recorded duty-officer action; median response 47 min vs 15 min required. *Criterion:* Procedure v4. *Evidence:* V1 (log extract) contradicts the A2 assertion from the Head of Repairs. *How learned:* log extract Feb 2026, all records; two interviews. *Consequence:* emergencies such as gas leaks may wait unreviewed; tenant-safety and regulatory exposure. *Recommendation (must fix before go):* on-screen alerting and a second night officer, or route all emergencies to a staffed contact centre — Head of Repairs, 4 weeks; verify by 2-week log re-test.

F2 · Missed emergencies concentrate among vulnerable tenants. *Condition:* 5 of 40 true emergencies classified routine; 3 involved vulnerable tenants. Vulnerability flags are not available to the model. *Evidence:* V1 (pilot test), V2 (data-flow documentation). *Consequence:* the people most at risk are most likely to be delayed. *Recommendation (must fix before go):* pass vulnerability flags as an override rule (any vulnerable tenant + heating/water/gas → emergency review) — Data & IT, 6 weeks; re-test on 1,000 cases with ≤ 1 missed emergency.

F3 · EU data residency is asserted, not verified. *Condition:* no contract or data-processing agreement seen. *Evidence:* A2. *Consequence:* unknown transfer risk for tenant personal data. *Recommendation (must fix before go):* obtain and review the DPA with the DPO — Procurement, 2 weeks.

What was wrong with the AI-assisted checklist

- It converted **assertions into ticks** (oversight, residency) — the log extract contradicts the first; the second is unverified.
- It invented a **benchmark** ("industry benchmark of 90%") with no source.
- It used **overall accuracy**, hiding the asymmetric cost of missed emergencies (Course 2).
- It made a **legal determination** ("Art. 22 not applicable") outside its — and the assessor's — authority; this must be referred to the DPO / legal.
- It left **bias testing** blank without flagging it — yet the vulnerable-tenant pattern is exactly such an issue.

Lesson: AI-assisted tools can speed up checklist drafting and note organisation, but their outputs are **asserted claims (at best A2)** until the assessor verifies each one against evidence.

Your turn

1. Convert observation 8 (contractor feedback) into a finding or a claim, with grade.
2. What recommendation would you give overall — proceed, phase or not yet — and why?
3. Write the first two sentences of the sponsor summary.

Model answers (brief)

1. Asserted claim (A2): "Contractor managers report improved routing; not measured." Recommend a measurement: misrouted jobs per 100 before and after.
2. Not yet for emergency triage (F1, F2, F3 block it). A phase is possible: classify routine requests by trade only, with emergencies still handled by the current process, if the DPA is verified.
3. "We recommend not yet for AI triage of emergency repairs: in testing, the system missed 5 of 40 emergencies — 3 for vulnerable tenants — and the human safety net is not operating as designed. Three actions, deliverable within six weeks, would allow us to re-assess in July; meanwhile, AI routing of routine repairs could start once data residency is confirmed."

4.11 Consulting templates

T4.1 Evidence Register & Method Log

Purpose: Grade every item of evidence and record how it was learned.

#	Statement	Grade (V1/V2/ A1/A2/U)	Method	Source	Date / period	Sample	Limitations	Supports finding
E7	*71 of 312 emergencies lacked officer action*	*V1*	*Log extract*	*Repairs system*	*Feb 2026*	*All*	*None*	*F1*

T4.2 Finding Card

Purpose: Write each finding with its full consequence chain.

Finding ID and title: .....

Dimension: .....

Condition (specific, measured): .....

Criterion (standard / target state): .....

Evidence and grade (register refs): .....

How learned: .....

Effect → business consequence: .....

Recommendation (action · owner · effort · date): .....

Class: ☐ must fix before go ☐ fix in flight ☐ improve later .....

Verification of completion: .....

T4.3 Remediation Sequence Plan

Purpose: Sequence recommendations by dependency and class.

Seq.	Action	Finding	Class	Owner	Effort / cost	Depends on	Due	Verified by

Criteria for “yes” / next phase: .....

Re-assessment date or trigger: .....

T4.4 Readiness Report Outline & Stakeholder Message Map

Purpose: Structure the report and keep conclusions identical across audiences.

REPORT OUTLINE

- [ ] 1 Decision summary: recommendation, top findings, consequences, route to yes, decision required
- [ ] 2 Scope and baseline: target state, boundary, standard and date
- [ ] 3 Readiness by dimension with evidence grades
- [ ] 4 Findings (finding cards)
- [ ] 5 Sequenced remediation
- [ ] 6 Limitations
- [ ] Appendices: evidence register, method log, inventory, asserted claims

MESSAGE MAP

Audience	Their key question	Emphasis	Format	Conclusion (must be identical)
Sponsor				
Governance committee				
Compliance / legal / DPO				
Frontline managers				
Technical teams				

T4.5 AI-Assisted Output Verification Sheet

Purpose: Treat AI-generated checklist items as claims until verified.

AI-generated item / pre-fill	Type (fact / assertion / benchmark / legal view)	Source cited?	Verification performed	Result	Grade after verification

- ☐ No legal determinations accepted without legal / DPO review
- ☐ No benchmarks accepted without a traceable source
- ☐ Blank items flagged, not skipped
- ☐ Assertions converted to ticks have been re-opened

4.12 Best practices and pitfalls

✓ DO THIS

- Lead with the decision and the route to yes.
- Grade every evidence item; keep claims and findings apart.
- Record method, source, date and sample for each finding.
- Complete the consequence chain for every material finding.
- Describe gaps against the required state for this use case.
- Sequence and classify recommendations; name owners.
- Tailor emphasis by audience; keep conclusions identical.

× AVOID THIS

- Reporting a senior person's assertion as a verified finding.
- Findings without a business consequence.
- Generic maturity levels as the headline.
- Long, unprioritized recommendation lists.
- "Not yet" without criteria or a re-assessment date.
- Softening the conclusion for the sponsor version.
- Accepting AI-generated ticks, benchmarks or legal views unverified.

4.13 Module 4 cheat sheet

MODULE 4 · ONE-PAGE SUMMARY

Reporting Readiness

EVIDENCE GRADES

V1 tested/observed · **V2** records · **A1** asserted + corroborated · **A2** asserted only · **U** unknown. Seniority does not raise a grade.

METHOD LOG

Method · source · date/period · sample & selection · performed by · limitations. Every finding traceable.

CONSEQUENCE CHAIN

Condition → effect → business consequence (value, cost/delay, regulatory, operational, people) → decision → action.

GAPS

Current vs **required state for this use case**, with grade and "blocking?" — not generic maturity levels.

RECOMMENDATIONS

Must fix before go · fix in flight · improve later. Action, owner, effort, date, dependency, verification. Sequence: ownership → blocking data/permissions → oversight & policy → training → reinforcement & value capture.

PROFESSIONAL "NOT YET"

Acknowledge what works · graded blockers · consequence of going now · route to yes · criteria · re-assessment date · ready subset (→ phase).

AUDIENCES

Sponsor · governance · compliance/legal/DPO · frontline managers · technical teams. **Same facts, same conclusion, different emphasis.**

REPORT STRUCTURE

Decision summary · scope & baseline · readiness by dimension · findings · sequenced remediation · limitations · appendices.

AI-ASSISTED CHECKLISTS

Useful for drafting; outputs are claims (≤ A2) until verified. Watch for: assertions turned into ticks, invented benchmarks, accuracy hiding asymmetric errors, legal determinations, silent blanks.

4.14 Knowledge check

Answer all questions before turning the page.

K1 **The Head of Operations states that "all outputs are reviewed". No logs or samples have been examined. How should this be reported?**

(A) As an asserted claim (A2), attributed, with what would verify it
(B) Omitted, because it is unverified
(C) As a verified finding, because the source is senior
(D) As a V2 finding

K2 **Which evidence is graded V1?**

(A) Two managers agreeing in interviews
(B) A vendor statement
(C) The assessor's own re-performance of a 50-memo sample
(D) A signed procedure requiring review

K3 **Why document how each finding was learned?**

(A) To make the report longer and more impressive
(B) So others can check, challenge and repeat the work, and so limitations are visible
(C) To shift responsibility to interviewees
(D) Because the EU AI Act requires a method log in every report

K4 **Which finding is best connected to a business consequence?**

(A) "Data lineage is immature."
(B) "Lineage documentation is at maturity level 2."
(C) "Lineage tools should be purchased."
(D) "Without lineage for sensor data, the team cannot trace a 14% asset mismatch to its source, so failure predictions for those assets cannot be trusted and €300k of projected savings is at risk."

K5 **How should gaps be expressed in a readiness report?**

(A) As industry benchmark comparisons
(B) As positions on a generic 1–5 maturity scale
(C) As a single overall readiness score
(D) As the distance between current state and the required state for this use case, with evidence grade and whether it blocks

K6 **Which element is essential to a professional "not yet"?**

(A) A list of every minor issue found
(B) Criteria for "yes", named remediation with owners, and a re-assessment date
(C) An apology for delaying the project
(D) A recommendation to cancel the use case

K7 **The sponsor version of a report says “proceed”, while the compliance version implies “not yet”. What is wrong?**

- (A) The sponsor should not see the report.
- (B) The compliance version is too long.
- (C) The report has lost integrity: conclusions must be identical across audiences.
- (D) Nothing — audiences need different messages.

K8 **An AI-assisted checklist ticks “GDPR Article 22 — not applicable”. What should the assessor do?**

- (A) Treat it as an unverified claim and refer the question to the DPO or legal counsel.
- (B) Delete the item.
- (C) Accept it; the AI analysed the notes.
- (D) Change it to “applicable” to be safe.

4.15 Knowledge check: answers and explanations

K1 **Answer: A** Review: 4.2

The Head of Operations states that "all outputs are reviewed". No logs or samples have been examined. How should this be reported?

✓ **A** As an asserted claim (A2), attributed, with what would verify it — **Correct.** Uncorroborated statements are claims until tested.

✗ B Omitted, because it is unverified — **Wrong.** Claims should be recorded and graded, not hidden.

✗ C As a verified finding, because the source is senior — **Wrong.** Seniority does not change the evidence grade.

✗ D As a V2 finding — **Wrong.** V2 requires documentary or system records.

K2 **Answer: C** Review: 4.2

Which evidence is graded V1?

✗ A Two managers agreeing in interviews — **Wrong.** This is corroborated assertion, A1.

✗ B A vendor statement — **Wrong.** This is an assertion.

✓ **C** The assessor's own re-performance of a 50-memo sample — **Correct.** Direct test by the assessor is V1.

✗ D A signed procedure requiring review — **Wrong.** Records showing intent are V2 at most, and do not show operation.

K3 **Answer: B** Review: 4.3

Why document how each finding was learned?

✗ A To make the report longer and more impressive — **Wrong.** Length is not the aim.

✓ **B** So others can check, challenge and repeat the work, and so limitations are visible — **Correct.** Traceability underpins credibility and re-assessment.

✗ C To shift responsibility to interviewees — **Wrong.** Not the purpose.

✗ D Because the EU AI Act requires a method log in every report — **Wrong.** Not the reason.

K4 **Answer: D** Review: 4.4

Which finding is best connected to a business consequence?

✗ A "Data lineage is immature." — **Wrong.** No consequence.

✗ B "Lineage documentation is at maturity level 2." — **Wrong.** A generic level, not a consequence.

✗ C "Lineage tools should be purchased." — **Wrong.** A solution without a consequence.

✓ **D** "Without lineage for sensor data, the team cannot trace a 14% asset mismatch to its source, so failure predictions for those assets cannot be trusted and €300k of projected savings is at risk." — **Correct.** It completes the chain from condition to consequence.

K5 **Answer: D** Review: 4.5

How should gaps be expressed in a readiness report?

✗ A As industry benchmark comparisons — **Wrong.** Benchmarks may not reflect the required state.

✗ B As positions on a generic 1–5 maturity scale — **Wrong.** Generic levels do not show what this use case needs.

✗ C As a single overall readiness score — **Wrong.** Averages hide the weakest link.

✓ **D** As the distance between current state and the required state for this use case, with evidence grade and whether it blocks — **Correct.** This makes gaps actionable.

K6 **Answer: B** Review: 4.7

Which element is essential to a professional "not yet"?

✗ (A) A list of every minor issue found — **Wrong.** Detail belongs in the appendix.

✓ (B) Criteria for "yes", named remediation with owners, and a re-assessment date — **Correct.** These make "not yet" a plan rather than a refusal.

✗ (C) An apology for delaying the project — **Wrong.** Not essential, and it undermines the message.

✗ (D) A recommendation to cancel the use case — **Wrong.** That is a different recommendation.

K7 **Answer: C** Review: 4.8

The sponsor version of a report says "proceed", while the compliance version implies "not yet". What is wrong?

✗ (A) The sponsor should not see the report. — **Wrong.** The sponsor is the decision-holder.

✗ (B) The compliance version is too long. — **Wrong.** Length is not the problem.

✓ (C) The report has lost integrity: conclusions must be identical across audiences. — **Correct.** Same facts, same conclusion, different emphasis.

✗ (D) Nothing — audiences need different messages. — **Wrong.** Emphasis can differ; conclusions must not.

K8 **Answer: A** Review: 4.10

An AI-assisted checklist ticks "GDPR Article 22 — not applicable". What should the assessor do?

✓ (A) Treat it as an unverified claim and refer the question to the DPO or legal counsel. — **Correct.** Legal determinations are outside both the tool's and the assessor's authority.

✗ (B) Delete the item. — **Wrong.** The question is relevant and must be answered properly.

✗ (C) Accept it; the AI analysed the notes. — **Wrong.** AI output is an assertion and this is a legal determination.

✗ (D) Change it to "applicable" to be safe. — **Wrong.** Guessing in either direction is not appropriate.

R1 Course at a glance: the master cheat sheet

ALL FOUR MODULES ON ONE PAGE

Assess AI Readiness

M1 · DIMENSIONS OF READINESS

Readiness = can **this** capability work in **this** context now. Seven dimensions: **data · technology · ownership · adoption · governance · measurement · trust** — a chain, not an average. Model-level risk → testing; organization-level risk → inventory and governance review. Trustworthiness in context. Opening moves: bound inventory (embedded, shadow, adjacent) · classify scrutiny · name evidence · record standard and date. NIST AI RMF and EU AI Act as context.

M2 · DATA AND TECHNOLOGY

Fit for purpose. Quality (profile, don't ask) · governance · provenance vs lineage · access and permission-aware retrieval · single source of truth · labels and GenAI content · **fidelity** · **can / should / permitted** · tiered controls that operate · technology for production · owners. Finding = condition · criterion · evidence · consequence · recommendation.

M3 · ORGANIZATIONAL READINESS

Target state first. Map roles and **control points** (still independent?). **Conditions, not assurances**: proficiency · oversight (competence, authority, time, information, independence, feedback) · accountability · policy clarity · reinforcement · frontline managers · shadow AI · resistance · rework · **value capture**. Finding: **proceed · phase · not yet** with sequenced, owned remediation.

M4 · REPORTING

Grades **V1 · V2 · A1 · A2 · U**; verified vs asserted. Method log. Consequence chain. Gaps vs required state. Recommendations: must fix before go · fix in flight · improve later. Professional "not yet". Same conclusion, tailored emphasis. AI-assisted checklists = claims until verified.

R2 Template index

Code	Template	Use when
T1.1	Readiness Dimension Scan	First week
T1.2	Inventory Bounding & Scrutiny Register	Scoping
T1.3	Governance Evidence Request List	Evidence requests
T1.4	Assessment Baseline Record	Before scoring
T2.1	Data Quality & Fidelity Profile	Data fieldwork
T2.2	Lineage, Provenance & Labels Record	Data fieldwork
T2.3	Access Triad & Tier Controls Check	Access and controls
T2.4	Technology Readiness & Ownership Checklist	Technology
T2.5	Data Readiness Finding	Writing findings
T3.1	Organizational Target State & Control-Point Map	Organizational scope
T3.2	Operating Condition Test Sheet	Testing assurances
T3.3	Human Oversight Design Check	Oversight
T3.4	Adoption & Resistance Scan	Adoption

Code	Template	Use when
T3.5	Organizational Readiness Finding	Closing the module finding
T4.1	Evidence Register & Method Log	Throughout
T4.2	Finding Card	Reporting
T4.3	Remediation Sequence Plan	Reporting
T4.4	Report Outline & Stakeholder Message Map	Reporting
T4.5	AI-Assisted Output Verification Sheet	Using AI tools

R3 Glossary

A

Accountable business owner
Named business role answerable for an AI system's outcomes, with authority to change, pause or retire it.

Adoption (dimension)
Whether people will use an AI capability as intended, supported by process, skills and incentives.

AI literacy
Role-appropriate knowledge and skills to use and oversee AI; an EU AI Act duty (since 2026: take measures to support it) for providers and deployers.

AI-assisted checklist
Checklist drafted or pre-filled by an AI tool; its outputs are claims until verified.

Assertion / asserted claim
A statement someone made that the assessor has not independently established (grades A1/A2).

Assessment baseline
Record of the standards, versions, target state, scope and dates that readiness is measured against.

Assurance
A statement of intention ("we will train everyone"); not evidence of an operating condition.

Automation bias
Tendency to over-accept automated outputs, especially fluent ones under time pressure.

C

Can / should / permitted
Three distinct access questions: technical reach; necessity and proportionality; legal, contractual and policy basis.

Consequence chain
Condition → effect → business consequence → decision → action.

Control point
A step where something is checked, approved or decided (e.g., four-eyes review, sign-off).

D

Data fidelity
How faithfully recorded data represents what actually happened.

Data governance
Roles, policies and processes keeping data fit for use over time.

Data lineage
The path data takes through systems and transformations to reach its use.

Data provenance
The origin of data and the rights and conditions attached to it.

Data quality
Accuracy, completeness, consistency, timeliness, validity and uniqueness of data.

Deployer
Under the EU AI Act, an organization using an AI system under its authority.

DPIA
Data Protection Impact Assessment under GDPR for high-risk processing.

E

Embedded AI
AI features inside existing software, often enabled through updates.

Evidence grade
Rating of evidence strength, e.g., V1, V2, A1, A2, U.

Evidence register
List of every evidence item with grade, method, source, date, sample and the finding it supports.

F

Finding
A reported condition with criterion, graded evidence, method, consequence and recommendation.

Fix in flight
Recommendation class: material but closable during early operation with monitoring.

FRIA
Fundamental-rights impact assessment required of certain deployers of high-risk AI systems under the EU AI Act.

Frontline manager
Team-level manager who sets daily priorities and reinforces (or undermines) adoption.

G

Governance (dimension)
Rules of use, risk classification, oversight, decision rights and documentation existing and operating.

H

Human oversight specification
Document defining who reviews which outputs, with what competence, authority, time and information, and how to override or stop.

Human-in-the-loop
A design in which a human reviews or approves AI outputs; effective only with competence, authority, time, information and independence.

I

Inventory bounding
Defining organizational, use-case, system-type, lifecycle and time scope of the AI inventory for an assessment.

L

Labelled outcome records
Historical examples paired with recorded outcomes from which ML systems learn.

M

Maturity gap
Distance between current state and the required state for a specific use case.

Measurement (dimension)
Baselines, success metrics, monitoring and feedback loops defined and running.

Method log
Record of how each finding was learned: method, source, date, sample, performer, limitations.

Model-level risk
Risk arising from model behaviour — accuracy, robustness, bias, hallucination, drift — established by testing.

Must fix before go
Recommendation class: blocking condition on which the decision depends.

N

NIST AI RMF
Voluntary AI risk framework (Govern, Map, Measure, Manage) with characteristics of trustworthy AI.

Not yet
Readiness finding: blocking gaps across the scope; carries remediation, criteria for yes and a re-assessment date.

O

Operating condition
Something verifiably true now (records, logs, observation, scenario tests).

Organization-level risk
Risk from ownership, oversight, rules, inventory, logging, shadow AI and use in practice — established by structural inventory and governance review.

Oversharing
Over-broad permissions on content that AI assistants make discoverable.

Ownership (dimension)
Named owners for business outcome, system and data, with authority.

P

Permission-aware retrieval
Retrieval that enforces the end user's permissions rather than a broad service account's.

Phase
Readiness finding: proceed in the ready part of scope while remediating the rest.

Pre-deployment assessment
Readiness assessment before a deployment or scaling commitment.

Proceed
Readiness finding: no blocking gaps in the target scope; remaining gaps closed in flight.

R

Readiness assessment
Diagnosis of whether an organization can make a specific AI capability work, in context, now.

Reinforcement
Targets, incentives, routines and leadership signals that sustain new behaviour.

Rework
Effort spent correcting or redoing AI outputs; a hidden cost.

Rubber-stamping
Superficial approval of AI outputs; signs include very short review times and near-zero edits.

S

Scrutiny tier
Depth of review warranted by an inventory item: light, standard, enhanced, escalate.

Shadow AI
AI tools used without organizational sanction.

Single source of truth
Designated system of record (or reconciled master view) for each critical data entity.

System owner
Business role accountable for a deployed AI system's purpose, performance and use.

T

Target state
The defined operating state readiness is measured against: process, autonomy, oversight, scale, roles, value mechanism.

Technology (dimension)
Infrastructure, integration, security, run tooling, logging, scale and fallback for production.

Tiered controls
Controls scaled to data sensitivity tiers (public, internal, confidential, restricted).

Trust (dimension)
Justified, calibrated trust of users and affected people — neither blind reliance nor rejection.

Trustworthiness in context
Trustworthiness as a property of model + data + people + process + oversight + purpose together.

V

Value capture
Designed mechanism and owner converting AI benefits into realised cost, revenue or redeployed capacity.

Verified finding
A condition the assessor established by test, observation or records (grades V1/V2).

W

Weakest link
The dimension whose gap limits overall readiness; readiness is a chain, not an average.

R4 Mock exam

Format: 36 multiple-choice questions · one best answer each · **Time:** 60 minutes · **Suggested pass mark:** 70% (25 correct)

Questions 1–9 cover Module 1, 10–17 Module 2, 18–25 Module 3, 26–33 Module 4 and 34–37 are integrated scenarios. Answer every question before checking the answer key, which explains why each option is right or wrong.

Q1 **What distinguishes an AI readiness assessment from an AI maturity assessment?**

(A) They are the same.
(B) Readiness asks whether a specific capability can work in a specific context now; maturity rates general AI capability.
(C) Maturity assessments are only for regulators.
(D) Readiness uses a five-level scale.

Q2 **Which list gives the seven dimensions of readiness used in this course?**

(A) Govern, map, measure, manage, monitor, report, improve
(B) Value, risk, effort, fit, feasibility, sponsorship, clarity
(C) Strategy, people, process, technology, data, culture, finance
(D) Data, technology, ownership, adoption, governance, measurement, trust

Q3 **A capability scores strongly on six dimensions but has no named owner. What should the assessment conclude about ownership?**

(A) It is likely a blocking weak link that drives the finding.
(B) It is minor because six dimensions are strong.
(C) Ownership is a technology issue.
(D) Ownership can be decided after go-live.

Q4 **Which evidence establishes organization-level risk?**

(A) Red-team results on prompt injection
(B) Sub-group accuracy tests
(C) Inventory records, ownership registers, oversight specifications and observation of review practice
(D) Benchmark leaderboards

Q5 **Why is trustworthiness treated as a property of a system in context?**

(A) Because the same model may be trustworthy in one use and not another, depending on data, people, process, oversight and purpose
(B) Because trust is purely subjective
(C) Because models cannot be tested
(D) Because regulators certify systems, not models

Q6 **What is the first step in bounding a use-case inventory for an assessment?**

A Purchase a discovery tool
B Score the target system
C Interview the vendor
D Define organizational, use-case, system-type, lifecycle and time scope — including embedded, shadow and adjacent AI

Q7 **A proposed system analyses employees' emotions during video calls. Which scrutiny tier fits?**

A Enhanced
B Escalate — refer to legal before anything else
C Light
D Standard

Q8 **Why should the standard and date be recorded before scoring?**

A Because vendors require it
B To make the report look formal
C So the finding can be reproduced, compared at re-assessment and defended as standards change
D Because scores expire after one year

Q9 **Which is an EU AI Act expectation relevant to deployer readiness for high-risk systems?**

A Deployers must train their own foundation models.
B Deployers are exempt if the provider is certified.
C Only public bodies have obligations.
D Deployers must assign human oversight to people with the necessary competence and authority, and keep logs under their control.

Q10 **Which statement about data readiness is correct?**

A The same data can be ready for one use case and not for another.
B Data is either AI-ready or not, regardless of use.
C Passing a governance audit proves readiness.
D Only GenAI needs data readiness.

Q11 **A contract is silent on whether building owners' sensor data may be used for AI. Which access question is open?**

A None
B Can
C Permitted
D Should

Q12 **Before connecting a GenAI assistant to a document platform, which check most directly prevents oversharing?**

A Model accuracy testing
B A permissions review of connected repositories plus permission-aware retrieval
C A cost model
D AI literacy training

Q13 **Two systems hold different versions of the asset register. What should be required before training a model on asset data?**

(A) A designated system of record or reconciled master view
(B) Nothing — models handle inconsistency
(C) Train two models
(D) Use whichever is larger

Q14 **Agents pick "Other" as the reason code to save time. Which readiness issue does this create?**

(A) Uniqueness
(B) Timeliness
(C) Fidelity — the recorded data does not reflect reality
(D) Provenance

Q15 **Health data will be used in an AI pipeline. What shows that tier controls are ready?**

(A) A vendor's security brochure
(B) A policy stating that health data is restricted
(C) Staff awareness of GDPR
(D) Evidence that controls operate: DPIA approved, pseudonymisation applied in the pipeline, isolated environment, access logs

Q16 **Which is a technology readiness gap?**

(A) Policy unclear to users
(B) No manual fallback if the AI is switched off
(C) No named system owner
(D) Labels exist only for audited cases

Q17 **Which elements make a data readiness finding actionable?**

(A) A maturity level
(B) Condition, criterion, graded evidence, business consequence and recommendation with owner, date and go-live impact
(C) Observation and opinion
(D) A vendor recommendation

Q18 **Which is an operating condition?**

(A) "We will train all staff."
(B) "Our managers are committed."
(C) "Humans are always in the loop."
(D) "Training records show 84% completion; 76% passed an error-spotting scenario."

Q19 **Which conditions make human oversight effective?**

(A) A signature field on every output
(B) A long policy document
(C) Competence, authority, time, information, independence and feedback
(D) High model accuracy

Q20 **Median review time dropped to under a minute and the edit rate to 1%, while sampled outputs contain uncorrected errors. What does this show?**

- A Rubber-stamping driven by automation bias
- B Excellent efficiency
- C Policy is clear
- D The model improved

Q21 **Who most directly determines day-to-day adoption of a new AI capability?**

- A The data science team
- B The board
- C The vendor
- D Frontline managers, through priorities, targets and routines

Q22 **Shadow AI use is found in the team that will receive a sanctioned tool. How should it be read?**

- A As irrelevant
- B Only as a policy breach
- C As both an adoption signal (demand, skills) and a governance gap
- D As proof the sanctioned tool will fail

Q23 **A pilot saved time but neither costs nor backlog changed. Which condition was missing?**

- A Value capture by design
- B Data quality
- C Permission-aware retrieval
- D Model monitoring

Q24 **Which remediation should usually come first in the sequence?**

- A Value tracking
- B Naming the accountable owner and decision rights
- C KPI changes
- D Training

Q25 **Blocking gaps exist in every region, including the pilot region. Which finding fits?**

- A Phase
- B Proceed
- C Cancel
- D Not yet, with named remediation, criteria for yes and a re-assessment date

Q26 **A team lead says “every memo is checked”; logs show 23% of memos have no review record. How should this be reported?**

- A As a V1/V2 finding (logs) that contradicts an A2 assertion, reported with both
- B As unknown
- C As verified oversight
- D Omit the assertion

Q27 **Which field belongs in a method log?**

(A) The vendor's price
(B) Sample size and selection
(C) The assessor's opinion of the client
(D) The project budget

Q28 **What makes a finding worth an executive's attention?**

(A) Many sources
(B) Technical detail
(C) A complete chain from condition to business consequence and decision
(D) A high maturity score

Q29 **Which recommendation class fits "appoint a system owner" when no owner exists?**

(A) Optional
(B) Must fix before go
(C) Fix in flight
(D) Improve later

Q30 **Which statement best represents a professional "not yet"?**

(A) "Not yet: three blocking conditions (graded), consequences of going now, actions with owners, criteria for yes, re-assessment on 15 May."
(B) "Proceed with caution."
(C) "The organization is not ready."
(D) "We cannot recommend AI at this time."

Q31 **How should a readiness report differ between the sponsor and the frontline managers?**

(A) Managers should not receive it
(B) The sponsor gets only good news
(C) Different conclusions to suit each audience
(D) Different emphasis and format, same facts and conclusion

Q32 **An AI-assisted checklist states "accuracy exceeds industry benchmark of 90%" with no source. What should the assessor do?**

(A) Accept it
(B) Cite the AI tool as the source
(C) Treat it as an unverified claim; find a traceable source or remove it, and assess error asymmetry rather than overall accuracy
(D) Raise the benchmark to 95%

Q33 **Integrated: A model tests at 96% accuracy, but there is no owner, no oversight design and the policy is unclear. What is the right overall finding?**

(A) Not yet (or phase if a ready subset exists), with ownership, oversight design and policy clarity as sequenced remediation
(B) Retest the model
(C) Proceed with a disclaimer
(D) Proceed — model risk is low

Q34 **Integrated: A GenAI assistant will index a repository that includes confidential client contracts shared company-wide, and the vendor hosts outside the EU by default. Which are blocking findings?**

(A) Neither — the content is internal
(B) Both: oversharing must be fixed before connection, and hosting / transfer terms verified with the DPO
(C) Only hosting
(D) Only oversharing

Q35 **Integrated: The sponsor wants the report to say "ready" because the board meets next week; the evidence shows blocking gaps. What should the assessor do?**

(A) Write two reports
(B) Write "ready with minor issues"
(C) Deliver a professional "not yet" or "phase" with route to yes and dates, and help the sponsor frame it for the board
(D) Delay the report

Q36 **Integrated: Which sequence reflects this course from start to report?**

(A) Score → bound → report
(B) Test the model → proceed
(C) Report → assess → bound
(D) Bound inventory and classify scrutiny → name evidence and record standard and date → assess data, technology and organizational conditions → grade evidence → report findings with sequenced remediation

R5 Mock exam: answer key and explanations

Quick answer grid

1 · B	**2** · D	**3** · A	**4** · C	**5** · A	**6** · D	**7** · B	**8** · C	**9** · D	**10** · A
11 · C	**12** · B	**13** · A	**14** · C	**15** · D	**16** · B	**17** · B	**18** · D	**19** · C	**20** · A
21 · D	**22** · C	**23** · A	**24** · B	**25** · D	**26** · A	**27** · B	**28** · C	**29** · B	**30** · A
31 · D	**32** · C	**33** · A	**34** · B	**35** · C	**36** · D				

Score	Interpretation	Next step
31–36	Exam-ready	Review missed items; run a real assessment with the templates
25–30	Pass level	Re-read cheat sheets for modules where you missed two or more
below 25	Not yet	Re-study weak modules using the “Review” references; retake in one week

Full explanations

Q1 **Answer: B** Review: M1 · 1.1

What distinguishes an AI readiness assessment from an AI maturity assessment?

✗ (A) They are the same. — **Wrong.** They answer different questions.

✓ (B) Readiness asks whether a specific capability can work in a specific context now; maturity rates general AI capability. — **Correct.** Readiness is always readiness for something.

✗ (C) Maturity assessments are only for regulators. — **Wrong.** Incorrect.

✗ (D) Readiness uses a five-level scale. — **Wrong.** Scales are not the distinction.

Q2 **Answer: D** Review: M1 · 1.2

Which list gives the seven dimensions of readiness used in this course?

✗ (A) Govern, map, measure, manage, monitor, report, improve — **Wrong.** Partly NIST functions; not the dimensions.

✗ (B) Value, risk, effort, fit, feasibility, sponsorship, clarity — **Wrong.** These are Course 2 screening dimensions.

✗ (C) Strategy, people, process, technology, data, culture, finance — **Wrong.** A generic transformation list.

✓ (D) Data, technology, ownership, adoption, governance, measurement, trust — **Correct.** These are the seven dimensions.

Q3 **Answer: A** Review: M1 · 1.2–1.3

A capability scores strongly on six dimensions but has no named owner. What should the assessment conclude about ownership?

✓ (A) It is likely a blocking weak link that drives the finding. — **Correct.** The weakest material link decides.

✗ (B) It is minor because six dimensions are strong. — **Wrong.** Readiness is a chain; a missing owner can stall everything.

✗ (C) Ownership is a technology issue. — **Wrong.** It is an organizational dimension.

✗ (D) Ownership can be decided after go-live. — **Wrong.** Without an owner nobody can act on incidents or changes.

Q4 **Answer: C** Review: M1 · 1.4

Which evidence establishes organization-level risk?

- ✗ (A) Red-team results on prompt injection — **Wrong.** Model-/system-level testing.
- ✗ (B) Sub-group accuracy tests — **Wrong.** Model-level.
- ✓ (C) Inventory records, ownership registers, oversight specifications and observation of review practice — **Correct.** Structural inventory and governance review.
- ✗ (D) Benchmark leaderboards — **Wrong.** Model-level, and generic.

Q5 **Answer: A** Review: M1 · 1.5

Why is trustworthiness treated as a property of a system in context?

- ✓ (A) Because the same model may be trustworthy in one use and not another, depending on data, people, process, oversight and purpose — **Correct.** Context determines whether reliance is justified.
- ✗ (B) Because trust is purely subjective — **Wrong.** Trustworthiness has testable characteristics.
- ✗ (C) Because models cannot be tested — **Wrong.** Models can be tested; context still matters.
- ✗ (D) Because regulators certify systems, not models — **Wrong.** Not the reason.

Q6 **Answer: D** Review: M1 · 1.6

What is the first step in bounding a use-case inventory for an assessment?

- ✗ (A) Purchase a discovery tool — **Wrong.** A tool may help but does not define scope.
- ✗ (B) Score the target system — **Wrong.** Scoring comes after bounding and baseline.
- ✗ (C) Interview the vendor — **Wrong.** The vendor cannot bound the client's inventory.
- ✓ (D) Define organizational, use-case, system-type, lifecycle and time scope — including embedded, shadow and adjacent AI — **Correct.** Bounding sets what is in scope.

Q7 **Answer: B** Review: M1 · 1.7

A proposed system analyses employees' emotions during video calls. Which scrutiny tier fits?

- ✗ (A) Enhanced — **Wrong.** Still assumes the use can proceed.
- ✓ (B) Escalate — refer to legal before anything else — **Correct.** Possible prohibited workplace practice.
- ✗ (C) Light — **Wrong.** Far too low.
- ✗ (D) Standard — **Wrong.** Too low for a possible prohibited practice.

Q8 **Answer: C** Review: M1 · 1.9

Why should the standard and date be recorded before scoring?

- ✗ (A) Because vendors require it — **Wrong.** Not relevant.
- ✗ (B) To make the report look formal — **Wrong.** Not the reason.
- ✓ (C) So the finding can be reproduced, compared at re-assessment and defended as standards change — **Correct.** Ready = against standard S, for target T, on date D.
- ✗ (D) Because scores expire after one year — **Wrong.** No such rule.

Q9 **Answer: D** Review: M1 · 1.10

Which is an EU AI Act expectation relevant to deployer readiness for high-risk systems?

✗ (A) Deployers must train their own foundation models. — **Wrong.** No such requirement.

✗ (B) Deployers are exempt if the provider is certified. — **Wrong.** Deployer duties remain.

✗ (C) Only public bodies have obligations. — **Wrong.** Private deployers have obligations too.

✓ (D) Deployers must assign human oversight to people with the necessary competence and authority, and keep logs under their control. — **Correct.** These are deployer expectations (confirm dates).

Q10 **Answer: A** Review: M2 · 2.1

Which statement about data readiness is correct?

✓ (A) The same data can be ready for one use case and not for another. — **Correct.** Purpose, latency, history and labels differ by use case.

✗ (B) Data is either AI-ready or not, regardless of use. — **Wrong.** Readiness is fitness for a specific purpose.

✗ (C) Passing a governance audit proves readiness. — **Wrong.** The audit's criteria may not fit the use case.

✗ (D) Only GenAI needs data readiness. — **Wrong.** All AI depends on fit data.

Q11 **Answer: C** Review: M2 · 2.9

A contract is silent on whether building owners' sensor data may be used for AI. Which access question is open?

✗ (A) None — **Wrong.** The permitted question is unresolved.

✗ (B) Can — **Wrong.** The data is technically reachable.

✓ (C) Permitted — **Correct.** The legal and contractual basis is unclear and must be determined.

✗ (D) Should — **Wrong.** Proportionality is a separate question that may also need answering.

Q12 **Answer: B** Review: M2 · 2.5

Before connecting a GenAI assistant to a document platform, which check most directly prevents oversharing?

✗ (A) Model accuracy testing — **Wrong.** Accuracy is unrelated to permissions.

✓ (B) A permissions review of connected repositories plus permission-aware retrieval — **Correct.** This prevents the assistant surfacing content users shouldn't see.

✗ (C) A cost model — **Wrong.** Unrelated.

✗ (D) AI literacy training — **Wrong.** Training does not fix permissions.

Q13 **Answer: A** Review: M2 · 2.6

Two systems hold different versions of the asset register. What should be required before training a model on asset data?

✓ (A) A designated system of record or reconciled master view — **Correct.** A single source of truth prevents learning from contradictory data.

✗ (B) Nothing — models handle inconsistency — **Wrong.** They learn whatever they are given.

✗ (C) Train two models — **Wrong.** This duplicates the inconsistency.

✗ (D) Use whichever is larger — **Wrong.** Size does not determine correctness.

Q14 **Answer: C** Review: M2 · 2.8

Agents pick "Other" as the reason code to save time. Which readiness issue does this create?

✗ (A) Uniqueness — **Wrong.** Not duplication.

✗ (B) Timeliness — **Wrong.** Not a lag problem.

✓ (C) Fidelity — the recorded data does not reflect reality — **Correct.** Default and catch-all values make labels noise.

✗ (D) Provenance — **Wrong.** The origin is known; faithfulness is the issue.

Q15 **Answer: D** Review: M2 · 2.10

Health data will be used in an AI pipeline. What shows that tier controls are ready?

✗ (A) A vendor's security brochure — **Wrong.** An assertion.

✗ (B) A policy stating that health data is restricted — **Wrong.** Policy shows intent, not operation.

✗ (C) Staff awareness of GDPR — **Wrong.** Necessary but insufficient.

✓ (D) Evidence that controls operate: DPIA approved, pseudonymisation applied in the pipeline, isolated environment, access logs — **Correct.** Operating controls are the readiness test.

Q16 **Answer: B** Review: M2 · 2.11

Which is a technology readiness gap?

✗ (A) Policy unclear to users — **Wrong.** An organizational gap.

✓ (B) No manual fallback if the AI is switched off — **Correct.** Resilience and fallback are part of technology readiness.

✗ (C) No named system owner — **Wrong.** An ownership gap.

✗ (D) Labels exist only for audited cases — **Wrong.** A data/label gap.

Q17 **Answer: B** Review: M2 · 2.13

Which elements make a data readiness finding actionable?

✗ (A) A maturity level — **Wrong.** Generic and not actionable.

✓ (B) Condition, criterion, graded evidence, business consequence and recommendation with owner, date and go-live impact — **Correct.** These make a finding actionable.

✗ (C) Observation and opinion — **Wrong.** Insufficient.

✗ (D) A vendor recommendation — **Wrong.** Not a finding.

Q18 **Answer: D** Review: M3 · 3.4

Which is an operating condition?

✗ (A) "We will train all staff." — **Wrong.** Assurance.

✗ (B) "Our managers are committed." — **Wrong.** Assertion.

✗ (C) "Humans are always in the loop." — **Wrong.** Assertion until tested.

✓ (D) "Training records show 84% completion; 76% passed an error-spotting scenario." — **Correct.** Verifiable condition.

Q19 **Answer: C** Review: M3 · 3.6

Which conditions make human oversight effective?

- ✗ (A) A signature field on every output — **Wrong.** Signatures can be rubber-stamped.
- ✗ (B) A long policy document — **Wrong.** Not sufficient.
- ✓ (C) Competence, authority, time, information, independence and feedback — **Correct.** These conditions let reviewers catch and correct errors.
- ✗ (D) High model accuracy — **Wrong.** Accuracy does not make oversight effective.

Q20 **Answer: A** Review: M3 · 3.6

Median review time dropped to under a minute and the edit rate to 1%, while sampled outputs contain uncorrected errors. What does this show?

- ✓ (A) Rubber-stamping driven by automation bias — **Correct.** Classic indicators.
- ✗ (B) Excellent efficiency — **Wrong.** Errors are passing through.
- ✗ (C) Policy is clear — **Wrong.** Unrelated.
- ✗ (D) The model improved — **Wrong.** The sample contradicts this.

Q21 **Answer: D** Review: M3 · 3.9

Who most directly determines day-to-day adoption of a new AI capability?

- ✗ (A) The data science team — **Wrong.** They build but do not manage operations.
- ✗ (B) The board — **Wrong.** Too distant from daily work.
- ✗ (C) The vendor — **Wrong.** Not inside daily operations.
- ✓ (D) Frontline managers, through priorities, targets and routines — **Correct.** They decide what teams actually do.

Q22 **Answer: C** Review: M3 · 3.10

Shadow AI use is found in the team that will receive a sanctioned tool. How should it be read?

- ✗ (A) As irrelevant — **Wrong.** It is directly relevant.
- ✗ (B) Only as a policy breach — **Wrong.** It also signals demand and skills.
- ✓ (C) As both an adoption signal (demand, skills) and a governance gap — **Correct.** It is a readiness signal in both directions.
- ✗ (D) As proof the sanctioned tool will fail — **Wrong.** Not necessarily.

Q23 **Answer: A** Review: M3 · 3.11

A pilot saved time but neither costs nor backlog changed. Which condition was missing?

- ✓ (A) Value capture by design — **Correct.** Benefit is absorbed unless someone banks it.
- ✗ (B) Data quality — **Wrong.** Unrelated.
- ✗ (C) Permission-aware retrieval — **Wrong.** Unrelated.
- ✗ (D) Model monitoring — **Wrong.** Unrelated to banking value.

Q24 **Answer: B** Review: M3 · 3.12

Which remediation should usually come first in the sequence?

- ✗ (A) Value tracking — **Wrong.** Later.
- ✓ (B) Naming the accountable owner and decision rights — **Correct.** Without an owner, nothing else gets fixed.
- ✗ (C) KPI changes — **Wrong.** Important, but later.
- ✗ (D) Training — **Wrong.** Training without an owner and design is premature.

Q25 **Answer: D** Review: M3 · 3.12

Blocking gaps exist in every region, including the pilot region. Which finding fits?

✗ (A) Phase — **Wrong.** There is no ready subset.

✗ (B) Proceed — **Wrong.** Blocking gaps exist.

✗ (C) Cancel — **Wrong.** The gaps are remediable.

✓ (D) Not yet, with named remediation, criteria for yes and a re-assessment date — **Correct.** Blockers across scope call for not yet, with a route to yes.

Q26 **Answer: A** Review: M4 · 4.2

A team lead says "every memo is checked"; logs show 23% of memos have no review record. How should this be reported?

✓ (A) As a V1/V2 finding (logs) that contradicts an A2 assertion, reported with both — **Correct.** Graded evidence makes the contradiction visible.

✗ (B) As unknown — **Wrong.** Evidence exists.

✗ (C) As verified oversight — **Wrong.** The logs contradict the claim.

✗ (D) Omit the assertion — **Wrong.** Recording it shows why the finding matters.

Q27 **Answer: B** Review: M4 · 4.3

Which field belongs in a method log?

✗ (A) The vendor's price — **Wrong.** Not a method element.

✓ (B) Sample size and selection — **Correct.** Needed so others can check and repeat the work.

✗ (C) The assessor's opinion of the client — **Wrong.** Not a method element.

✗ (D) The project budget — **Wrong.** Not a method element.

Q28 **Answer: C** Review: M4 · 4.4

What makes a finding worth an executive's attention?

✗ (A) Many sources — **Wrong.** Quantity alone is not enough.

✗ (B) Technical detail — **Wrong.** Detail belongs elsewhere.

✓ (C) A complete chain from condition to business consequence and decision — **Correct.** The consequence makes it decision-relevant.

✗ (D) A high maturity score — **Wrong.** Not relevant.

Q29 **Answer: B** Review: M4 · 4.6

Which recommendation class fits "appoint a system owner" when no owner exists?

✗ (A) Optional — **Wrong.** Not optional.

✓ (B) Must fix before go — **Correct.** The decision depends on it.

✗ (C) Fix in flight — **Wrong.** Operating without an owner is not acceptable.

✗ (D) Improve later — **Wrong.** Ownership is blocking.

Q30 **Answer: A** Review: M4 · 4.7

Which statement best represents a professional "not yet"?

✓ **A** "Not yet: three blocking conditions (graded), consequences of going now, actions with owners, criteria for yes, re-assessment on 15 May." — **Correct.** This contains all the elements.

✗ B "Proceed with caution." — **Wrong.** This avoids the finding.

✗ C "The organization is not ready." — **Wrong.** No route to yes.

✗ D "We cannot recommend AI at this time." — **Wrong.** Vague and final-sounding.

Q31 **Answer: D** Review: M4 · 4.8

How should a readiness report differ between the sponsor and the frontline managers?

✗ A Managers should not receive it — **Wrong.** They need what changes for their teams.

✗ B The sponsor gets only good news — **Wrong.** This undermines integrity.

✗ C Different conclusions to suit each audience — **Wrong.** Conclusions must be identical.

✓ **D** Different emphasis and format, same facts and conclusion — **Correct.** Tailor emphasis; never tailor conclusions.

Q32 **Answer: C** Review: M4 · 4.10

An AI-assisted checklist states "accuracy exceeds industry benchmark of 90%" with no source. What should the assessor do?

✗ A Accept it — **Wrong.** Unsourced benchmarks may be invented.

✗ B Cite the AI tool as the source — **Wrong.** An AI tool is not a source for a benchmark.

✓ **C** Treat it as an unverified claim; find a traceable source or remove it, and assess error asymmetry rather than overall accuracy — **Correct.** AI outputs are claims until verified.

✗ D Raise the benchmark to 95% — **Wrong.** No basis.

Q33 **Answer: A** Review: M1 · 1.4, M3 · 3.12

Integrated: A model tests at 96% accuracy, but there is no owner, no oversight design and the policy is unclear. What is the right overall finding?

✓ **A** Not yet (or phase if a ready subset exists), with ownership, oversight design and policy clarity as sequenced remediation — **Correct.** Organization-level gaps block regardless of model quality.

✗ B Retest the model — **Wrong.** The model is not the problem.

✗ C Proceed with a disclaimer — **Wrong.** A disclaimer does not close blocking gaps.

✗ D Proceed — model risk is low — **Wrong.** Model-level evidence does not establish organization-level readiness.

Q34 **Answer: B** Review: M2 · 2.5, 2.10

Integrated: A GenAI assistant will index a repository that includes confidential client contracts shared company-wide, and the vendor hosts outside the EU by default. Which are blocking findings?

✗ A Neither — the content is internal — **Wrong.** Confidential client data is exposed and transferred.

✓ **B** Both: oversharing must be fixed before connection, and hosting / transfer terms verified with the DPO — **Correct.** Both are must-fix-before-go.

✗ C Only hosting — **Wrong.** Oversharing is also blocking.

✗ D Only oversharing — **Wrong.** Hosting is also blocking.

Q35 **Answer: C** Review: M4 · 4.7–4.8

Integrated: The sponsor wants the report to say "ready" because the board meets next week; the evidence shows blocking gaps. What should the assessor do?

✗ (A) Write two reports — **Wrong.** Conclusions must be identical.

✗ (B) Write "ready with minor issues" — **Wrong.** This misrepresents the evidence.

✓ (C) Deliver a professional "not yet" or "phase" with route to yes and dates, and help the sponsor frame it for the board — **Correct.** Integrity with a constructive plan.

✗ (D) Delay the report — **Wrong.** This avoids the decision.

Q36 **Answer: D** Review: M1–M4

Integrated: Which sequence reflects this course from start to report?

✗ (A) Score → bound → report — **Wrong.** Bounding and baseline come before scoring.

✗ (B) Test the model → proceed — **Wrong.** Ignores the organization-level dimensions.

✗ (C) Report → assess → bound — **Wrong.** Reverse order.

✓ (D) Bound inventory and classify scrutiny → name evidence and record standard and date → assess data, technology and organizational conditions → grade evidence → report findings with sequenced remediation — **Correct.** This is the course arc.

R6 Further reading and next steps

Useful public companions to this course. Check for the latest versions.

Resource	Why it is useful
NIST AI Risk Management Framework (AI RMF 1.0), Playbook and Generative AI Profile	Govern–Map–Measure–Manage expectations; trustworthiness characteristics
EU AI Act — Regulation (EU) 2024/1689 (deployer obligations, AI literacy, Annex III)	Readiness expectations for deployers; read with the 2026 AI Omnibus (Regulation (EU) 2026/1744)
ISO/IEC 42001:2023 and ISO/IEC 23894:2023	Management-system and risk-management expectations useful as standards in the baseline
GDPR and national guidance on AI and data protection; DPIA guidance	Permitted-use questions and tiered controls
Literature on automation bias and human oversight	Designing oversight that actually catches errors

Practice assignments

1. Pick a live or planned AI capability; complete T1.1–T1.4 for it, including at least one embedded and one shadow item.
2. Profile one real data extract (T2.1) and write one data readiness finding with T2.5.
3. Convert five assurances you have heard into tested conditions (T3.2) and write a proceed / phase / not yet finding (T3.5).
4. Draft a two-page readiness report with T4.1–T4.4, then ask a colleague to challenge every evidence grade.

ENTERPRISE AI TRANSFORMATION AND CONSULTING PROFESSIONAL

COURSE 4 · STUDY BOOK

Design AI Transformation Roadmaps

How to turn prioritized opportunities and readiness findings into a phased, fundable enterprise roadmap: sequenced by checkable preconditions, funded across horizons, framed for workflows and agents, and pressure-tested before execution.

4 MODULES · CASE STUDIES · TEMPLATES · CHEAT SHEETS · MOCK EXAM

Study edition · 2026

Includes full answer explanations

How to use this book

This study book accompanies **Course 4, Design AI Transformation Roadmaps**, of the Enterprise AI Transformation and Consulting Professional certificate. It is written for consultants, AI strategy and transformation leads, portfolio and investment owners, and enterprise architects who must turn approved AI priorities into a credible, phased plan.

Each module follows the same rhythm so that you always know where you are:

Element	What you get	How to use it
Module opener	Learning objectives, section list, study time	Read first; return to it to self-test at the end
Overview & key takeaways	The module on one screen	Skim before, re-read after
Core sections	Concepts, tables, diagrams, worked examples	Study actively: annotate, restate in your own words
Callout boxes	Key ideas, examples, pitfalls, exam traps, consultant tips	Exam traps are the most frequently tested ideas
Case study	A realistic enterprise scenario with model answers	Answer the questions before reading the model answers
Templates	Fill-in consulting tools (T-codes), one per page or less	Copy into your own engagements
Do / Avoid	Best practices and pitfalls side by side	Use as a pre-meeting checklist
Cheat sheet	One-page module summary	Revise from these in the final week
Knowledge check	Scenario questions with explanations for **every** option	Understanding why the wrong answers are wrong matters as much as getting the right one

Callout key

KEY IDEA
The idea to remember if you remember nothing else.

EXAM TRAP
A frequently tested distinction or a tempting wrong answer.

EXAMPLE
A worked illustration of the concept.

CONSULTANT'S TIP
Practical wording or technique from the field.

PITFALL
A common mistake made in real engagements.

DEFINITION
A precise definition worth learning word for word.

> **NOTE · Builds on Courses 1–3**
> This course turns the outputs of earlier courses into a plan: the prioritized portfolio and its dispositions (Course 2), the readiness findings — proceed, phase, not yet — and their remediation sequences (Course 3), and the deliverable arc, phase gates and decision rights (Course 1). Reused concepts are recalled briefly rather than re-taught.

Suggested study plan (four weeks)

Week	Study	Practise
1	Module 1	Draft a roadmap canvas and precondition cards for three initiatives you know
2	Module 2	Classify ten initiatives by horizon, test portfolio balance, write fund/defer/scale conditions
3	Module 3	Map one workflow, isolate its agentic surface and frame it for the roadmap
4	Module 4, then revise all cheat sheets and the glossary	Run a premortem and control-load test on your roadmap; take the mock exam (60 min) and review every explanation

> **NOTE · About the examples**
> All organizations in case studies and examples are fictional, and their figures are illustrative. Regulatory and framework references are for orientation only; always check current law and guidance, and refer legal interpretation to qualified specialists.

Contents

FRONT **How to use this book** 2

MODULE 1 **Roadmap Fundamentals** 6
What an AI transformation roadmap is — and is not 7
The anatomy of a roadmap 8
Milestones and dependencies 8
Sequencing with checkable preconditions, not calendar dates 9
Sequencing use cases by business value and readiness 10
Horizon planning 10
Phase gates and pilot-to-scale decisions 11
Capacity and downstream work 11
Operating-model choices: centralized, federated, and who owns what 12
Using demonstrated value to support responsible scaling 13
Case study: Tavira Logistics Group 14
Consulting templates 16
Best practices and pitfalls 18
Module 1 cheat sheet 19
Knowledge check 20
Knowledge check: answers and explanations 22

MODULE 2 **Sequencing the Investment Portfolio Across Horizons** 24
Horizons by proof-maturity, not calendar 25
Classifying initiatives 25
Portfolio balance across horizons 26
Phased funding and pilot-to-scale funding gates 27
Capacity constraints 27
Dependency mapping 28
Sequencing investment under uncertainty 28
How early work funds and de-risks later commitments 29
Finance, business-unit, portfolio and AI leadership perspectives 29
Stating observable conditions for defer, fund or scale 29
Case study: Bergmann Industrial 31
Consulting templates 32
Best practices and pitfalls 34
Module 2 cheat sheet 35
Knowledge check 36
Knowledge check: answers and explanations 38

MODULE 3 **AI-Enabled Workflow and Agent Opportunities** 40
Process mapping and workflow redesign — before automation 41
The automation spectrum 42
Identifying AI automation candidates 42
Isolating the agentic surface 43
Deployment layers 43
Citizen-developer governance 44
Permissions, guardrails and system connections 44
Ongoing upkeep, human escalation and verification 45
Reach, reversibility, workflow ownership and residual work 45

Writing a roadmap-ready opportunity frame 46
Case study: Vireo Telecom 47
Consulting templates 49
Best practices and pitfalls 51
Module 3 cheat sheet 52
Knowledge check 53
Knowledge check: answers and explanations 55

MODULE 4 **Pressure-Testing the Roadmap 57**
Planning vs testing 58
Incorrect assumptions vs fragile dependencies 59
Testing assumptions 59
Dependency mapping and sequencing risk 59
Premortem analysis 60
Red-team review 60
Stage-gate review of the roadmap itself 61
Testing controls against operational load 61
Failure modes and concrete consequences 62
Choosing a disposition 62
Evaluating roadmap value, feasibility, resilience and execution risk 63
Case study: Aurelia Pharma 64
Consulting templates 66
Best practices and pitfalls 69
Module 4 cheat sheet 70
Knowledge check 71
Knowledge check: answers and explanations 73

BACK **Course at a glance: the master cheat sheet 75**
Template index 75
Glossary 77
Mock exam 80
Mock exam: answer key and explanations 86
Further reading and next steps 94

MODULE 1

Roadmap Fundamentals

Turning prioritized AI opportunities and organizational readiness into a phased enterprise plan: milestones, dependencies, horizons and gates — sequenced by checkable preconditions rather than calendar dates, within real capacity, under a chosen operating model.

YOU WILL BE ABLE TO

- → Explain what an AI transformation roadmap is and what it is built from
- → Structure phases, workstreams, milestones, dependencies and gates
- → Sequence initiatives with checkable preconditions rather than calendar dates
- → Sequence use cases by business value and readiness, and plan horizons
- → Design phase gates and pilot-to-scale decisions
- → Account for delivery capacity, absorption capacity and downstream work
- → Compare centralized and federated operating models and assign governance, platform and execution ownership
- → Use demonstrated value to support responsible scaling

IN THIS MODULE

1. What a roadmap is
2. Anatomy of a roadmap
3. Milestones and dependencies
4. Preconditions, not calendar dates
5. Sequencing by value and readiness
6. Horizon planning
7. Phase gates and pilot-to-scale decisions
8. Capacity and downstream work
9. Operating-model choices
10. Demonstrated value and responsible scaling
11. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4–5 hours

An AI transformation roadmap turns two earlier outputs — a **prioritized opportunity portfolio** (Course 2) and **readiness findings** (Course 3) — into a phased plan the enterprise can execute, fund and govern. This module teaches the structure of a good roadmap and the discipline that makes it credible: initiatives start when **checkable preconditions** are met, scale when **value is demonstrated**, and stay within the organization's real **capacity** under a clear **operating model**.

KEY TAKEAWAYS

1. A roadmap is a **sequenced plan of use cases and enablers** with milestones, dependencies and gates — not a Gantt chart of projects or a wish list.
2. Sequence with **checkable preconditions** ("starts when the data gate is passed and an owner is named"), not calendar dates; dates are estimates derived from preconditions.
3. Sequence use cases by **value × readiness**: now, prepare, learn, later.
4. Plan **horizons** by proof-maturity: prove, scale, extend.
5. **Pilot-to-scale** is a separate decision with its own criteria; scaling readiness ≠ pilot readiness.
6. Account for **delivery capacity**, **absorption capacity** and the **downstream work** every deployment creates.
7. Choose an **operating model**: who owns governance and standards, the shared platform and local execution.

1.1 What an AI transformation roadmap is — and is not

A roadmap is...	A roadmap is not...
A phased plan connecting use cases to the enablers they depend on	A list of approved AI projects
Sequenced by preconditions, dependencies and evidence	A Gantt chart of fixed dates
Bounded by delivery and absorption capacity	A statement of ambition
Punctuated by gates where decisions are taken on evidence	A commitment to deliver everything listed
Owned, with clear operating-model responsibilities	A technology architecture
Revised as evidence arrives	A document frozen at approval

Input	From	Used for
Prioritized portfolio with dispositions (fund now, prepare, bundle, defer)	Course 2	What goes on the roadmap, and in what broad order
Readiness findings (proceed / phase / not yet) and remediation sequences	Course 3	Preconditions and enabler workstreams
Capability inventory and gap conditions	Course 1	Shared enablers and duplication
Strategy, risk appetite, budget envelope	Leadership	Ambition, constraints, balance
Delivery and operational capacity	Delivery, operations	How much can run in parallel

1.2 The anatomy of a roadmap

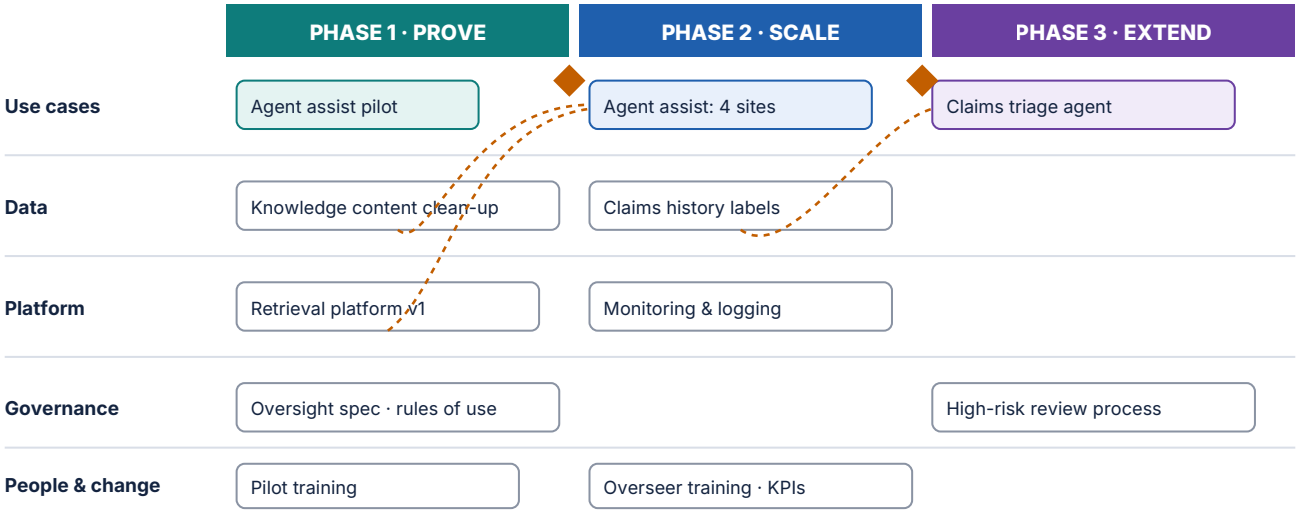


Figure 1.1 — Phases, workstreams, initiatives, dependencies (dashed) and gates (diamonds)

Element	Definition	Good practice
Phase / horizon	A stage of the transformation defined by its purpose (prove, scale, extend)	Defined by proof-maturity, not calendar quarters
Workstream	A lane of related work: use cases, data, platform, governance, people and change	Enablers get their own lanes — never hidden inside use cases
Initiative	A use case or enabler with an owner, scope and outcome	Each has preconditions and exit criteria
Milestone	A verifiable point of progress	Prefer evidence and decision milestones over activity milestones
Dependency	A relationship where one item needs another	Typed, owned, and visible on the map
Gate	A decision point with pre-agreed criteria and one decision-holder	Outcomes: go, go with conditions, pivot, extend, stop

1.3 Milestones and dependencies

Milestone type	Example	Value
Activity	“Training delivered”	Weak: says work happened, not that anything changed
Delivery	“Retrieval platform v1 in production”	Medium: a capability exists
Evidence	“Pilot shows ≥ 25% handling-time reduction on 2,000 cases”	Strong: reduces uncertainty
Decision	“Gate 2: scale decision taken by COO”	Strong: commits or releases resources

Dependency type	Example	Typical fragility
Data	Labelled claims history needed before triage model	Quality and labelling take longer than planned
Platform	Retrieval platform before any knowledge assistant	Shared platform team overloaded
Governance / policy	Oversight spec approved before customer-facing use	Legal review queue
People / skills	Trained overseers before go-live	Training competes with operations
Process	Redesigned approval step before agent acts	Process owner not engaged
Vendor / external	Vendor feature release; contract renewal	Outside the organization's control
Regulatory	Obligations applying from a given date	Dates or interpretations change

CONSULTANT'S TIP

Distinguish **hard** dependencies (cannot start without it) from **soft** ones (could start, but at higher cost or risk). Many roadmaps fail because soft dependencies were treated as hard (needless delay) or hard ones as soft (stalled pilots).

1.4 Sequencing with checkable preconditions, not calendar dates

"Claims triage starts in Q3" is a guess dressed as a plan. It will be true only if a dozen unstated conditions happen to be met by then. A precondition-based roadmap states those conditions explicitly — and they become the real management agenda.

Calendar-based	Precondition-based
"Customer chatbot: Q2."	"Customer assistant starts build when: (1) knowledge content ≥ 90% current and owner-approved; (2) rules of use approved by Legal; (3) contact-centre director accepts system ownership; (4) retrieval platform v1 in production."
"Scale agent assist: Q4."	"Scale to all sites when the pilot shows ≥ 20% handling-time reduction, edit rate 10–40% (oversight working), and site leads have completed overseer training."

DEFINITION · Checkable precondition

A condition that is **observable** (someone can see whether it is true), **verifiable** (evidence exists or can be produced), **owned** (a named person is responsible for making it true) and **binary or thresholded** (no "sufficiently good").

Weak precondition	Checkable version
"Data is ready."	"Data readiness gate (T2.3, Course 2) passed for claims 2022–2025."
"Stakeholders aligned."	"Works council informed and agreement on log use signed."
"Governance in place."	"Oversight specification v1 approved by Head of Risk; overseers named per site."
"Platform available."	"Retrieval platform v1 in production with permission-aware access and logging."

KEY IDEA

Dates do not disappear: they become **forecasts** derived from the expected time to satisfy preconditions — and they move when preconditions slip, visibly and for a reason.

EXAM TRAP

A frequently tested distinction: when asked how to sequence an initiative that depends on data remediation, the best answer ties its start to the **completion of that remediation (a checkable condition)** — not to a quarter that "should" leave enough time.

1.5 Sequencing use cases by business value and readiness

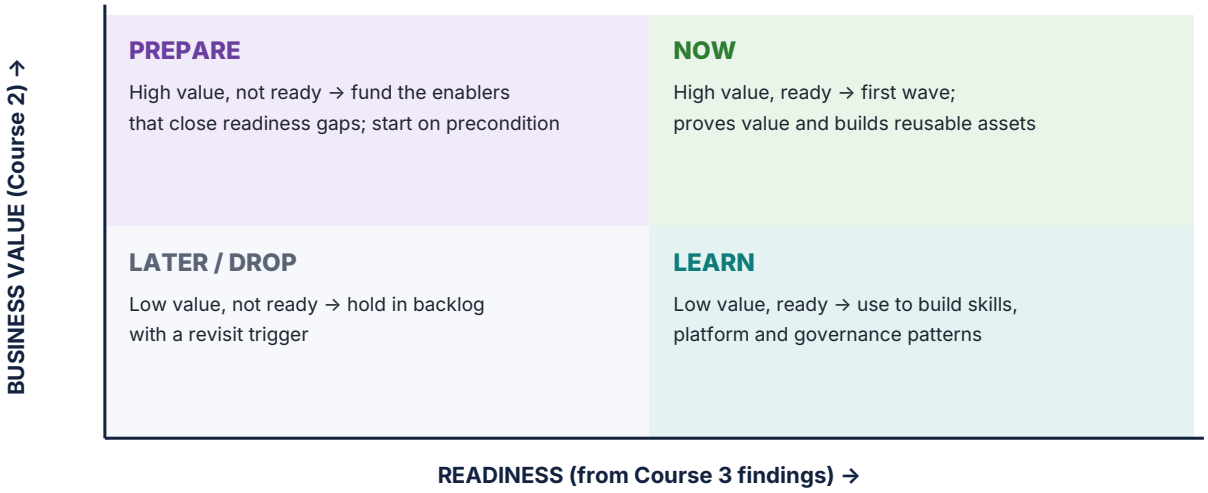


Figure 1.2 — Value × readiness sequencing

Readiness findings from Course 3 translate directly: **proceed** → "now" (if value supports it); **phase** → "now" for the ready part, "prepare" for the rest; **not yet** → "prepare" with the remediation as enabler initiatives and the finding's criteria as preconditions.

BEST PRACTICE · Sequencing heuristics

- Put **enablers** that unblock several use cases early.
- Prefer a first wave that builds **reusable assets** (platform, governance patterns, skills).
- Go **low-risk before high-risk** within the same pattern.
- Include at least one **visible value** item in the first phase.

1.6 Horizon planning

Horizon	Purpose	Typical content	Advances when…
H1 · Prove	Demonstrate value in own context; build first reusable assets	Ready, high-value use cases; core enablers; governance basics	Evidence milestones met
H2 · Scale	Extend proven capabilities across units, volumes, channels	Rollouts; platform hardening; operating-model build-out	Scale gates passed

Horizon	Purpose	Typical content	Advances when…
H3 · Extend / transform	New capabilities, higher autonomy, redesigned processes	Agentic workflows; new products; cross-functional redesign	Foundations and evidence from H1–H2 in place

> **NOTE**
> Horizons are defined by **proof-maturity** and dependency position, not by year. An initiative can move from H3 to H2 early if its preconditions are met — or stay in H3 for years. Module 2 develops horizons for investment.

1.7 Phase gates and pilot-to-scale decisions

A pilot that succeeds proves value at small scale under attentive conditions. Scaling changes almost everything: more users (less enthusiastic), more volume (more edge cases, more review load), more sites (different data and practices), and run costs that now matter. Pilot-to-scale is therefore a **separate decision** with its own criteria.

Scale criterion	Question
Value demonstrated	Did the pilot beat the baseline on the agreed metric, net of the ledger?
Oversight operating	Do review indicators show real checking (not rubber-stamping)? Will it hold at scale volume?
Run cost known	Unit cost at scale volume and support costs estimated from pilot data?
Owner capacity	Can the business owner and overseers absorb the scaled workload?
Risk at scale	Does risk change with reach (more people, more decisions)? Re-classified if needed?
Generalisation	Do other sites' data and practices resemble the pilot's? Tested?
Support model	Help, incident handling and upkeep defined for the scaled population?

> **PITFALL · Scaling the pilot conditions, not the pilot**
> Pilots often succeed because of conditions that will not scale: hand-picked users, a dedicated data scientist on call, clean data from one site. Before scaling, list the pilot's special conditions and check which ones will exist at scale.

1.8 Capacity and downstream work

Two kinds of capacity limit a roadmap. **Delivery capacity**: the people who build, integrate, test and govern (data engineers, platform team, legal and risk reviewers, change managers). **Absorption capacity**: the organization's ability to take on change and the new work AI creates (review queues, exception handling, training time, support).

Downstream work created by each deployment	Who carries it
Monitoring performance and drift	Technical owner / platform team
Reviewing outputs and handling exceptions	Business overseers
Maintaining knowledge content, prompts, connectors	Content owners, platform team
Handling incidents and complaints	Business owner, support, risk

Downstream work created by each deployment	Who carries it
Re-validation after model or vendor changes	Technical owner, risk
Training new staff	Business teams, L&D

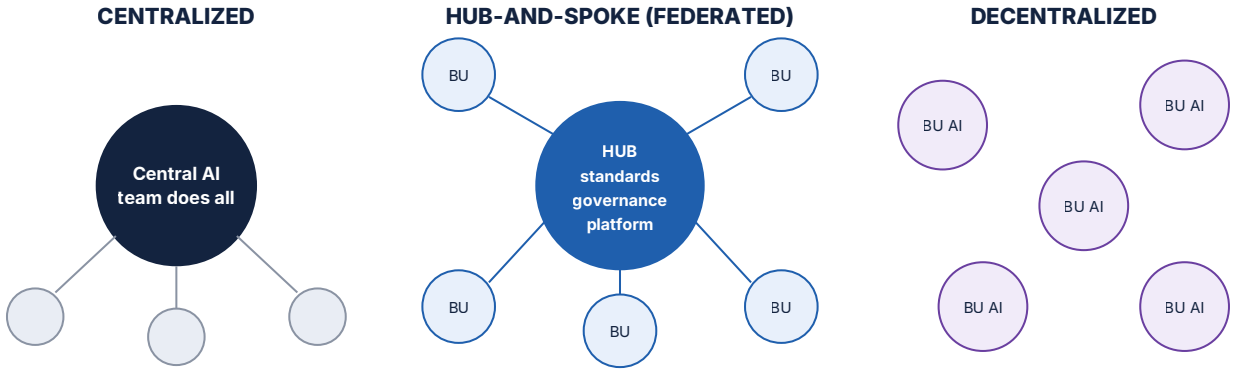
KEY IDEA

Every initiative that goes live **reduces** capacity available for the next one, because it creates permanent run work. A roadmap that plans only build capacity will slow down phase by phase — unless downstream work is planned and staffed.

CONSULTANT'S TIP · Limit work in progress

Set a maximum number of concurrent initiatives per scarce resource (e.g., legal review, data engineering, change management). Starting fewer initiatives and finishing them usually delivers value sooner than starting everything at once.

1.9 Operating-model choices: centralized, federated, and who owns what



Most enterprises converge on a hub-and-spoke: central standards and platform, local execution.

Figure 1.3 — Operating-model archetypes

Model	Strengths	Weaknesses	Fits when
Centralized	Consistent standards; scarce talent pooled; strong governance	Bottleneck; distance from business; slow local adoption	Early stage; high regulation; small organization
Federated (hub-and-spoke)	Central standards and platform with local ownership and speed	Requires clear decision rights; coordination effort	Most large enterprises after initial phase
Decentralized	Speed and business proximity	Duplication, inconsistent controls, tool sprawl	Rarely deliberate; often the unplanned starting state

Responsibility	Typically owned by (federated model)
AI policy, risk taxonomy, standards, approval thresholds	Central governance (hub)
Shared platform: models access, retrieval, monitoring, logging, security patterns	Central platform team

Responsibility	Typically owned by (federated model)
Use-case ownership, process redesign, adoption, value capture	Business units (spokes)
Oversight in operations	Business units, to central standards
Portfolio intake, prioritization, roadmap	Portfolio office with BU representation
Assurance and audit	Second and third lines, independent of delivery

> **EXAM TRAP**
> "Federated" does not mean "every business unit sets its own rules". In a federated model, **standards and governance are central**; **execution is local**. An answer giving business units their own risk thresholds describes decentralization, not federation.

1.10 Using demonstrated value to support responsible scaling

Demonstrated value is what earns the right to scale — and it must be demonstrated in a way that supports the scaling decision: against a baseline, net of moved costs, with oversight shown to work, and with a credible view of what changes at scale.

Responsible scaling practice	Why
Release scaling funds against evidence milestones	Links money to proof
Scale in steps (sites, volumes, autonomy levels)	Each step tests generalisation
Re-check risk classification when reach grows	More people affected may change the tier
Keep a fallback and stop mechanism	Scaling increases the cost of failure
Track value capture, not just usage	Usage without captured value does not fund the next phase

1.11 Case study: Tavira Logistics Group

CASE STUDY **From an approved portfolio to a precondition-based roadmap**

Tavira Logistics Group (fictional) · contract logistics · 11 countries

Inputs

Approved portfolio: customer-service agent assist (fund now), document extraction for customs (fund now), demand forecasting (prepare — master data), autonomous rebooking agent for delayed shipments (prepare — high-risk actions, no process owner). Readiness: agent assist *phase* (ready in two countries); extraction *proceed*; forecasting *not yet* (master data); rebooking *not yet*.

The first draft (rejected)

The PMO's draft placed all four in a 2026 Gantt chart with quarterly start dates, shared one data-engineering team across all of them in Q2, and had no enabler workstreams.

The redesigned roadmap (excerpt)

Initiative	Horizon	Starts when (preconditions)	Scales when
Agent assist — PT & ES	H1	Ready now (proceed for these countries)	≥ 20% handling-time reduction; oversight indicators healthy
Customs extraction	H1	Ready now	Field accuracy ≥ 95% on own documents; exception queue staffed
Master-data harmonisation (enabler)	H1	Data owner named; scope agreed	—
Agent assist — other 9 countries	H2	Knowledge content translated and owner-approved per country	Per-country gate
Demand forecasting	H2	Master-data harmonisation complete for top 3 regions	Forecast error beats baseline by ≥ 10%
Rebooking agent	H3	Rebooking process owner named; action limits and approval thresholds approved; extraction and forecasting live	Autonomy raised level by level on evidence

Operating model: hub-and-spoke — central AI governance and retrieval/extraction platform; country operations own use cases and oversight. Capacity: data-engineering work-in-progress limited to two initiatives; a run team is staffed before extraction goes live.

Discussion questions

1. Why was the quarterly Gantt chart rejected?
2. How did readiness findings become preconditions?
3. Why is the rebooking agent in H3 despite strong value?
4. What downstream work does customs extraction create, and why staff it before go-live?

Model answers (brief)

1. Dates hid unstated conditions, overloaded one team and left out enablers, so the plan would slip without explanation.
2. Each “not yet” remediation became an enabler initiative, and each finding’s “criteria for yes” became a start precondition.
3. It depends on two other capabilities and on governance for autonomous actions; its preconditions cannot be met earlier.
4. An exception queue for low-confidence documents, monitoring and re-validation. Without staffing, the exceptions would pile up and erode the value.

1.12 Consulting templates

T1.1 Roadmap Canvas

Purpose: Lay out phases, workstreams, initiatives, dependencies and gates on one page. **Use when:** first draft of any roadmap

Workstream	H1 · Prove	Gate 1 criteria	H2 · Scale	Gate 2 criteria	H3 · Extend
Use cases					
Data					
Platform					
Governance					
People & change					
Operating model					

Key cross-lane dependencies: .....

WIP limits per scarce resource: .....

T1.2 Initiative Precondition Card

Purpose: Replace a start date with checkable preconditions.

Initiative / owner / horizon: .....

Source disposition (Course 2) and readiness finding (Course 3): .....

#	Precondition (observable, verifiable, thresholded)	Owner	Evidence that it is met	Forecast date	Status

Exit / scale criteria: .....

Downstream work this creates and who staffs it: .....

T1.3 Dependency Register

Purpose: Type, own and track every dependency.

ID	Needs (initiative)	Depends on	Type (data / platform / governance / people / process / vendor / regulatory)	Hard / soft	Owner	Fragility (H/M/L)	Mitigation

T1.4 Operating Model & Ownership Matrix

Purpose: Assign governance, platform and local execution responsibilities.

Chosen model (centralized / hub-and-spoke / decentralized) and why: .....

Responsibility	Central governance	Platform team	Business unit	Portfolio office	Assurance
Policy, standards, thresholds					
Shared platform & patterns					
Use-case ownership & value capture					
Operational oversight					
Intake, prioritization, roadmap					
Monitoring & incidents					
Audit & assurance					

T1.5 Pilot-to-Scale Gate Checklist

Purpose: Decide scaling on its own criteria.

- [ ] Value beat baseline on agreed metric, net of the ledger
- [ ] Oversight indicators healthy and will hold at scale volume
- [ ] Unit run cost and support cost at scale estimated from pilot data
- [ ] Business owner and overseers can absorb scaled workload
- [ ] Risk re-classified for increased reach
- [ ] Pilot's special conditions listed; those absent at scale tested
- [ ] Support, incident handling and upkeep defined
- [ ] Fallback and stop mechanism in place
- [ ] Scaling steps (sites / volume / autonomy) and per-step criteria defined

Decision: go / go with conditions / extend pilot / stop — by whom, date: .....

1.13 Best practices and pitfalls

✓ DO THIS

- Build from the portfolio and readiness findings.
- Give enablers their own workstreams.
- Start initiatives on checkable preconditions.
- Prefer evidence and decision milestones.
- Treat pilot-to-scale as a separate decision.
- Plan and staff downstream run work; limit WIP.
- Make operating-model ownership explicit.

× AVOID THIS

- A Gantt chart of approved projects.
- Hiding data and governance work inside use cases.
- Quarter-based starts with unstated conditions.
- Activity milestones ("training delivered").
- Scaling the pilot's special conditions.
- Planning build capacity only.
- "Federated" meaning every unit sets its own rules.

1.14 Module 1 cheat sheet

MODULE 1 · ONE-PAGE SUMMARY

Roadmap Fundamentals

ROADMAP =

Phased plan of use cases + enablers · milestones · dependencies · gates · capacity-bounded · owned · revised on evidence. Inputs: portfolio (C2), readiness (C3), inventory (C1).

ANATOMY

Phases/horizons · workstreams (use cases, data, platform, governance, people) · initiatives · milestones (activity < delivery < evidence/decision) · dependencies (typed, hard/soft) · gates.

PRECONDITIONS, NOT DATES

Checkable = **observable · verifiable · owned · thresholded**. "Starts when…" Dates become forecasts derived from preconditions and move visibly.

VALUE × READINESS

Now (high, ready) · prepare (high, not ready) · learn (low, ready) · later/drop. Proceed → now; phase → split; not yet → prepare.

HORIZONS

H1 prove · H2 scale · H3 extend/transform — by proof-maturity and dependency position, not by year.

PILOT → SCALE

Value net of ledger · oversight holds at volume · run cost · owner capacity · risk at reach · generalisation · support · fallback. Don't scale pilot conditions.

CAPACITY

Delivery + absorption. Downstream run work (monitoring, review, upkeep, incidents, re-validation, training). Limit WIP.

OPERATING MODEL

Centralized · **hub-and-spoke (federated)** · decentralized. Federated: **central** policy, standards, platform, assurance; **local** use-case ownership, oversight, adoption, value capture. Responsible scaling: funds on evidence, stepwise, re-classify risk, fallback, track value capture.

1.15 Knowledge check

Answer all questions before turning the page. Explanations for every option follow.

K1 **Which best describes an AI transformation roadmap?**

(A) A statement of AI ambition
(B) A target technology architecture
(C) A phased plan connecting use cases to enablers, sequenced by preconditions and dependencies, bounded by capacity and punctuated by gates
(D) A Gantt chart of approved AI projects with start dates

K2 **Which is a checkable precondition?**

(A) “Q3 2026.”
(B) “Data is ready.”
(C) “Stakeholders are aligned.”
(D) “Oversight specification v1 approved by the Head of Risk and overseers named per site.”

K3 **A forecasting use case depends on master-data harmonisation that has no fixed end date. How should its start be set?**

(A) Start when harmonisation is complete for the agreed regions, with a forecast date derived from that work
(B) Start now in parallel to save time
(C) Remove it from the roadmap
(D) Q4 — that should leave enough time

K4 **A use case has high value but a “not yet” readiness finding. Where does it go?**

(A) Learn
(B) Prepare: enabler initiatives for its remediation, with its criteria for “yes” as start preconditions
(C) Now
(D) Drop

K5 **Which statement about horizons is correct?**

(A) Horizons apply only to platforms.
(B) H3 contains only low-value items.
(C) H1 is this year, H2 next year, H3 the year after.
(D) Horizons reflect proof-maturity and dependency position: prove, scale, extend.

K6 **A pilot succeeded with hand-picked users and a data scientist on call. What should happen before scaling?**

(A) List the pilot’s special conditions, test which will exist at scale, and apply separate scale criteria
(B) Scale immediately to all sites
(C) Run the pilot again with the same users
(D) Cancel, because the pilot was unrepresentative

K7 **Why does a roadmap that plans only build capacity tend to slow down phase by phase?**

- (A) Because teams get tired
- (B) Because budgets shrink
- (C) Because every deployment creates permanent downstream run work that consumes capacity
- (D) Because AI models degrade instantly

K8 **In a federated (hub-and-spoke) operating model, who typically owns risk thresholds and standards?**

- (A) The vendor
- (B) Central governance (the hub), with business units executing to those standards
- (C) Each business unit separately
- (D) The data science team

1.16 Knowledge check: answers and explanations

K1 **Answer: C** Review: 1.1

Which best describes an AI transformation roadmap?

- ✗ A — A statement of AI ambition — **Wrong.** Ambition without sequencing is not a roadmap.
- ✗ B — A target technology architecture — **Wrong.** Architecture is one input, not the roadmap.
- ✓ C — A phased plan connecting use cases to enablers, sequenced by preconditions and dependencies, bounded by capacity and punctuated by gates — **Correct.** This captures the roadmap's structure and discipline.
- ✗ D — A Gantt chart of approved AI projects with start dates — **Wrong.** A Gantt chart of projects omits enablers, preconditions, gates and capacity.

K2 **Answer: D** Review: 1.4

Which is a checkable precondition?

- ✗ A — "Q3 2026." — **Wrong.** A date, not a condition.
- ✗ B — "Data is ready." — **Wrong.** Not observable or thresholded.
- ✗ C — "Stakeholders are aligned." — **Wrong.** Vague and unverifiable.
- ✓ D — "Oversight specification v1 approved by the Head of Risk and overseers named per site." — **Correct.** Observable, verifiable, owned and binary.

K3 **Answer: A** Review: 1.4

A forecasting use case depends on master-data harmonisation that has no fixed end date. How should its start be set?

- ✓ A — Start when harmonisation is complete for the agreed regions, with a forecast date derived from that work — **Correct.** Ties the start to a checkable condition; dates become forecasts.
- ✗ B — Start now in parallel to save time — **Wrong.** The hard dependency would make the work stall.
- ✗ C — Remove it from the roadmap — **Wrong.** It has value; it needs sequencing, not removal.
- ✗ D — Q4 — that should leave enough time — **Wrong.** A guessed date hides the dependency.

K4 **Answer: B** Review: 1.5

A use case has high value but a "not yet" readiness finding. Where does it go?

- ✗ A — Learn — **Wrong.** That quadrant is for low-value, ready items.
- ✓ B — Prepare: enabler initiatives for its remediation, with its criteria for "yes" as start preconditions — **Correct.** Readiness remediation becomes enabler work and preconditions.
- ✗ C — Now — **Wrong.** Not ready, so it cannot start.
- ✗ D — Drop — **Wrong.** Its value justifies preparation.

K5 **Answer: D** Review: 1.6

Which statement about horizons is correct?

- ✗ A — Horizons apply only to platforms. — **Wrong.** They apply to all initiatives.
- ✗ B — H3 contains only low-value items. — **Wrong.** H3 often holds high-value, high-dependency items.
- ✗ C — H1 is this year, H2 next year, H3 the year after. — **Wrong.** Horizons are defined by proof-maturity, not calendar.
- ✓ D — Horizons reflect proof-maturity and dependency position: prove, scale, extend. — **Correct.** An initiative moves horizon as evidence and preconditions change.

K6 **Answer: A** Review: 1.7

A pilot succeeded with hand-picked users and a data scientist on call. What should happen before scaling?

✓ **A** List the pilot's special conditions, test which will exist at scale, and apply separate scale criteria — **Correct.** Pilot-to-scale is a separate decision.

✗ B Scale immediately to all sites — **Wrong.** The pilot's special conditions may not exist at scale.

✗ C Run the pilot again with the same users — **Wrong.** That repeats the same bias.

✗ D Cancel, because the pilot was unrepresentative — **Wrong.** The evidence is useful; it needs a scale test.

K7 **Answer: C** Review: 1.8

Why does a roadmap that plans only build capacity tend to slow down phase by phase?

✗ A Because teams get tired — **Wrong.** Not the structural reason.

✗ B Because budgets shrink — **Wrong.** Not necessarily.

✓ **C** Because every deployment creates permanent downstream run work that consumes capacity — **Correct.** Monitoring, review, upkeep and incidents accumulate.

✗ D Because AI models degrade instantly — **Wrong.** Incorrect.

K8 **Answer: B** Review: 1.9

In a federated (hub-and-spoke) operating model, who typically owns risk thresholds and standards?

✗ A The vendor — **Wrong.** Vendors do not own the client's standards.

✓ **B** Central governance (the hub), with business units executing to those standards — **Correct.** Standards are central; execution is local.

✗ C Each business unit separately — **Wrong.** That describes decentralization.

✗ D The data science team — **Wrong.** Standards sit with governance.

MODULE 2

Sequencing the Investment Portfolio Across Horizons

Placing AI initiatives across near-, mid- and longer-term horizons by proof-maturity, dependencies, risk and readiness; funding in phases through pilot-to-scale gates; and stating the observable conditions for defer, fund and scale decisions.

YOU WILL BE ABLE TO

- Place initiatives across horizons using proof-maturity, dependencies, risk and readiness — not calendar dates alone
- Test portfolio balance across horizons
- Design phased funding with pilot-to-scale gates
- Sequence investments under capacity constraints and uncertainty
- Map dependencies, including enabler investments
- Explain how early work can fund and de-risk later commitments
- Reconcile finance, business-unit, portfolio and AI leadership perspectives
- State observable conditions for defer, fund and scale decisions

IN THIS MODULE

1. Horizons by proof-maturity
2. Classifying initiatives
3. Portfolio balance across horizons
4. Phased funding and pilot-to-scale gates
5. Capacity constraints
6. Dependency mapping
7. Sequencing under uncertainty
8. How early work funds and de-risks later work
9. Stakeholder perspectives on funding
10. Defer, fund and scale conditions
11. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4 hours

A roadmap becomes real when money follows it. This module treats the roadmap as an **investment portfolio** spread across horizons: how to place each initiative by how much it has been proven and what it depends on, how to release funding in tranches tied to evidence, how to sequence under capacity limits and uncertainty, and how to state — in observable terms — when to defer, fund or scale.

KEY TAKEAWAYS

1. Place initiatives in horizons by **proof-maturity, dependency position, risk and readiness** — the calendar follows.
2. Test **portfolio balance**: value timing, risk, learning, capacity and enabler share.
3. Fund in **tranches** released at **pilot-to-scale gates**; the amount at risk is the tranche, not the programme.
4. Identify the **binding capacity constraint** and sequence around it.
5. Map **dependencies**; fund enablers explicitly.
6. Under uncertainty, buy **information cheaply** first; keep later commitments **reversible** until evidence arrives.
7. Early work **funds** later work only if value is **captured**, and **de-risks** it through reusable assets and evidence.
8. Write **observable conditions** for defer, fund and scale decisions.

2.1 Horizons by proof-maturity, not calendar

Horizon	Proof-maturity	Typical evidence	Funding posture
Near-term (H1)	Proven in own context, or ready to prove now	Pilot or production evidence on own data; readiness = proceed	Fund delivery and scale steps
Mid-term (H2)	Promising; key conditions still to be met	PoC, comparable references, phase findings	Fund enablers and pilots; scale on evidence
Longer-term (H3)	Option value; depends on earlier assets	Market signals, early experiments	Fund small options and learning; no large commitments

KEY IDEA

Calendar dates alone hide *why* an initiative sits where it does. “H2 because forecasting needs harmonised master data and a pilot result” is defensible; “H2 because it’s next year” is not.

2.2 Classifying initiatives

Classification input	Question	Pushes towards
Proof-maturity	What is the strongest evidence that this works here? (evidence ladder, Course 1)	Strong → H1; weak → H2/H3

Classification input	Question	Pushes towards
Dependency position	Does it depend on unfinished enablers or other initiatives?	Many open dependencies → later
Risk	Use-based risk tier (Course 2) and reach	High-risk → later, after controls are proven on lower-risk uses
Readiness	Proceed / phase / not yet (Course 3)	Not yet → enablers first
Value timing	When does value arrive, and how certain is it?	Early, certain value → H1

EXAMPLE · Classifying four initiatives

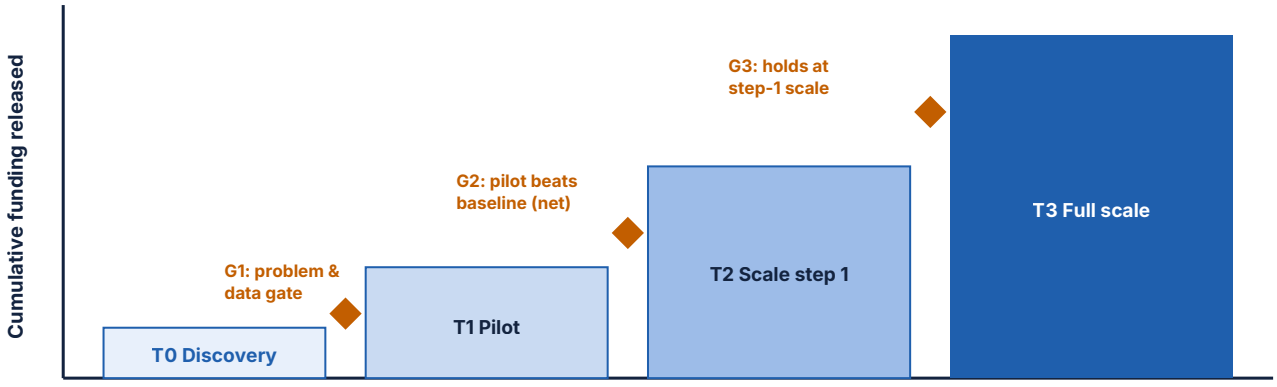
Initiative	Proof	Dependencies	Risk	Readiness	Horizon
Invoice extraction	Production at sister company; PoC on own data	None open	Low	Proceed	H1
Agent assist	Pilot in 2 sites	Content per country	Moderate	Phase	H1 (2 sites) / H2 (rest)
Demand forecasting	References only	Master data	Low	Not yet	H2
Autonomous replenishment agent	Vendor demo	Forecasting + oversight standard	High	Not assessed	H3 (option)

2.3 Portfolio balance across horizons

A portfolio concentrated in H1 delivers early value but builds nothing for later; one concentrated in H3 burns money on options with no early proof. Many organizations use an indicative split (for example, roughly 60–70% near-term, 20–30% mid-term, about 10% longer-term) — treat any such ratio as a starting point to discuss, not a rule.

Balance test	Question	Warning sign
Value timing	Does value arrive in every period?	No measurable value for 18 months
Risk	Is aggregate risk within appetite and governance capacity?	Several high-risk uses launching together
Learning	Does the portfolio buy information about the most uncertain bets?	Large H3 commitments with no experiments
Enabler share	Are enablers funded explicitly?	Enablers hidden in use-case budgets or unfunded
Capacity	Does each scarce resource stay within its limit?	One team named on most initiatives
Concentration	Too much on one vendor, platform or unit?	Single-vendor dependency across horizons

2.4 Phased funding and pilot-to-scale funding gates



Each tranche is released only when the previous gate's observable conditions are met; the maximum loss is the tranche at risk, not the programme.

Figure 2.1 — Tranche-based funding released at gates

Tranche	Funds	Gate to release next tranche
T0 · Discovery	Problem framing, data gate, readiness assessment	Problem and value hypothesis confirmed; data gate passed; owner named
T1 · Pilot	Time-boxed trial on own data with real users	Pilot beats baseline net of ledger; oversight indicators healthy
T2 · Scale step	First expansion (sites, volume)	Results hold at step-1 scale; run cost confirmed
T3 · Full scale	Enterprise rollout and run	Value capture tracked; support model operating

CONSULTANT'S TIP

Ask finance to approve the **envelope** for the whole initiative but **release** by tranche. Sponsors get confidence that money exists; finance gets control over when it is spent.

2.5 Capacity constraints

Money is rarely the binding constraint in AI portfolios. More often it is a scarce **resource**: data engineers, the platform team, legal and risk reviewers, subject-matter experts, change managers — or the business's capacity to absorb change.

Step	Action
1 · Identify	List each scarce resource and its available capacity per period
2 · Load	Map each initiative's demand on each resource by tranche
3 · Find the binding constraint	The resource most over-committed
4 · Sequence around it	Prioritize initiatives by value per unit of the binding resource; set WIP limits
5 · Relieve it	Hire, contract, automate or simplify — often the highest-return investment in the portfolio

EXAMPLE

Twelve initiatives each need four weeks of legal review in the same half-year; legal has capacity for five. Without sequencing, all twelve stall in the legal queue. With sequencing, five proceed, seven start on a precondition — and a reusable review template cuts later reviews to two weeks.

2.6 Dependency mapping

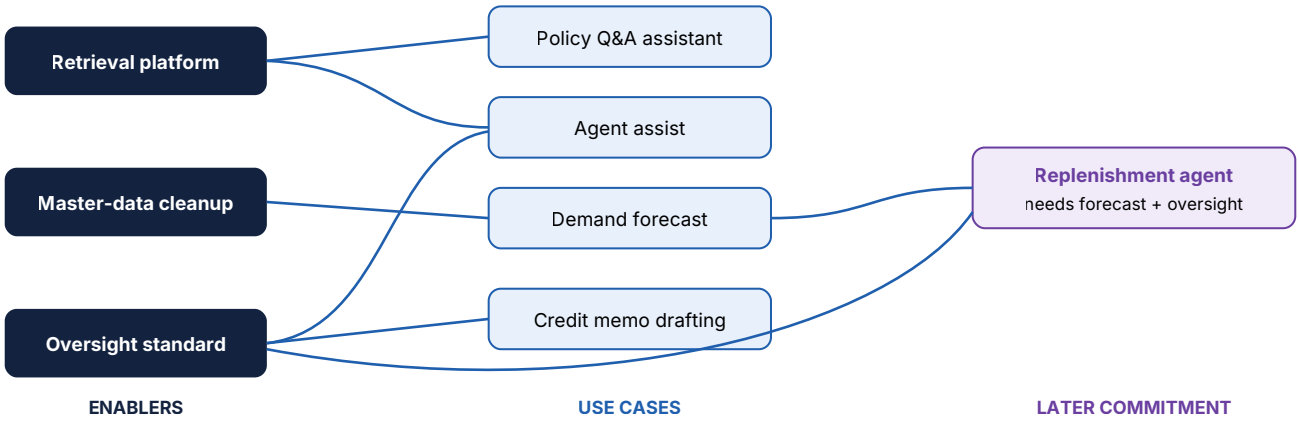


Figure 2.2 — Enablers, use cases and a later commitment

A dependency map shows which **enabler investments** unlock which use cases. Enablers that unlock several use cases deserve early funding even though they create no value directly; the value they unlock belongs in their justification.

PITFALL · Orphaned enablers

Enablers are often unfunded because no single business case "owns" them. Fund them as portfolio-level investments, justified by the use cases they unlock, with their own gates.

2.7 Sequencing investment under uncertainty

Principle	Application
Buy information cheaply	Small experiments and pilots on the biggest uncertainties before large commitments
Keep options open	Prefer short contract terms, modular architecture and reversible choices until evidence arrives
Stage commitments	Commit in tranches; each gate is a real option to continue, change or stop
Know the stop rule	Define in advance what result ends funding
Separate reversible from irreversible	Move fast on reversible decisions; slow down on irreversible ones (long contracts, decommissioning legacy, workforce changes)

KEY IDEA

Under uncertainty the value of early work is partly **information**: a cheap pilot that shows a big bet will not work is a good investment.

2.8 How early work funds and de-risks later commitments

Mechanism	How it works	Condition
Self-funding	Captured benefits from H1 initiatives are allocated to H2 budgets	Benefits are **captured** (cost actually removed or redeployed), not just measured
Reusable assets	Platform, governance patterns, data products and skills built in H1 reduce later cost and time	Assets designed for reuse, not one-off
Evidence	H1 results reduce uncertainty for similar later initiatives	Comparable task, data and context
Credibility	Delivered results increase sponsor and finance confidence	Results reported honestly, including failures

EXAM TRAP

"H1 saves €2m, which will fund H2" is valid only if the €2m is **captured** — removed from budgets or redeployed. Time saved but not banked funds nothing.

2.9 Finance, business-unit, portfolio and AI leadership perspectives

Perspective	Cares most about	Typical concern	What helps
Finance	Cash, returns, accountability, cost control	Open-ended spend; benefits that never appear in budgets	Tranches, captured-value tracking, stop rules
Business unit	Local value, control, speed	Central priorities delaying local needs	Clear criteria, local ownership, visible slots
Portfolio office	Balance, capacity, dependencies	Over-commitment; hidden dependencies	Capacity loading, dependency map, WIP limits
AI leadership	Platform, standards, capability building	Enablers unfunded; duplication	Portfolio-level enabler funding, reuse

CONSULTANT'S TIP · Making the tension productive

Put all four perspectives in the same funding session with the same facts: horizon classification, balance tests, capacity load and gate conditions. Disagreements then become about specific criteria rather than about whose priorities matter more.

2.10 Stating observable conditions for defer, fund or scale

Decision	Observable condition (examples)
Defer	"Defer until master-data match rate ≥ 95% for top-3 regions (currently 81%), reviewed monthly."
Fund (next tranche)	"Fund pilot when: data gate passed, owner named, oversight spec approved, baseline measured."
Scale	"Scale to 4 sites when: handling time −20% vs baseline over 8 weeks; edit rate 10–40%; unit cost ≤ €0.12; site leads trained."

Decision	Observable condition (examples)
Stop	“Stop if pilot shows < 10% improvement after 10 weeks or any critical safety incident.”

BEST PRACTICE

Write conditions so that a person who was not in the meeting could check them: metric, threshold, data source, period, who confirms.

2.11 Case study: Bergmann Industrial

CASE STUDY **Twelve initiatives, one funding committee**

Bergmann Industrial (fictional) · machinery manufacturer · 7 plants

Situation

The roadmap lists twelve AI initiatives. The CFO wants an annual budget decision; plant directors want their projects started; the AI lead warns that the retrieval platform and data foundations are unfunded.

What the advisor did

- **Classified** all twelve by proof, dependencies, risk and readiness: 4 near-term, 5 mid-term, 3 longer-term options.
- **Tested balance:** no H3 experiments for the two biggest bets (autonomous scheduling; AI-designed spare parts); enablers 0% of budget; legal review loaded at 240%.
- **Restructured funding:** a three-year envelope approved; tranches released at gates. Enablers (platform, master data, oversight standard) funded at portfolio level as 22% of year-one spend.
- **Relieved the constraint:** a reusable AI legal-review template and an external counsel framework, reducing review time from 4 to 2 weeks.
- **Self-funding rule:** 50% of captured savings from the two H1 initiatives (verified by finance) released to the H2 pool.
- **Conditions:** each initiative given observable defer / fund / scale / stop conditions in a single register reviewed quarterly.

Discussion questions

1. Why were the two big bets given small experiments rather than full funding or no funding?
2. Why fund enablers at portfolio level?
3. Why did the self-funding rule depend on finance verification?
4. Which perspective is most served by tranche release, and how do the others benefit?

Model answers (brief)

1. They are the most uncertain; cheap experiments buy information before irreversible commitments, keeping the option open.
2. They unlock several use cases but belong to none; portfolio funding prevents them from being orphaned.
3. Only captured savings can fund later work; verification prevents counting unbanked time as money.
4. Finance (control and stop rules); BUs gain a transparent path to funding; the portfolio office gains capacity control; AI leadership gains funded enablers.

2.12 Consulting templates

T2.1 Horizon Classification Sheet

Purpose: Place each initiative by proof, dependencies, risk and readiness.

Initiative	Strongest evidence (ladder level)	Open dependencies	Risk tier	Readiness finding	Value timing	Horizon	Rationale (one line)

T2.2 Portfolio Balance Test

Purpose: Check balance across horizons before funding.

Test	Current portfolio	Target / guideline	Issue	Adjustment
Spend by horizon (H1/H2/H3)				
Value timing by period				
High-risk items concurrent				
Experiments on biggest uncertainties				
Enabler share of spend				
Load on binding resource				
Vendor / platform concentration				

T2.3 Funding Tranche & Gate Plan

Purpose: Release funding in tranches tied to evidence.

Initiative / total envelope / owner: .....

Tranche	Amount	Funds what	Release gate (observable conditions)	Stop rule	Decision-holder
T0 Discovery					
T1 Pilot					
T2 Scale step					
T3 Full scale					

T2.4 Capacity Load & Dependency Map

Purpose: Find the binding constraint and the enablers that unlock most value.

CAPACITY LOAD

Scarce resource	Capacity / period	Demand (sum of initiatives)	Load %	Binding?	Relief option

ENABLER → USE CASES

Enabler	Unlocks	Value unlocked (range)	Funded at portfolio level?	Gate

T2.5 Stakeholder Perspective Grid

Purpose: Surface and reconcile funding perspectives.

Perspective	Top concern for this portfolio	Evidence that would satisfy it	Proposed mechanism	Agreed?
Finance				
Business units				
Portfolio office				
AI leadership				

T2.6 Defer / Fund / Scale / Stop Condition Register

Purpose: Make every funding decision checkable.

Initiative	Current status	Next decision	Observable condition (metric · threshold · source · period)	Confirmed by	Review date

2.13 Best practices and pitfalls

✓ DO THIS

- Classify horizons by proof, dependencies, risk and readiness.
- Test balance, including enabler share and experiments.
- Approve envelopes, release tranches at gates.
- Find and relieve the binding capacity constraint.
- Fund enablers at portfolio level.
- Buy information cheaply before irreversible commitments.
- Write observable defer / fund / scale / stop conditions.

× AVOID THIS

- Horizons as calendar years.
- All spend in H1, or big unproven H3 bets.
- Full funding up front for uncertain initiatives.
- Assuming money is the only constraint.
- Leaving enablers to individual business cases.
- Counting uncaptured savings as self-funding.
- Vague conditions (“when ready”).

2.14 Module 2 cheat sheet

MODULE 2 · ONE-PAGE SUMMARY

Sequencing the Investment Portfolio Across Horizons

HORIZONS

H1 proven/ready · H2 promising, conditions open · H3 options. Placed by **proof-maturity, dependencies, risk, readiness**, value timing.

BALANCE TESTS

Spend by horizon · value timing · concurrent high-risk · experiments on big bets · enabler share · binding resource load · concentration.

TRANCHE FUNDING

T0 discovery → G1 (problem, data gate, owner) → T1 pilot → G2 (beats baseline net; oversight healthy) → T2 scale step → G3 (holds at scale; run cost) → T3 full scale. Approve envelope; release by tranche; stop rules.

CAPACITY

Identify → load → binding constraint → sequence by value per unit of it → relieve it.

DEPENDENCIES

Map enablers → use cases. Fund enablers at portfolio level, justified by value unlocked.

UNDER UNCERTAINTY

Buy information cheaply · keep options open · stage commitments · stop rules · fast on reversible, slow on irreversible.

EARLY FUNDS LATER

Self-funding only from **captured** value · reusable assets · comparable evidence · credibility.

PERSPECTIVES & CONDITIONS

Finance (control) · BU (local value) · portfolio (balance, capacity) · AI leadership (platform, standards). Conditions: metric · threshold · source · period · who confirms — for **defer, fund, scale, stop**.

2.15 Knowledge check

Answer all questions before turning the page.

K1 **Which rationale best justifies placing an initiative in the mid-term horizon?**

(A) "The sponsor prefers to wait."
(B) "It's planned for next year."
(C) "It has PoC evidence but depends on master-data harmonisation and a pilot result."
(D) "It is expensive."

K2 **A portfolio has 95% of spend on near-term initiatives and no funded enablers. What is the main risk?**

(A) Too much risk
(B) Nothing is built for later horizons; mid-term initiatives will be blocked by missing enablers
(C) Too little early value
(D) Finance will object to early value

K3 **Why release funding in tranches tied to gates?**

(A) Because budgets are annual
(B) To slow projects down
(C) To avoid stop decisions
(D) So the amount at risk at any time is limited to the current tranche, and money follows evidence

K4 **Twelve initiatives each need legal review, and legal can handle five per half-year. What is the best portfolio response?**

(A) Treat legal review as the binding constraint: sequence by value per review slot, set a WIP limit, and invest in relieving the constraint
(B) Start all twelve anyway
(C) Skip legal review for low-value items
(D) Cancel seven initiatives

K5 **A retrieval platform unlocks four use cases but has no business case of its own. How should it be funded?**

(A) Split across the four use cases' budgets
(B) As a portfolio-level enabler investment justified by the value it unlocks, with its own gates
(C) Not at all until a use case pays for it
(D) By the vendor

K6 **Which approach best sequences investment under high uncertainty about a large autonomous-agent bet?**

(A) Fund fully now to move fast
(B) Sign a five-year platform contract to secure a discount
(C) Run a small, time-boxed experiment on the main uncertainty with a defined stop rule, keeping later commitments reversible
(D) Wait until competitors prove it

K7 **H1 initiatives "save 30,000 hours". Can this fund H2?**

(A) Yes, at the average salary cost
(B) Yes, if the sponsor agrees
(C) No, savings can never fund other work
(D) Only to the extent the value is captured — cost removed or capacity redeployed — and verified

K8 **Which is an observable scale condition?**

(A) "Scale to 4 sites when handling time is −20% vs baseline over 8 weeks, edit rate is 10–40% and unit cost is ≤ €0.12, confirmed by the ops controller."
(B) "Scale in Q3."
(C) "Scale once the model is good enough."
(D) "Scale when the business is comfortable."

2.16 Knowledge check: answers and explanations

K1 **Answer: C** Review: 2.1–2.2

Which rationale best justifies placing an initiative in the mid-term horizon?

- ✗ (A) "The sponsor prefers to wait." — **Wrong.** Preference is not a placement criterion.
- ✗ (B) "It's planned for next year." — **Wrong.** A calendar label, not a rationale.
- ✓ (C) "It has PoC evidence but depends on master-data harmonisation and a pilot result." — **Correct.** Proof-maturity and dependency position justify the placement.
- ✗ (D) "It is expensive." — **Wrong.** Cost alone does not determine the horizon.

K2 **Answer: B** Review: 2.3

A portfolio has 95% of spend on near-term initiatives and no funded enablers. What is the main risk?

- ✗ (A) Too much risk — **Wrong.** Near-term items are usually lower risk.
- ✓ (B) Nothing is built for later horizons; mid-term initiatives will be blocked by missing enablers — **Correct.** Balance requires enabler funding and learning for later horizons.
- ✗ (C) Too little early value — **Wrong.** Early value is plentiful.
- ✗ (D) Finance will object to early value — **Wrong.** Unlikely.

K3 **Answer: D** Review: 2.4

Why release funding in tranches tied to gates?

- ✗ (A) Because budgets are annual — **Wrong.** Tranches are evidence-based, not calendar-based.
- ✗ (B) To slow projects down — **Wrong.** Not the purpose.
- ✗ (C) To avoid stop decisions — **Wrong.** Gates make stop decisions possible.
- ✓ (D) So the amount at risk at any time is limited to the current tranche, and money follows evidence — **Correct.** Tranches convert large commitments into evidence-based steps.

K4 **Answer: A** Review: 2.5

Twelve initiatives each need legal review, and legal can handle five per half-year. What is the best portfolio response?

- ✓ (A) Treat legal review as the binding constraint: sequence by value per review slot, set a WIP limit, and invest in relieving the constraint — **Correct.** This is the capacity-constraint method.
- ✗ (B) Start all twelve anyway — **Wrong.** They will queue and stall.
- ✗ (C) Skip legal review for low-value items — **Wrong.** Bypassing governance is not acceptable.
- ✗ (D) Cancel seven initiatives — **Wrong.** Sequencing is better than cancelling.

K5 **Answer: B** Review: 2.6

A retrieval platform unlocks four use cases but has no business case of its own. How should it be funded?

- ✗ (A) Split across the four use cases' budgets — **Wrong.** This often leaves it unfunded or fragmented.
- ✓ (B) As a portfolio-level enabler investment justified by the value it unlocks, with its own gates — **Correct.** Prevents orphaned enablers.
- ✗ (C) Not at all until a use case pays for it — **Wrong.** This blocks all four use cases.
- ✗ (D) By the vendor — **Wrong.** Not a sound funding mechanism.

K6 **Answer: C** Review: 2.7

Which approach best sequences investment under high uncertainty about a large autonomous-agent bet?

✗ (A) Fund fully now to move fast — **Wrong.** This maximises exposure.

✗ (B) Sign a five-year platform contract to secure a discount — **Wrong.** An irreversible commitment made before evidence exists.

✓ (C) Run a small, time-boxed experiment on the main uncertainty with a defined stop rule, keeping later commitments reversible — **Correct.** Buys information cheaply and keeps the option open.

✗ (D) Wait until competitors prove it — **Wrong.** Waiting forfeits learning.

K7 **Answer: D** Review: 2.8

H1 initiatives "save 30,000 hours". Can this fund H2?

✗ (A) Yes, at the average salary cost — **Wrong.** Unbanked time is not cash.

✗ (B) Yes, if the sponsor agrees — **Wrong.** Agreement does not create cash.

✗ (C) No, savings can never fund other work — **Wrong.** Captured savings can.

✓ (D) Only to the extent the value is captured — cost removed or capacity redeployed — and verified — **Correct.** Self-funding requires captured value.

K8 **Answer: A** Review: 2.10

Which is an observable scale condition?

✓ (A) "Scale to 4 sites when handling time is −20% vs baseline over 8 weeks, edit rate is 10–40% and unit cost is ≤ €0.12, confirmed by the ops controller." — **Correct.** Metric, threshold, period and confirmer are all stated.

✗ (B) "Scale in Q3." — **Wrong.** A date, not a condition.

✗ (C) "Scale once the model is good enough." — **Wrong.** Not thresholded.

✗ (D) "Scale when the business is comfortable." — **Wrong.** Not observable.

MODULE 3

AI-Enabled Workflow and Agent Opportunities

Identifying AI-enabled workflow opportunities and framing agent use cases for the roadmap: choosing the right kind of automation, isolating the agentic surface, and accounting for permissions, guardrails, upkeep, escalation and residual work — framing, not build.

YOU WILL BE ABLE TO

- Use process mapping and workflow redesign to find AI-enabled workflow opportunities
- Distinguish ordinary automation, predictive AI, generative AI, assistants and multi-step agentic workflows
- Isolate the agentic surface where judgment justifies autonomy
- Account for deployment layers and citizen-developer governance
- Account for permissions, guardrails, system connections, upkeep, escalation and verification
- Use reach, reversibility, workflow ownership and residual work to shape cost, risk and feasibility
- Write a roadmap-ready agent opportunity frame

IN THIS MODULE

1. Redesign before automating
2. The automation spectrum
3. Identifying automation candidates
4. Isolating the agentic surface
5. Deployment layers
6. Citizen-developer governance
7. Permissions, guardrails and connections
8. Upkeep, escalation and verification
9. Reach, reversibility, ownership, residual work
10. The roadmap-ready frame
11. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4–5 hours

Agents and AI-enabled workflows are the fastest-growing category on enterprise AI roadmaps — and the easiest to frame badly. This module teaches you to find workflow opportunities through process mapping and redesign, choose the **least autonomous technique that does the job**, isolate the narrow **agentic surface** where judgment justifies autonomy, and frame each opportunity with its real cost and risk: permissions, guardrails, upkeep, escalation and residual work. The focus is **opportunity identification and roadmap-ready framing**, not technical build.

KEY TAKEAWAYS

1. **Redesign** the workflow before automating it; don't automate waste.
2. Use the **automation spectrum**: ordinary automation → predictive AI → generative AI → assistant → agentic workflow. Pick the least autonomous option that works.
3. Isolate the **agentic surface** — the steps where variable inputs and judgment justify autonomy; keep deterministic steps deterministic.
4. Account for **deployment layer** and **citizen-developer governance**.
5. Every agent needs **permissions, guardrails and system connections** designed — and generates **upkeep**, **escalation** and **verification** work.
6. **Reach, reversibility, workflow ownership and residual work** shape cost, risk and feasibility.

3.1 Process mapping and workflow redesign — before automation

Mapping a workflow (Course 2, Module 1) often shows that steps exist only to compensate for other problems: re-checking data entered badly upstream, chasing approvals that add no control, reformatting between systems. Automating these steps with AI makes waste faster. **Redesign first**: remove, simplify or move steps — then ask what is left for automation and AI.

Redesign question	Possible outcome
Does this step need to exist?	Remove it (e.g., redundant approval)
Can the input be captured right at source?	Fix upstream (e.g., structured form instead of e-mail)
Can steps run in parallel?	Shorter cycle time without AI
Can the decision be made by a rule?	Ordinary automation, not AI
Does a step need judgment on variable inputs?	Candidate for AI — and possibly the agentic surface

3.2 The automation spectrum

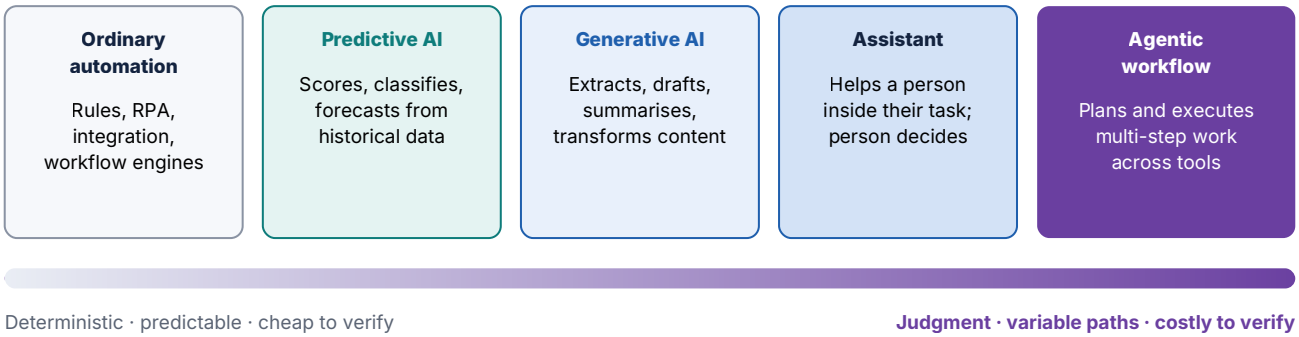


Figure 3.1 — From ordinary automation to agentic workflows

Type	Fits when…	Example	Watch-out
Ordinary automation	Inputs are structured and rules are explicit and stable	Route invoices by amount; sync records between systems	Brittle when rules change; but cheap and auditable
Predictive AI	A pattern in historical data predicts an outcome	Churn risk, failure prediction, fraud score	Needs labelled outcomes (label test)
Generative AI	Unstructured content must be read, drafted or transformed	Extract contract clauses; draft replies	Plausible errors; needs review design
Assistant	A person does the task and benefits from help in context	Agent assist; copilot in documents	Value depends on adoption and review
Agentic workflow	Multi-step work across systems with variable paths that require judgment	Diagnose and fix order fallout; resolve IT access requests	Permissions, guardrails, upkeep, escalation, verification

KEY IDEA · The least-autonomy principle

Choose the **least autonomous technique that achieves the outcome**. Every step up the spectrum adds cost to build, verify, govern and maintain. Autonomy must earn its place.

EXAM TRAP

Scenarios often describe a fixed, rule-based task (e.g., “if order value > €10k, route to manager”) and offer “deploy an AI agent” as an answer. The correct answer is **ordinary automation**: the rules are explicit and no judgment is involved.

3.3 Identifying AI automation candidates

Signal in the workflow	Points to
High volume of similar but not identical cases	AI classification, extraction or drafting
Unstructured inputs (e-mails, documents, notes, images)	Generative AI / deep learning
Experts making pattern judgments repeatedly	Predictive AI, or assistant support
Multi-step work across several systems with case-by-case paths	Agentic workflow candidate

Signal in the workflow	Points to
Many hand-offs where context is lost	Assistant or agent that carries context
Exceptions consuming disproportionate expert time	Agent handles routine; experts handle exceptions

3.4 Isolating the agentic surface

Most “agent” opportunities are workflows in which only a few steps genuinely need autonomy. The **agentic surface** is the part of the workflow where inputs vary, several paths are possible, and choosing among them requires judgment that rules cannot capture — *and* where acting without waiting for a person creates enough value to justify the controls.

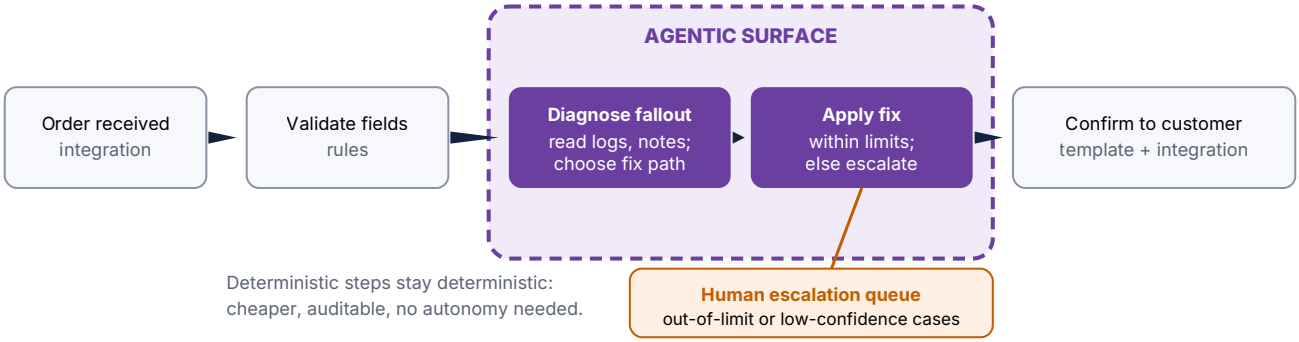


Figure 3.2 — Isolating the agentic surface in an order-fallout workflow

Test for each step	Agentic surface if...
Input variability	Inputs vary widely and are partly unstructured
Path variability	Several possible next actions, chosen case by case
Judgment	The choice cannot be reduced to stable rules
Value of acting	Waiting for a human materially costs time, money or experience
Controllability	Actions can be limited, logged and, ideally, reversed

BEST PRACTICE

Keep the surface **narrow**: surround it with deterministic steps (validation, templated confirmations, integrations) and an escalation route. A narrow surface is cheaper to verify, easier to govern and simpler to maintain.

3.5 Deployment layers

Layer	Description	Roadmap implications
Embedded in existing software	Agents or assistants offered within CRM, ITSM, office or ERP suites	Fast; limited control; vendor roadmap dependency; must enter intake
Low-code / platform agents	Built by business or IT on an enterprise agent platform	Speed and reach; requires citizen-developer governance

Layer	Description	Roadmap implications
Custom-built	Engineered by technical teams on the organization's platform	Most control and cost; for differentiating or high-risk workflows
Personal assistants	Individual productivity assistants with user-level access	Broad adoption; oversharing and policy risks

3.6 Citizen-developer governance

Low-code agent builders let business users create agents quickly. Without governance, they also create a new form of sprawl — dozens of agents with broad permissions, no owner and no upkeep plan.

Governance element	Good practice
Tiers of permission	Personal (own data, read-only) · team (shared data, approval) · enterprise (production actions, full review)
Registry	Every shared agent registered with owner, purpose, connections and permissions
Approved templates and connectors	Pre-approved patterns with guardrails built in
Review before promotion	Agents promoted to wider use pass a risk and quality review
Lifecycle	Owner re-certifies periodically; orphaned agents retired
Training	Builders trained on data, permissions and failure modes

3.7 Permissions, guardrails and system connections

Design element	Questions to frame for the roadmap
Permissions	Which systems and data can the agent read? Which can it write to? Under whose identity? Is it least privilege?
Action limits	Value, volume and type limits (e.g., refunds ≤ €200; no deletions)
Approval thresholds	Which actions require human approval (autonomy level 3)?
Guardrails	Input and output checks, blocked actions, policy rules, content filters, rate limits
System connections	Which APIs/connectors; their owners; stability; change notifications
Logging	Every action, input and decision logged and retained
Kill switch and rollback	Who can stop it; how actions are reversed
Security	Exposure to prompt injection through content the agent reads; tool misuse

PITFALL · Borrowed identity

Agents running under a broad service account can do far more than any user who asks them to act. Frame permissions explicitly in the opportunity — they drive security review effort and risk tier.

3.8 Ongoing upkeep, human escalation and verification

Upkeep item	Why it recurs
Prompts and instructions	Process and policy changes; new edge cases
Connectors and APIs	Upstream system changes break integrations
Model updates	Provider updates change behaviour; re-validation needed
Knowledge content	Policies, products and procedures change
Evaluation sets	Must reflect current cases to detect regressions
Permissions	Roles and systems change

Escalation & verification design	Framing question
Escalation triggers	Low confidence, out-of-limit actions, policy flags, customer requests for a human
Escalation capacity	Who handles escalations, how many per day, at what cost?
Verification	How are agent actions checked — pre-approval, sampling, reconciliation, outcome monitoring?
Feedback	How corrections improve the agent

> **KEY IDEA**
> Upkeep, escalation and verification are **permanent run costs**. They belong in the opportunity frame and the business case — they are often what separates an attractive demo from a viable roadmap item.

3.9 Reach, reversibility, workflow ownership and residual work

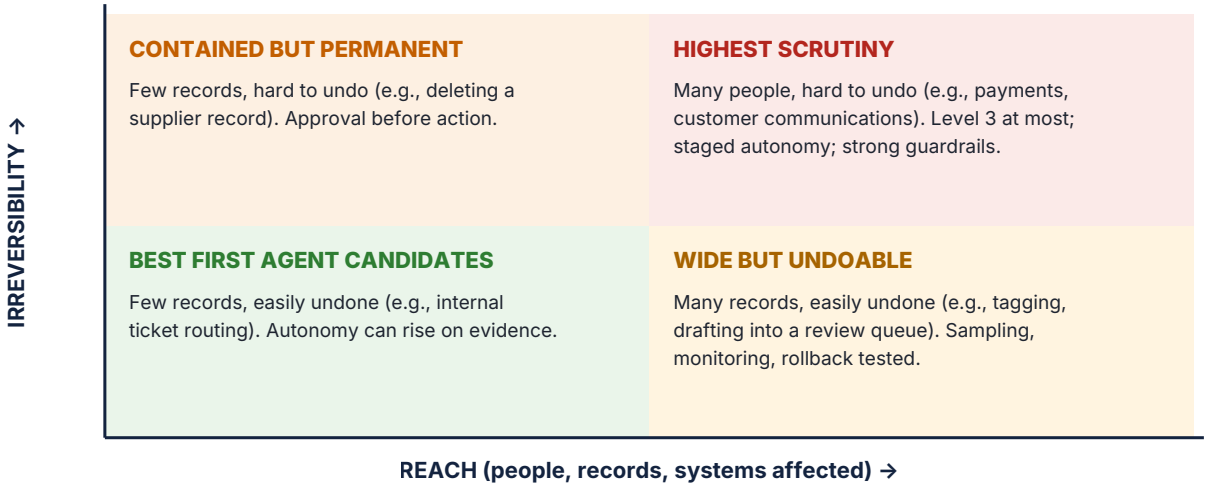


Figure 3.3 — Reach × irreversibility shapes autonomy and controls

Factor	Question	Effect on cost, risk, feasibility
Reach	How many people, records, systems and how much value does each action touch?	Higher reach → higher risk tier, more verification, staged rollout
Reversibility	Can an action be undone fully, quickly and cheaply?	Irreversible → approval before action; lower autonomy
Workflow ownership	Does one owner control the whole workflow, or does it cross functions?	Cross-functional → harder governance; needs a named end-to-end owner; lower feasibility
Residual work	What remains for humans: exceptions, approvals, verification, upkeep?	Large residual → smaller net value; must be staffed

EXAMPLE

An agent that fixes order fallout touches order, billing and provisioning systems owned by three departments. Without a named end-to-end owner, nobody can approve its permissions or answer for its actions — the opportunity is **not roadmap-ready** until ownership is resolved, whatever its technical feasibility.

3.10 Writing a roadmap-ready opportunity frame

Frame element	Content
Workflow and owner	End-to-end workflow, redesigned; named owner
Automation type per step	Ordinary automation / predictive / generative / assistant / agentic
Agentic surface	Steps with autonomy and target autonomy level
Deployment layer	Embedded / low-code / custom; builder governance
Permissions and guardrails	Read/write scope, limits, approvals, kill switch
Reach and reversibility	Quadrant; staged autonomy plan
Residual work and upkeep	Escalation volume and staffing; upkeep items and owners
Value (net)	Gross value minus residual work, upkeep and run costs
Preconditions	For the roadmap (Module 1)

3.11 Case study: Vireo Telecom

CASE STUDY **"An agent to fix all order fallout"**

Vireo Telecom (fictional) · broadband and mobile provider · 3.2m customers

Situation

About 9% of new orders "fall out" — they stop between ordering, provisioning and billing. A team of 60 resolves them manually. A vendor proposes an autonomous agent "to fix all order fallout end to end".

Framing work

- **Redesign first:** 35% of fallout came from an address-validation gap at order entry — fixed at source with ordinary automation (structured address lookup).
- **Spectrum:** of the remaining fallout, 40% followed six known patterns with explicit rules — ordinary automation; 45% required reading system logs and notes to diagnose — the **agentic surface**; 15% were complex exceptions — human experts with an assistant.
- **Permissions:** agent may read order, provisioning and billing systems; may write only re-submission and appointment-booking actions; no billing adjustments above €50; approval required for cancellations.
- **Reach/reversibility:** re-submissions reversible and contained → autonomy level 4 after a level-3 phase; customer credits less reversible and wider → level 3 only.
- **Ownership:** no one owned end-to-end order-to-activation; the COO appointed an owner as a precondition.
- **Residual work and upkeep:** ~25 escalations/day for 8 specialists; connectors to three systems with monthly releases; evaluation set refreshed quarterly.

Roadmap result

H1: address fix and rules automation (no AI); H1–H2: agent at level 3 on the agentic surface, preconditioned on owner, permissions approval and escalation team; H2: level 4 for re-submissions after 8 weeks of evidence.

Discussion questions

1. Why did redesign come first, and what share of the problem did it remove?
2. Why was the agentic surface limited to diagnosis and fix within limits?
3. How did reach and reversibility shape autonomy levels?
4. Why is the end-to-end owner a precondition?

Model answers (brief)

1. Fixing the address gap at source removed 35% of the fallout cheaply and without AI; automating it would have made waste faster.
2. Only those steps needed judgment over variable inputs; the rule-based patterns and confirmations were cheaper and more auditable as ordinary automation.
3. Reversible, contained re-submissions could reach level 4; credits are harder to reverse and touch money, so they stay at level 3.

4. The workflow crossed three departments. Without one owner, nobody can approve permissions, answer for actions or staff escalations.

3.12 Consulting templates

T3.1 Workflow Redesign & Automation Type Map

Purpose: Redesign first, then choose the least autonomous technique per step.

Workflow / end-to-end owner: .....

Step	Remove / simplify / move?	Input type	Path variability	Judgment needed?	Automation type (rule / predictive / generative / assistant / agentic / human)	Why

T3.2 Agentic Surface Worksheet

Purpose: Isolate the steps where autonomy is justified.

Step	Input variability	Path variability	Judgment beyond rules	Value of acting without waiting	Controllable (limit, log, reverse)	On surface?

Target autonomy level now / later and evidence to raise it: .....

Escalation route and triggers: .....

T3.3 Permissions, Guardrails & Connections Sheet

Purpose: Frame what the agent may do and how it is controlled.

System	Owner	Read	Write (which actions)	Limits	Approval needed?	Logged?

- [ ] Runs under least-privilege identity (not a broad service account)
- [ ] Blocked actions listed
- [ ] Guardrails on inputs and outputs defined
- [ ] Kill switch owner named; rollback tested
- [ ] Prompt-injection exposure considered for content the agent reads
- [ ] Deployment layer and builder tier (personal / team / enterprise) recorded

T3.4 Residual Work & Upkeep Ledger

Purpose: Price the permanent work an agent creates.

RESIDUAL WORK

Item	Volume / period	Who	Time per item	Annual cost
Escalations				
Approvals				
Verification / sampling				
Incident handling				

UPKEEP

Item	Trigger / frequency	Owner	Effort / period
Prompts & instructions			
Connectors & APIs			
Model updates & re-validation			
Knowledge content			
Evaluation set refresh			

T3.5 Agent Opportunity Frame (roadmap-ready)

Purpose: Summarise the opportunity for the roadmap.

Workflow and named end-to-end owner: .....

Redesign changes made first: .....

Agentic surface and target autonomy (now → later): .....

Deployment layer and builder governance tier: .....

Permissions summary and key limits: .....

Reach × reversibility quadrant: .....

Residual work and upkeep (annual cost): .....

Net value (range): .....

Risk tier and specialists to involve: .....

Roadmap preconditions: .....

3.13 Best practices and pitfalls

✓ DO THIS

- Redesign the workflow before automating it.
- Choose the least autonomous technique that works.
- Keep the agentic surface narrow.
- Govern citizen-built agents with tiers and a registry.
- Frame permissions, limits, approvals and kill switch.
- Price upkeep, escalation and verification as run costs.
- Require an end-to-end workflow owner.

× AVOID THIS

- Automating waste.
- Using agents for rule-based tasks.
- “End-to-end agent” for whole workflows.
- Letting low-code agents proliferate unregistered.
- Broad service-account permissions.
- Valuing the demo without residual work.
- Cross-functional agents with no owner.

3.14 Module 3 cheat sheet

MODULE 3 · ONE-PAGE SUMMARY

AI-Enabled Workflow and Agent Opportunities

REDESIGN FIRST

Remove · fix at source · parallelise · rule where possible · AI only for judgment on variable inputs.

SPECTRUM

Ordinary automation → predictive → generative → assistant → agentic. **Least autonomy that works.**

AGENTIC SURFACE

Input variability · path variability · judgment beyond rules · value of acting now · controllable. Keep it narrow, wrapped in deterministic steps + escalation.

DEPLOYMENT LAYERS

Embedded · low-code platform · custom · personal assistants. Citizen-developer governance: tiers, registry, templates, review, lifecycle, training.

PERMISSIONS & GUARDRAILS

Read/write scope · least-privilege identity · action limits · approval thresholds · guardrails · connections & owners · logging · kill switch & rollback · prompt-injection exposure.

UPKEEP & ESCALATION

Prompts · connectors · model updates · content · eval sets · permissions. Escalation triggers and capacity; verification design. Permanent run costs.

FRAMING FACTORS

Reach · **reversibility** · **workflow ownership** · **residual work** → cost, risk, feasibility, autonomy level.

ROADMAP-READY FRAME

Workflow + owner · redesign · automation type per step · surface & autonomy · layer · permissions · reach × reversibility · residual work & upkeep · net value · risk tier · preconditions.

3.15 Knowledge check

Answer all questions before turning the page.

K1 **Invoices over €10,000 must be routed to a manager; others are approved automatically. Which technique fits?**

- (A) Predictive AI
- (B) Generative AI
- (C) Ordinary automation (a rule in a workflow engine)
- (D) An AI agent

K2 **What is the "least-autonomy principle"?**

- (A) Never use agents.
- (B) Use the newest technology available.
- (C) Always keep a human in every step.
- (D) Choose the least autonomous technique that achieves the outcome, because each step up adds build, verification, governance and upkeep cost.

K3 **Which step most clearly belongs on the agentic surface?**

- (A) Checking that a postcode field is filled
- (B) Reading logs and notes to diagnose why an order stalled, then choosing among several fixes within limits
- (C) Sending a templated confirmation
- (D) Copying an order ID between systems

K4 **A team proposes "an agent that handles the entire workflow end to end". What is the best framing advice?**

- (A) Map and redesign the workflow, keep deterministic steps deterministic, and confine autonomy to a narrow agentic surface with escalation
- (B) Build it on the lowest-cost platform.
- (C) Approve it; end-to-end maximises value.
- (D) Reject agents entirely.

K5 **Business users are building dozens of low-code agents with broad data access. What governance is most needed?**

- (A) Vendor-managed governance
- (B) Tiered permissions, an agent registry with owners, approved templates, review before promotion and lifecycle re-certification
- (C) Ban low-code platforms.
- (D) Annual audits only

K6 **An agent runs under a service account with write access to all customer records. Which framing issue is this?**

- (A) Deployment layer
- (B) Upkeep
- (C) Residual work
- (D) Permissions: least privilege, scoped actions and limits are not designed

K7 **An agent's actions affect many customers and are hard to undo (e.g., issuing credits). What autonomy approach fits?**

- (A) Approval before action (level 3 at most), staged rollout and strong guardrails
- (B) Full autonomy to maximise speed
- (C) No controls, because credits are small
- (D) Sampling after the fact only

K8 **A workflow crosses order, billing and provisioning departments with no single owner. How does this affect the opportunity?**

- (A) The vendor can own it
- (B) Each department should build its own agent
- (C) It lowers feasibility; a named end-to-end owner should be a roadmap precondition
- (D) No effect if the technology works

3.16 Knowledge check: answers and explanations

K1 **Answer: C** Review: 3.2

Invoices over €10,000 must be routed to a manager; others are approved automatically. Which technique fits?

- ✗ (A) Predictive AI — **Wrong.** There is nothing to predict.
- ✗ (B) Generative AI — **Wrong.** No content needs to be generated.
- ✓ (C) Ordinary automation (a rule in a workflow engine) — **Correct.** Explicit, stable rules call for deterministic automation.
- ✗ (D) An AI agent — **Wrong.** No judgment or variable path is involved.

K2 **Answer: D** Review: 3.2

What is the "least-autonomy principle"?

- ✗ (A) Never use agents. — **Wrong.** Agents are appropriate where judgment justifies autonomy.
- ✗ (B) Use the newest technology available. — **Wrong.** Novelty is not a criterion.
- ✗ (C) Always keep a human in every step. — **Wrong.** Too absolute; some steps can be automated.
- ✓ (D) Choose the least autonomous technique that achieves the outcome, because each step up adds build, verification, governance and upkeep cost. — **Correct.** Autonomy must earn its place.

K3 **Answer: B** Review: 3.4

Which step most clearly belongs on the agentic surface?

- ✗ (A) Checking that a postcode field is filled — **Wrong.** A deterministic validation rule.
- ✓ (B) Reading logs and notes to diagnose why an order stalled, then choosing among several fixes within limits — **Correct.** Variable inputs, several paths and judgment.
- ✗ (C) Sending a templated confirmation — **Wrong.** Deterministic.
- ✗ (D) Copying an order ID between systems — **Wrong.** Integration.

K4 **Answer: A** Review: 3.1, 3.4

A team proposes "an agent that handles the entire workflow end to end". What is the best framing advice?

- ✓ (A) Map and redesign the workflow, keep deterministic steps deterministic, and confine autonomy to a narrow agentic surface with escalation — **Correct.** A narrow surface is cheaper, safer and easier to maintain.
- ✗ (B) Build it on the lowest-cost platform. — **Wrong.** Platform choice does not fix the scope problem.
- ✗ (C) Approve it; end-to-end maximises value. — **Wrong.** This ignores the cost of verifying and governing deterministic steps as autonomous ones.
- ✗ (D) Reject agents entirely. — **Wrong.** Some steps may justify autonomy.

K5 **Answer: B** Review: 3.6

Business users are building dozens of low-code agents with broad data access. What governance is most needed?

- ✗ (A) Vendor-managed governance — **Wrong.** The organization must govern its own agents.
- ✓ (B) Tiered permissions, an agent registry with owners, approved templates, review before promotion and lifecycle re-certification — **Correct.** Citizen-developer governance makes speed safe.
- ✗ (C) Ban low-code platforms. — **Wrong.** This drives the activity underground.
- ✗ (D) Annual audits only — **Wrong.** Too slow and too late.

K6 **Answer: D** Review: 3.7

An agent runs under a service account with write access to all customer records. Which framing issue is this?

✗ (A) Deployment layer — **Wrong.** Not the core issue.

✗ (B) Upkeep — **Wrong.** Not the issue.

✗ (C) Residual work — **Wrong.** Not the issue.

✓ (D) Permissions: least privilege, scoped actions and limits are not designed — **Correct.** A borrowed broad identity lets the agent do far more than needed.

K7 **Answer: A** Review: 3.9

An agent's actions affect many customers and are hard to undo (e.g., issuing credits). What autonomy approach fits?

✓ (A) Approval before action (level 3 at most), staged rollout and strong guardrails — **Correct.** This is the highest-scrutiny quadrant.

✗ (B) Full autonomy to maximise speed — **Wrong.** High reach and irreversibility call for strong controls.

✗ (C) No controls, because credits are small — **Wrong.** Reach makes small amounts add up.

✗ (D) Sampling after the fact only — **Wrong.** Too weak for irreversible, wide-reach actions.

K8 **Answer: C** Review: 3.9

A workflow crosses order, billing and provisioning departments with no single owner. How does this affect the opportunity?

✗ (A) The vendor can own it — **Wrong.** The vendor cannot own the client's workflow accountability.

✗ (B) Each department should build its own agent — **Wrong.** This fragments the workflow and duplicates effort.

✓ (C) It lowers feasibility; a named end-to-end owner should be a roadmap precondition — **Correct.** Cross-functional workflows need an owner before an agent can be governed.

✗ (D) No effect if the technology works — **Wrong.** Ownership determines who can approve permissions and answer for actions.

MODULE 4

Pressure-Testing the Roadmap

Turning an approved plan into a credible, fundable roadmap: testing assumptions, exposing fragile dependencies, running premortems and red-team reviews, loading controls with real volumes, naming concrete consequences and choosing the right disposition for each weakness.

YOU WILL BE ABLE TO

- Distinguish planning from testing and explain why independent review matters
- Separate incorrect assumptions from fragile dependencies
- Test assumptions and map dependencies for sequencing risk
- Run premortem, red-team and stage-gate reviews
- Test controls against operational load
- Identify failure modes and state concrete consequences
- Choose dispositions: correction, resequencing, fallback design, de-scoping, deferral or removal
- Evaluate a roadmap's value, feasibility, resilience and execution risk

IN THIS MODULE

1. Planning vs testing
2. Assumptions vs dependencies
3. Testing assumptions
4. Dependency mapping and sequencing risk
5. Premortem analysis
6. Red-team review
7. Stage-gate review of the roadmap
8. Testing controls against operational load
9. Failure modes and concrete consequences
10. Choosing a disposition
11. Evaluating the roadmap
12. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4 hours

An approved roadmap is a hypothesis about the future. Before money and people are committed to it, it should be attacked — deliberately, independently and concretely. This module teaches the techniques that expose hidden weaknesses: assumption tests, dependency fragility analysis, premortems, red-team and stage-gate reviews, control-load tests and failure-mode analysis — and the dispositions that turn each weakness into a better plan.

KEY TAKEAWAYS

1. **Planning** builds the case for a roadmap; **testing** looks for how it fails. Different mindsets, ideally different people.
2. Separate **incorrect assumptions** (beliefs that may be false) from **fragile dependencies** (true today, easily broken). They need different tests and fixes.
3. Use **premortems**, **red-team reviews** and **stage-gate reviews** to surface weaknesses before execution.
4. **Test controls against operational load**: oversight that works in a pilot can fail silently at scale.
5. Express each weakness as a **concrete consequence**: what breaks, when, for whom, at what cost.
6. Choose a **disposition**: correct, resequence, design a fallback, de-scope, defer or remove.
7. Evaluate the roadmap on **value, feasibility, resilience and execution risk**.

4.1 Planning vs testing

	Planning	Testing
Purpose	Build a coherent, persuasive plan	Find where and how the plan fails
Mindset	Constructive, optimistic, integrative	Sceptical, adversarial, specific
Typical question	“How do we make this work?”	“What would have to be true — and is it?”
Bias risk	Optimism, confirmation, sunk cost	Over-criticism without alternatives
Who	Roadmap team	Independent reviewers, red team, people closest to execution

KEY IDEA

The people who built a plan are the least able to see its weaknesses — not from lack of skill, but because they have spent weeks making it coherent. **Independence** is the core ingredient of a useful test.

4.2 Incorrect assumptions vs fragile dependencies

INCORRECT ASSUMPTION

A belief the plan relies on that may be **false**

- "Agents will accept 70% of drafts."
- "Claims history has usable labels."
- "Unit cost stays at €0.08."

Test: evidence, experiment, reference class

Fix: correct the plan to reality

FRAGILE DEPENDENCY

Something that is **true today** but could easily break

- One platform engineer knows the connectors.
- Vendor feature promised for Q2.
- ERP upgrade programme must finish first.

Test: single points of failure, chains, slack

Fix: redundancy, fallback, resequence

Figure 4.1 — Two different kinds of weakness

EXAM TRAP

"The vendor will deliver the agent feature in Q2" is a **dependency** (it may be true today as a commitment, but it is fragile). "Users will accept 70% of drafts" is an **assumption** (it may simply be false). Questions often test whether you would treat a fragile dependency with an assumption test — the right response is redundancy, fallback or resequencing.

4.3 Testing assumptions

1. **Surface** every assumption the plan relies on: value, adoption, data, cost, capacity, timing, regulation, vendor behaviour.
2. **Rank** by impact (what happens to the plan if false) × uncertainty (how little evidence exists).
3. **Test** the top-ranked ones: existing evidence, reference classes, small experiments, expert challenge.
4. **Correct** the plan where an assumption fails; record the result either way.

Assumption	Impact if false	Evidence today	Test
Agents accept 70% of drafts	High — value halves	Vendor claim (weak)	2-week trial with 10 agents on own data
Unit cost €0.08 per interaction	Medium	Pilot at low volume	Load test at projected volume; vendor price schedule
Legal review takes 4 weeks	Medium — delays H2	Past AI reviews took 7–11	Reference class from last 5 reviews
Works council agrees in 6 weeks	High — blocks rollout	None	Early briefing; ask for indicative timeline

4.4 Dependency mapping and sequencing risk

Fragility pattern	What to look for	Typical mitigation
Single point of failure	One person, team, vendor or system on which many items depend	Redundancy, documentation, second supplier
Long chains	A → B → C → D: any slip cascades	Shorten chains; decouple; buffers

Fragility pattern	What to look for	Typical mitigation
External dependencies	Vendor releases, other programmes, regulation	Fallback designs; contractual commitments; resequence
Parallel starts on one resource	Many items starting together on the same team	WIP limits; resequence
No slack on the critical path	Every item on the longest chain is tightly scheduled	Buffers; move non-critical work off the path
Hidden dependencies	Assumed availability of data, platforms, people	Explicit dependency register (Module 1)

4.5 Premortem analysis

A **premortem** asks a team to imagine that the roadmap has already failed — and to explain why. Because the failure is "given", people feel free to name risks they would otherwise keep to themselves.

Step	Action
1 · Set the scene	"It is 18 months from now. The AI roadmap has failed badly. Write down why."
2 · Individual writing	Each participant writes reasons silently (5–10 minutes) — avoids anchoring on the loudest voice
3 · Round-robin	Collect one reason per person per round until exhausted
4 · Cluster and rank	Group similar reasons; rank by likelihood and impact
5 · Convert	Turn top reasons into failure modes, tests and dispositions

> ✦ **CONSULTANT'S TIP · Who to invite**
> Include people close to execution — team leads, platform engineers, legal reviewers, frontline managers — not only the planning team and sponsors. They know where plans really break.

4.6 Red-team review

A **red team** is an independent group tasked with attacking the roadmap: finding the strongest arguments that it will not deliver value, cannot be executed as sequenced, or creates unacceptable risk.

Design element	Good practice
Independence	Members did not build the plan; report to the decision-holder, not the roadmap owner
Mandate	Written scope: value, feasibility, resilience, risk; access to all documents and people
Rules	Attack the plan, not the people; every critique must be specific and evidenced
Output	A short list of material weaknesses with consequences and suggested dispositions
Response	The roadmap team responds to each point in writing: accepted, partly accepted, rejected with reason

4.7 Stage-gate review of the roadmap itself

The roadmap is reviewed at its own gates (e.g., before approval, before each horizon's funding release). Each review asks: are the preconditions for the next stage met? Did the assumptions hold? Have dependencies changed? Does the remaining plan still make sense?

> **NOTE**
> A roadmap stage-gate review differs from an initiative gate (Module 1): it looks at the **whole portfolio** — balance, capacity, dependencies and cumulative risk — not just one initiative's evidence.

4.8 Testing controls against operational load

Controls designed in a pilot — reviewing 10% of outputs, approving every agent action above €200, handling escalations within an hour — are usually tested at pilot volumes. At scale, the same control can require more people than exist. When that happens, controls rarely fail loudly: reviews get shorter, samples smaller, queues longer, until the control exists only on paper.

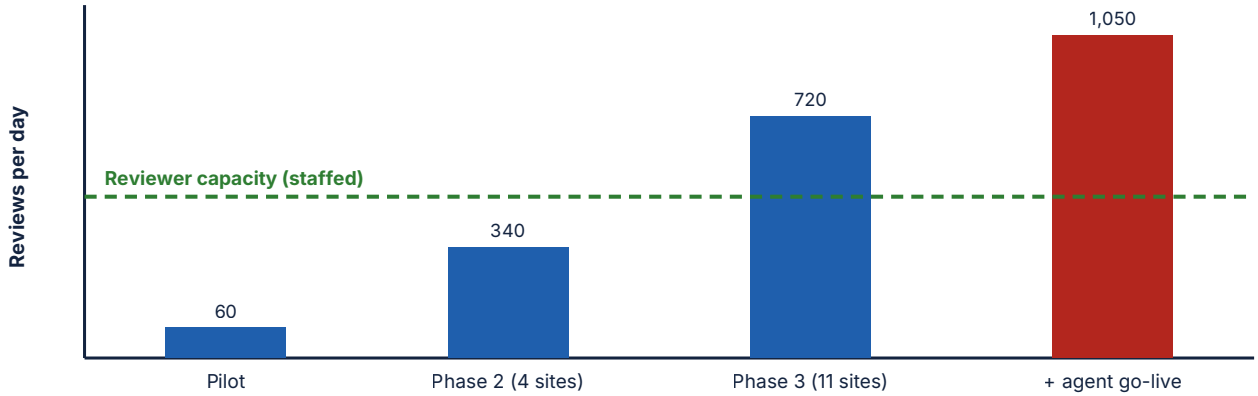


Figure 4.2 — Control load vs capacity across phases

Control	Load calculation	Question
Sampling review	Volume × sample rate × minutes per review	Reviewer hours available per phase?
Approval thresholds	Actions above threshold per day × approval time	Will approvers become a bottleneck or rubber-stamp?
Escalations	Volume × escalation rate × handling time	Specialist capacity? Customer wait times?
Incident handling	Expected incidents × investigation effort	Who investigates at 3 a.m.?
Monitoring	Number of systems × checks × frequency	Platform team capacity as systems multiply?

> **EXAMPLE**
> Agent assist at 11 sites: 120,000 interactions/month × 10% sample × 4 min = 800 reviewer hours/month ≈ 5 FTE. The plan staffed 1 FTE. Disposition: correct (staff 5 FTE) or redesign (risk-based sampling on 3% plus flagged cases) — documented before phase 3 funding.

4.9 Failure modes and concrete consequences

A weakness becomes actionable when it is stated as a **failure mode** with a **concrete consequence**: what breaks, when, for whom, at what cost — not "risk of delays".

Vague	Concrete consequence
"Data risk."	"If master-data harmonisation slips past March, forecasting misses the H2 gate; the €1.4m inventory saving moves at least two quarters and the replenishment agent (H3) loses its precondition."
"Key-person risk."	"If the sole connector engineer leaves, three live agents cannot be updated after the next ERP release; order fallout returns to manual handling (60 FTE) within weeks."
"Oversight risk."	"At phase 3 volume, sampling drops to under 1%, so a drift in reply quality could reach ~40,000 customers before detection."

Failure-mode field	Content
Failure mode	What fails
Cause	Why it would fail (assumption or dependency)
Consequence	Concrete effect: what, when, who, cost
Likelihood	H / M / L with reason
Detectability	Would we notice early? How?
Severity	Impact on value, risk, schedule
Disposition	Correct, resequence, fallback, de-scope, defer, remove

4.10 Choosing a disposition

Left to right: increasing impact on the roadmap's scope. Choose the lightest disposition that removes the consequence.

Figure 4.3 — Dispositions for roadmap weaknesses

Disposition	Use when	Example
Correction	An assumption is wrong and the plan can simply be fixed	Increase review staffing; correct cost estimate
Resequencing	Order creates avoidable risk or conflicts on resources	Move the agent after the platform hardening; stagger legal reviews
Fallback design	A fragile dependency cannot be removed	Manual process retained; second vendor option; feature flag
De-scoping	Full scope is not feasible, a smaller scope is	Two sites instead of eleven; level 3 instead of 4

Disposition	Use when	Example
Deferral	Preconditions cannot be met now but may be later	Defer until works agreement signed
Removal	The item does not survive testing on value, feasibility or risk	Drop the money pit; drop the duplicate

PITFALL · Accepting risk by silence

A weakness that is identified but given no disposition is accepted by default — usually by nobody in particular. Every material weakness needs a recorded disposition and an owner, even if the disposition is "accept, with monitoring".

4.11 Evaluating roadmap value, feasibility, resilience and execution risk

Criterion	Key questions	Evidence
Value	Does value survive assumption tests, the ledger and capture checks? Is it timed credibly?	Tested assumptions; captured-value plan
Feasibility	Can it be delivered within capacity, by component? Are preconditions achievable?	Capacity load; precondition forecasts
Resilience	Can it absorb a slip, a departure, a vendor change? Are there fallbacks?	Dependency fragility analysis; fallback designs
Execution risk	Do controls hold at load? Are owners and decision rights clear? Is sequencing risk managed?	Control-load tests; ownership matrix

4.12 Case study: Aurelia Pharma

CASE STUDY **Pressure-testing an approved roadmap before release of year-one funds**

Aurelia Pharma (fictional) · mid-size pharmaceutical company · commercial, medical and manufacturing functions

Situation

The executive board approved a three-horizon AI roadmap: medical-information assistant, field-force call notes, manufacturing deviation triage, and an H3 regulatory-submission drafting agent. Before releasing year-one funds, the CFO commissions an independent pressure test.

Findings (excerpt)

Finding	Type	Concrete consequence	Disposition
Medical-information assistant assumes 80% of answers need no medical review	Assumption (untested)	If review is required for most answers — likely in a regulated domain — value falls by ~70%	Correct: re-base value; run 4-week trial
All three H1 items need the same two platform engineers	Fragile dependency / SPOF	Any departure stalls all H1 items	Resequence + fallback: stagger; contractor framework
Deviation triage sampling 20% of AI classifications	Control load	At full volume = 3.5 FTE QA reviewers; plan has 0.5	Correct staffing or redesign sampling; precondition for scale
Submission drafting agent depends on unreleased vendor feature	Fragile dependency (external)	H3 timeline built on a vendor promise	Defer: start on feature release + evaluation on own documents
Call-notes tool overlaps the CRM's new embedded feature	Duplicate	Paying twice; two governance tracks	Remove: use embedded feature after review

Outcome

Year-one funding released for the corrected and resequenced plan; the CFO recorded the dispositions and named owners for each.

Discussion questions

1. Which findings were assumptions, which dependencies, and why does the distinction change the fix?
2. Why would the sampling control have failed silently?
3. Why defer, not remove, the submission agent?
4. Write the concrete consequence for the platform-engineer dependency in one sentence.

Model answers (brief)

1. The 80% no-review belief is an assumption, tested by a trial and fixed by correcting the plan. The shared engineers and the vendor feature are fragile dependencies, handled by staggering, a fallback and deferral.
2. Reviewers would cut samples and review time under load; the control would exist on paper only.
3. The value may be real; only the precondition (vendor feature, own-document evaluation) is missing — a trigger-based deferral keeps the option.
4. “If either platform engineer leaves in year one, all three H1 initiatives stall, delaying ~€2m of first-year value by at least two quarters.”

4.13 Consulting templates

T4.1 Assumption Register

Purpose: Surface, rank and test the beliefs the roadmap relies on.

#	Assumption	Impact if false (H/M/L)	Uncertainty (H/M/L)	Evidence today	Test	Owner	Result	Plan change

T4.2 Dependency Fragility Map

Purpose: Find single points of failure, chains and external dependencies.

Dependency	Items relying on it	Pattern (SPOF / chain / external / parallel load / no slack)	What breaks it	Early warning	Mitigation / fallback	Owner

T4.3 Premortem & Red-Team Worksheet

Purpose: Run a premortem and frame an independent red-team review.

PREMORTEM

Scenario: “It is ___ months from now and the roadmap has failed because…” .....

Participants (incl. people close to execution): .....

Top clustered reasons (ranked): .....

RED-TEAM CHARTER

Members and independence statement: .....

Reports to (decision-holder): .....

Scope (value / feasibility / resilience / risk): .....

Rules of engagement: .....

Output due / response due: .....

T4.4 Control Load Test

Purpose: Check that controls hold at each phase’s volume.

Control	Phase	Volume / period	Rate (sample / threshold / escalation)	Minutes each	Hours needed	Hours staffed	Gap	Disposition

T4.5 Failure Mode & Disposition Log

Purpose: Turn weaknesses into concrete consequences and decisions.

ID	Failure mode	Cause (assumption / dependency)	Concrete consequence (what, when, who, cost)	Likelihood	Detectability	Severity	Disposition	Owner

T4.6 Roadmap Review Scorecard

Purpose: Evaluate the roadmap before funding release.

Criterion	Rating (Strong / Adequate / Weak)	Key evidence	Material weaknesses	Conditions for release
Value				
Feasibility				
Resilience				
Execution risk				

Overall recommendation (release / release with conditions / revise and resubmit): .....

Reviewer(s) and independence: .....

4.14 Best practices and pitfalls

✓ DO THIS

- Separate planners from testers; use independent reviewers.
- Classify each weakness as assumption or dependency.
- Rank assumptions by impact × uncertainty and test the top ones.
- Run a premortem with people close to execution.
- Load-test every control at each phase's volume.
- State concrete consequences: what, when, who, cost.
- Record a disposition and owner for every material weakness.

× AVOID THIS

- Asking the plan's authors to test it.
- Treating fragile dependencies with assumption tests.
- Generic risk lists ("data risk", "people risk").
- Red teams that report to the roadmap owner.
- Assuming pilot-tested controls scale.
- Weaknesses identified but left without disposition.
- Defaulting to removal when a lighter disposition works.

4.15 Module 4 cheat sheet

MODULE 4 · ONE-PAGE SUMMARY

Pressure-Testing the Roadmap

PLANNING ≠ TESTING

Build the case vs find the failure. Independence is the key ingredient.

TWO WEAKNESSES

Assumption: may be false → test, correct. **Fragile dependency**: true now, easily broken → redundancy, fallback, resequence.

ASSUMPTION TESTING

Surface → rank (impact × uncertainty) → test (evidence, reference class, experiment) → correct and record.

FRAGILITY PATTERNS

Single point of failure · long chains · external · parallel load · no slack · hidden dependencies.

PREMORTEM

"It has failed — why?" Silent writing → round-robin → cluster & rank → failure modes. Invite executors.

RED TEAM & STAGE GATES

Independent, mandated, evidence-based critique reporting to the decision-holder; written responses. Roadmap gates review the whole portfolio.

CONTROL LOAD TEST

Volume × rate × minutes vs staffed hours, per phase. Controls fail **silently** under load (shorter reviews, smaller samples).

CONCRETE CONSEQUENCES

What breaks · when · for whom · cost. Fields: mode, cause, consequence, likelihood, detectability, severity, disposition.

DISPOSITIONS

Correct · resequence · fallback · de-scope · defer · remove — lightest that removes the consequence; owner for each.

EVALUATE THE ROADMAP

Value (survives tests, captured) · **feasibility** (capacity by component, preconditions) · **resilience** (fallbacks, slack) · **execution risk** (controls at load, owners, sequencing).

4.16 Knowledge check

Answer all questions before turning the page.

K1 **Why should the people who built a roadmap not be its only testers?**

(A) Having made the plan coherent, they are least able to see its weaknesses; independence surfaces what they miss.
(B) They lack technical skill.
(C) Planners are too busy.
(D) Testing is a compliance requirement.

K2 **"The vendor has committed to release the agent feature in Q2" is best classified as…**

(A) An incorrect assumption
(B) A fragile dependency — handled with fallback, resequencing or contractual protection
(C) A failure mode
(D) A control

K3 **Which assumption should be tested first?**

(A) High impact, low uncertainty
(B) Low impact, high uncertainty
(C) Low impact, low uncertainty
(D) High impact, high uncertainty

K4 **What makes a premortem effective?**

(A) Reasons are discussed aloud from the start
(B) Only senior sponsors participate
(C) Participants imagine the plan has failed and write reasons silently before sharing, including people close to execution
(D) It is run after execution

K5 **A red team reports to the roadmap owner, who edits its findings before the board sees them. What is wrong?**

(A) The board should not see critiques.
(B) Red teams should not write reports.
(C) Independence is compromised; the red team should report to the decision-holder, with the roadmap team responding in writing
(D) Nothing — the owner knows the plan best.

K6 **A 10% review sample worked in the pilot. At phase 3 it needs 5 FTE; 1 FTE is planned. What is most likely to happen without action?**

(A) Volume will fall.
(B) The control will degrade silently — shorter reviews and smaller samples — until it exists only on paper.
(C) The model will improve to compensate.
(D) Reviews will stop and someone will raise an alarm.

K7 **Which is a concrete consequence?**

(A) "There is data risk."
(B) "Stakeholders may be unhappy."
(C) "Delays are possible."
(D) "If harmonisation slips past March, forecasting misses its H2 gate and the €1.4m saving moves two quarters."

K8 **A use case's full 11-site rollout is infeasible within capacity, but two sites are. Which disposition fits?**

(A) De-scoping
(B) Removal
(C) Accept silently
(D) Correction

4.17 Knowledge check: answers and explanations

K1 **Answer: A** Review: 4.1

Why should the people who built a roadmap not be its only testers?

✓ **A** Having made the plan coherent, they are least able to see its weaknesses; independence surfaces what they miss. — **Correct.** Independence counters optimism and confirmation bias.

✗ **B** They lack technical skill. — **Wrong.** Skill is not the issue.

✗ **C** Planners are too busy. — **Wrong.** Not the core reason.

✗ **D** Testing is a compliance requirement. — **Wrong.** Not the reason.

K2 **Answer: B** Review: 4.2

"The vendor has committed to release the agent feature in Q2" is best classified as…

✗ **A** An incorrect assumption — **Wrong.** It may be true as a commitment; the issue is fragility.

✓ **B** A fragile dependency — handled with fallback, resequencing or contractual protection — **Correct.** True today but easily broken, and outside the organization's control.

✗ **C** A failure mode — **Wrong.** It is a cause of potential failure, not the failure itself.

✗ **D** A control — **Wrong.** Not a control.

K3 **Answer: D** Review: 4.3

Which assumption should be tested first?

✗ **A** High impact, low uncertainty — **Wrong.** Important but already well evidenced.

✗ **B** Low impact, high uncertainty — **Wrong.** Uncertain but unimportant.

✗ **C** Low impact, low uncertainty — **Wrong.** Least important.

✓ **D** High impact, high uncertainty — **Correct.** Rank by impact × uncertainty.

K4 **Answer: C** Review: 4.5

What makes a premortem effective?

✗ **A** Reasons are discussed aloud from the start — **Wrong.** This invites anchoring on the loudest voice.

✗ **B** Only senior sponsors participate — **Wrong.** This misses execution knowledge.

✓ **C** Participants imagine the plan has failed and write reasons silently before sharing, including people close to execution — **Correct.** Assumed failure plus independent writing surfaces hidden risks.

✗ **D** It is run after execution — **Wrong.** That is a postmortem.

K5 **Answer: C** Review: 4.6

A red team reports to the roadmap owner, who edits its findings before the board sees them. What is wrong?

✗ **A** The board should not see critiques. — **Wrong.** The decision-holder needs them.

✗ **B** Red teams should not write reports. — **Wrong.** They should, with specific evidence.

✓ **C** Independence is compromised; the red team should report to the decision-holder, with the roadmap team responding in writing — **Correct.** Independence and visible responses are core design elements.

✗ **D** Nothing — the owner knows the plan best. — **Wrong.** This compromises independence.

K6 **Answer: B** Review: 4.8

A 10% review sample worked in the pilot. At phase 3 it needs 5 FTE; 1 FTE is planned. What is most likely to happen without action?

✗ (A) Volume will fall. — **Wrong.** No reason to expect that.

✓ (B) The control will degrade silently — shorter reviews and smaller samples — until it exists only on paper. — **Correct.** This is why controls must be tested against operational load.

✗ (C) The model will improve to compensate. — **Wrong.** No mechanism for that.

✗ (D) Reviews will stop and someone will raise an alarm. — **Wrong.** Controls under load tend to degrade quietly.

K7 **Answer: D** Review: 4.9

Which is a concrete consequence?

✗ (A) "There is data risk." — **Wrong.** Vague.

✗ (B) "Stakeholders may be unhappy." — **Wrong.** Not concrete.

✗ (C) "Delays are possible." — **Wrong.** Vague.

✓ (D) "If harmonisation slips past March, forecasting misses its H2 gate and the €1.4m saving moves two quarters." — **Correct.** Names what breaks, when and the cost.

K8 **Answer: A** Review: 4.10

A use case's full 11-site rollout is infeasible within capacity, but two sites are. Which disposition fits?

✓ (A) De-scoping — **Correct.** Reduce scope to what is feasible, safely.

✗ (B) Removal — **Wrong.** Too severe; a smaller scope works.

✗ (C) Accept silently — **Wrong.** Every weakness needs a recorded disposition.

✗ (D) Correction — **Wrong.** Correction fixes a wrong assumption; here the scope must change.

R1 Course at a glance: the master cheat sheet

ALL FOUR MODULES ON ONE PAGE

Design AI Transformation Roadmaps

M1 · ROADMAP FUNDAMENTALS

Roadmap = use cases + enablers · milestones · dependencies · gates · capacity · owners. **Checkable preconditions, not dates**. Value × readiness: now · prepare · learn · later. Horizons H1 prove · H2 scale · H3 extend. **Pilot-to-scale is a separate decision**. Delivery + absorption capacity; downstream work; WIP limits. Operating model: hub-and-spoke — central standards, platform, assurance; local ownership, oversight, value capture.

M2 · INVESTMENT ACROSS HORIZONS

Place by **proof-maturity, dependencies, risk, readiness**. Balance: horizons, value timing, risk, learning, enablers, capacity, concentration. **Tranches** released at gates; stop rules. Binding constraint. Enablers funded at portfolio level. Under uncertainty: cheap information, reversible commitments. Self-funding only from captured value. Observable defer / fund / scale / stop conditions.

M3 · WORKFLOW AND AGENT OPPORTUNITIES

Redesign first. Spectrum: ordinary automation → predictive → generative → assistant → agentic; **least autonomy that works**. Narrow **agentic surface**. Deployment layers; citizen-developer governance. Permissions, limits, approvals, guardrails, connections, logging, kill switch. Upkeep, escalation, verification. **Reach · reversibility · ownership · residual work**.

M4 · PRESSURE-TESTING

Planning ≠ testing; independence. **Assumptions** (test, correct) vs **fragile dependencies** (redundancy, fallback, resequence). Premortem · red team · roadmap stage gates. **Control load tests**. Concrete consequences. Dispositions: correct · resequence · fallback · de-scope · defer · remove. Evaluate value · feasibility · resilience · execution risk.

R2 Template index

Code	Template	Use when
T1.1	Roadmap Canvas	First draft
T1.2	Initiative Precondition Card	Each initiative
T1.3	Dependency Register	Throughout
T1.4	Operating Model & Ownership Matrix	Operating model
T1.5	Pilot-to-Scale Gate Checklist	Scale decisions
T2.1	Horizon Classification Sheet	Placing initiatives
T2.2	Portfolio Balance Test	Before funding
T2.3	Funding Tranche & Gate Plan	Phased funding
T2.4	Capacity Load & Dependency Map	Constraints and enablers
T2.5	Stakeholder Perspective Grid	Funding sessions
T2.6	Defer / Fund / Scale / Stop Condition Register	Funding decisions
T3.1	Workflow Redesign & Automation Type Map	Workflow framing
T3.2	Agentic Surface Worksheet	Agent framing

Code	Template	Use when
T3.3	Permissions, Guardrails & Connections Sheet	Agent framing
T3.4	Residual Work & Upkeep Ledger	Net value
T3.5	Agent Opportunity Frame	Roadmap entry
T4.1	Assumption Register	Pressure test
T4.2	Dependency Fragility Map	Pressure test
T4.3	Premortem & Red-Team Worksheet	Independent review
T4.4	Control Load Test	Scale readiness
T4.5	Failure Mode & Disposition Log	Pressure test
T4.6	Roadmap Review Scorecard	Before funding release

R3 Glossary

A

Absorption capacity
The organization's ability to take on change and the new work AI creates (reviews, exceptions, training, support).

Agentic surface
The steps of a workflow where variable inputs and judgment justify autonomous action.

Agentic workflow
Multi-step work planned and executed by AI across tools with some autonomy.

Approval threshold
Limit above which an agent's action requires human approval.

Assistant
AI that helps a person within their task while the person decides and acts.

Assumption (roadmap)
A belief the plan relies on that may be false.

Automation spectrum
Ordinary automation → predictive AI → generative AI → assistant → agentic workflow.

B

Binding constraint
The scarce resource most over-committed by the portfolio.

C

Centralized operating model
A central AI team owns standards, platform and delivery.

Checkable precondition
An observable, verifiable, owned and thresholded condition that must be true before an initiative starts or advances.

Citizen developer
A business user building agents or automations on a low-code platform.

Concrete consequence
A precise statement of what breaks, when, for whom and at what cost.

Control load test
Comparing the effort a control requires at each phase's volume with staffed capacity.

Critical path
The longest chain of dependent work determining the earliest completion.

D

De-scoping
Disposition reducing scope to what is feasible and safe.

Decentralized operating model
Business units run AI independently; often the unplanned starting state.

Deferral
Disposition postponing an item until a trigger or precondition is met.

Delivery capacity
People and resources available to build, integrate, test and govern.

Dependency
A relationship in which one item needs another; typed and hard or soft.

Deployment layer
Where an agent runs: embedded in software, low-code platform, custom-built, personal assistant.

Disposition (roadmap)
Decision on a weakness: correct, resequence, fallback, de-scope, defer, remove.

Downstream work
Permanent run work created by each deployment: monitoring, review, upkeep, incidents, re-validation, training.

E

Enabler
An investment in data, platform, governance or people that unblocks use cases.

Escalation
Routing of cases an AI cannot or should not handle to a human.

Evidence milestone
A milestone that reduces uncertainty, e.g., pilot meets criteria.

F

Failure mode
A specific way in which the roadmap or an initiative could fail.

Fallback design
A planned alternative if a fragile dependency breaks.

Federated (hub-and-spoke) model
Central standards, governance and platform with local business execution.

Fragile dependency
Something true today on which the plan relies but which could easily break.

G

Guardrail
A technical or policy control constraining AI inputs, outputs or actions.

H

Horizon
A stage of the roadmap defined by proof-maturity: prove (H1), scale (H2), extend (H3).

K

Kill switch
Mechanism and named owner able to stop an AI system immediately.

L

Least privilege
Granting only the access strictly needed.

Least-autonomy principle
Choose the least autonomous technique that achieves the outcome.

M

Milestone
A verifiable point of progress: activity, delivery, evidence or decision.

O

Ordinary automation
Deterministic rules, RPA, integration and workflow engines.

P

Phase gate
A decision point with pre-agreed criteria and one decision-holder.

Pilot-to-scale decision
A separate decision, with its own criteria, to extend a proven pilot.

Portfolio balance
Distribution of spend and risk across horizons, value timing, enablers and capacity.

Premortem
Exercise imagining the plan has failed and explaining why, to surface risks.

Proof-maturity
How strongly an initiative has been shown to work in the organization's own context.

R

Reach
How many people, records, systems or how much value an AI action touches.

Real option
A small investment that preserves the right, not the obligation, to make a larger one later.

Red team
An independent group tasked with attacking a plan.

Removal
Disposition taking an item off the roadmap.

Resequencing
Disposition changing the order of initiatives to reduce risk or resource conflicts.

Residual work
Work that remains for humans after AI: exceptions, approvals, verification, upkeep.

Reversibility
Whether an AI action can be undone fully, quickly and cheaply.

Roadmap
A phased plan of use cases and enablers with milestones, dependencies and gates, bounded by capacity.

S

Self-funding
Using captured value from early initiatives to fund later ones.

Single point of failure
One person, team, vendor or system on which many items depend.

Stage-gate review (roadmap)
Portfolio-level review of preconditions, assumptions, dependencies and balance at roadmap gates.

Stop rule
A predefined result that ends funding.

T

Tranche
A portion of an approved envelope released when a gate's conditions are met.

U

Upkeep
Recurring maintenance of prompts, connectors, models, content, evaluation sets and permissions.

V

Value × readiness
Sequencing grid: now, prepare, learn, later.

Verification
Checking AI outputs or actions: pre-approval, sampling, reconciliation, outcome monitoring.

W

WIP limit
Maximum number of concurrent initiatives per scarce resource.

Workflow ownership
Whether a single named owner controls a workflow end to end.

Workflow redesign
Removing, simplifying or moving steps before automating.

R4 Mock exam

Format: 37 multiple-choice questions · one best answer each · **Time:** 60 minutes · **Suggested pass mark:** 70% (26 correct)

Questions 1–9 cover Module 1, 10–17 Module 2, 18–25 Module 3, 26–33 Module 4 and 34–37 are integrated scenarios. Answer every question before checking the answer key, which explains why each option is right or wrong.

Q1 **Which inputs does a roadmap primarily build on?**

- A Vendor product roadmaps
- B Competitors' announcements
- C The prioritized portfolio with dispositions and the readiness findings with remediation sequences
- D The IT budget alone

Q2 **Why do enablers deserve their own workstreams?**

- A Because hiding data, platform and governance work inside use cases leaves it unowned, unfunded and invisible as a dependency
- B Because enablers create value directly
- C To make the roadmap look larger
- D Because vendors require it

Q3 **Which milestone is strongest?**

- A "Platform installed"
- B "Pilot shows ≥ 25% handling-time reduction on 2,000 cases"
- C "Kick-off held"
- D "Training delivered"

Q4 **Which precondition is checkable?**

- A "Data mostly ready"
- B "Q2"
- C "Governance in place"
- D "Works council informed and agreement on log use signed"

Q5 **What happens to dates in a precondition-based roadmap?**

- A They become forecasts derived from preconditions and move visibly when preconditions slip.
- B They disappear entirely.
- C They are set by the vendor.
- D They are fixed and preconditions are ignored.

Q6 **A low-value use case is fully ready. Where does it sit on the value × readiness grid, and why might it still go early?**

- A Later/drop — never early
- B Prepare
- C Now — always
- D Learn — it can build reusable skills, platform and governance patterns

Q7 **Which scale criterion is most often overlooked?**

- A Whether the sponsor is enthusiastic
- B Whether the pilot had a name
- C Whether oversight and the owner's capacity will hold at scaled volume
- D Whether the pilot used AI

Q8 **Why limit work in progress per scarce resource?**

- A To satisfy auditors
- B Starting fewer initiatives and finishing them delivers value sooner than starting everything on a shared bottleneck
- C To reduce ambition
- D Because AI cannot run in parallel

Q9 **In a hub-and-spoke model, who owns use-case value capture?**

- A The business units (spokes)
- B Internal audit
- C The vendor
- D The central hub

Q10 **Which classification inputs place an initiative in a horizon?**

- A Vendor preference
- B Calendar year only
- C Proof-maturity, dependency position, risk, readiness and value timing
- D Sponsor seniority and budget size

Q11 **Which balance test would reveal that the two biggest bets have no experiments?**

- A Enabler share
- B Concentration
- C Value timing
- D Learning — does the portfolio buy information about the most uncertain bets?

Q12 **What is the main financial benefit of tranche-based funding?**

- A Faster delivery
- B The amount at risk is limited to the current tranche and money follows evidence
- C No need for gates
- D Lower total cost guaranteed

Q13 **Which is usually the binding constraint in AI portfolios?**

- A Often a scarce resource such as data engineers, platform teams, legal reviewers or absorption capacity
- B Always money
- C Model availability
- D Office space

Q14 **Why fund an enabler at portfolio level?**

- A Because enablers have no value
- B Because finance prefers it
- C Because they unlock several use cases but belong to none, so they risk being orphaned
- D To hide costs

Q15 **Which practice fits investment under high uncertainty?**

- A Full funding to move fast
- B Small experiments with stop rules; reversible commitments until evidence arrives
- C Waiting for certainty
- D Long contracts for discounts

Q16 **Under what condition can H1 savings fund H2?**

- A When the sponsor approves
- B Always
- C When they are measured
- D When they are captured — cost removed or capacity redeployed — and verified

Q17 **Which statement is an observable defer condition?**

- A "Defer until the CEO decides."
- B "Defer until master-data match rate ≥ 95% for the top-3 regions, reviewed monthly by the data owner."
- C "Defer until things improve."
- D "Defer to next year."

Q18 **What should happen before automating a workflow?**

- A Redesign: remove, simplify, fix at source, parallelise
- B Write a prompt
- C Train users
- D Select an agent platform

Q19 **A task requires reading varied customer e-mails and drafting replies that an agent then edits and sends. Which technique fits?**

- A Predictive AI
- B Fully autonomous agent
- C Assistant / generative drafting with human review
- D Ordinary automation

Q20 **Which statement about the agentic surface is correct?**

- A It should cover the whole workflow.
- B It is defined by the vendor.
- C It covers only deterministic steps.
- D It covers steps with variable inputs, multiple paths and judgment, where acting without waiting adds value and actions are controllable.

Q21 **Which is a citizen-developer governance element?**

- A Unlimited agent creation
- B An agent registry with owners, tiered permissions and review before promotion
- C Vendor-managed permissions only
- D No training needed

Q22 **Which permission design is best for an agent resolving IT access requests?**

- A Least-privilege identity, scoped write actions, limits, approval for sensitive groups, logging and a kill switch
- B Read-only everywhere
- C Domain-admin service account
- D The requester's own credentials

Q23 **Why must upkeep be in the opportunity frame?**

- A Vendors pay for it.
- B It is optional.
- C It is a one-off cost.
- D Prompts, connectors, model updates, content and evaluation sets require recurring effort that reduces net value

Q24 **Agent actions are contained to a few records and easily undone. What does this imply?**

- A No logging needed
- B It should be removed from the roadmap
- C A good first agent candidate; autonomy can rise on evidence
- D Highest scrutiny

Q25 **Residual work for an agent is estimated at 30 escalations per day. What must the frame include?**

- A Nothing — escalations are rare
- B The vendor's support hours
- C Who handles them, at what capacity and cost — deducted from net value and staffed as a precondition
- D A promise to reduce them later

Q26 **What distinguishes testing from planning?**

- A They are the same.
- B Testing uses spreadsheets.
- C Testing happens after execution.
- D Planning builds a coherent case; testing looks for how it fails, ideally independently

Q27 **"Users will accept 70% of AI drafts" is…**

- A A disposition
- B An assumption to test with evidence or a small experiment
- C A control
- D A fragile dependency

Q28 **Which pattern is a single point of failure?**

- (A) One platform engineer maintaining all connectors for live agents
- (B) Two vendors for one capability
- (C) Parallel teams
- (D) Documented runbooks

Q29 **What is the purpose of silent individual writing in a premortem?**

- (A) To create documentation
- (B) To save time
- (C) To assign blame
- (D) To avoid anchoring on the loudest or most senior voice so hidden risks surface

Q30 **To whom should a red team report?**

- (A) The roadmap owner
- (B) The vendor
- (C) The decision-holder, with the roadmap team responding in writing
- (D) No one

Q31 **A control requires 800 reviewer hours a month at phase 3; 160 are staffed. What is the right action?**

- (A) Record the gap as a failure mode and choose a disposition — staff up or redesign the sampling — as a precondition for phase 3
- (B) Accept — reviewers will speed up
- (C) Remove the control
- (D) Wait for incidents

Q32 **Which disposition fits a fragile external dependency that cannot be removed?**

- (A) Silence
- (B) Fallback design (and possibly resequencing)
- (C) Correction
- (D) Removal

Q33 **Which four criteria evaluate a roadmap in this course?**

- (A) Data, technology, people, process
- (B) Value, feasibility, resilience, execution risk
- (C) Strategy, structure, systems, style
- (D) Cost, speed, novelty, size

Q34 **Integrated: A sponsor wants an autonomous agent for a cross-functional workflow, starting in Q1. Readiness is "not yet" and there is no end-to-end owner. What belongs on the roadmap?**

- (A) Nothing — drop it
- (B) The agent in Q1 as requested
- (C) A vendor pilot to show progress
- (D) Redesign and ownership as preconditions, enablers in H1, the agent on the agentic surface in a later horizon, starting when preconditions are met

Q35 **Integrated: H1 shows strong pilot value, but captured savings are unverified and review capacity at scale is unknown. What should the H2 funding decision be?**

(A) Release the full H2 tranche
(B) Cancel H2
(C) Release on observable conditions: verified captured value and a passed control-load test for the scale step
(D) Fund H3 instead

Q36 **Integrated: A premortem reveals that three H1 initiatives depend on one platform team already at 120% load. Which combination fits?**

(A) Resequence under a WIP limit, add a fallback (contractor framework) and record the dependency as a fragility with an owner
(B) Start all three and add a buffer of two weeks
(C) Ignore it — the team will manage
(D) Remove all three

Q37 **Integrated: Which sequence reflects this course?**

(A) Design the roadmap with preconditions and an operating model → sequence investment across horizons with tranches → frame workflow and agent opportunities → pressure-test before funding release
(B) Pick an agent platform → roadmap
(C) Pressure-test → classify → design
(D) Fund → build → plan

R5 Mock exam: answer key and explanations

Quick answer grid

1 · C	**2** · A	**3** · B	**4** · D	**5** · A	**6** · D	**7** · C	**8** · B	**9** · A	**10** · C
11 · D	**12** · B	**13** · A	**14** · C	**15** · B	**16** · D	**17** · B	**18** · A	**19** · C	**20** · D
21 · B	**22** · A	**23** · D	**24** · C	**25** · C	**26** · D	**27** · B	**28** · A	**29** · D	**30** · C
31 · A	**32** · B	**33** · B	**34** · D	**35** · C	**36** · A	**37** · A			

Score	Interpretation	Next step
31–37	Exam-ready	Review missed items; run a real assessment with the templates
26–30	Pass level	Re-read cheat sheets for modules where you missed two or more
below 26	Not yet	Re-study weak modules using the “Review” references; retake in one week

Full explanations

Q1 **Answer: C** Review: M1 · 1.1

Which inputs does a roadmap primarily build on?

- ✗ (A) Vendor product roadmaps — **Wrong.** Context only.
- ✗ (B) Competitors’ announcements — **Wrong.** Signals, not inputs.
- ✓ (C) The prioritized portfolio with dispositions and the readiness findings with remediation sequences — **Correct.** These determine what goes on the roadmap and its preconditions.
- ✗ (D) The IT budget alone — **Wrong.** One constraint among several.

Q2 **Answer: A** Review: M1 · 1.2

Why do enablers deserve their own workstreams?

- ✓ (A) Because hiding data, platform and governance work inside use cases leaves it unowned, unfunded and invisible as a dependency — **Correct.** Separate lanes make enablers visible and owned.
- ✗ (B) Because enablers create value directly — **Wrong.** They mainly unlock value.
- ✗ (C) To make the roadmap look larger — **Wrong.** Not a reason.
- ✗ (D) Because vendors require it — **Wrong.** Irrelevant.

Q3 **Answer: B** Review: M1 · 1.3

Which milestone is strongest?

- ✗ (A) “Platform installed” — **Wrong.** Delivery milestone.
- ✓ (B) “Pilot shows ≥ 25% handling-time reduction on 2,000 cases” — **Correct.** Evidence milestone: it reduces uncertainty.
- ✗ (C) “Kick-off held” — **Wrong.** Activity.
- ✗ (D) “Training delivered” — **Wrong.** Activity milestone.

Q4 **Answer: D** Review: M1 · 1.4

Which precondition is checkable?

- ✗ A "Data mostly ready" — **Wrong.** Not thresholded.
- ✗ B "Q2" — **Wrong.** A date.
- ✗ C "Governance in place" — **Wrong.** Vague.
- ✓ D "Works council informed and agreement on log use signed" — **Correct.** Observable, verifiable and binary.

Q5 **Answer: A** Review: M1 · 1.4

What happens to dates in a precondition-based roadmap?

- ✓ A They become forecasts derived from preconditions and move visibly when preconditions slip. — **Correct.** Dates follow conditions.
- ✗ B They disappear entirely. — **Wrong.** They remain as forecasts.
- ✗ C They are set by the vendor. — **Wrong.** Not relevant.
- ✗ D They are fixed and preconditions are ignored. — **Wrong.** The reverse of the method.

Q6 **Answer: D** Review: M1 · 1.5

A low-value use case is fully ready. Where does it sit on the value × readiness grid, and why might it still go early?

- ✗ A Later/drop — never early — **Wrong.** Ready low-value items can build skills and patterns.
- ✗ B Prepare — **Wrong.** Prepare is for high value, not ready.
- ✗ C Now — always — **Wrong.** Now is for high value and ready.
- ✓ D Learn — it can build reusable skills, platform and governance patterns — **Correct.** The "learn" quadrant.

Q7 **Answer: C** Review: M1 · 1.7

Which scale criterion is most often overlooked?

- ✗ A Whether the sponsor is enthusiastic — **Wrong.** Not a criterion.
- ✗ B Whether the pilot had a name — **Wrong.** Irrelevant.
- ✓ C Whether oversight and the owner's capacity will hold at scaled volume — **Correct.** Controls and capacity are often tested only at pilot volume.
- ✗ D Whether the pilot used AI — **Wrong.** Trivial.

Q8 **Answer: B** Review: M1 · 1.8

Why limit work in progress per scarce resource?

- ✗ A To satisfy auditors — **Wrong.** Not the reason.
- ✓ B Starting fewer initiatives and finishing them delivers value sooner than starting everything on a shared bottleneck — **Correct.** WIP limits prevent queues and stalls.
- ✗ C To reduce ambition — **Wrong.** Not the purpose.
- ✗ D Because AI cannot run in parallel — **Wrong.** Incorrect.

Q9 **Answer: A** Review: M1 · 1.9

In a hub-and-spoke model, who owns use-case value capture?

✓ A The business units (spokes) — **Correct.** Local execution includes ownership, adoption and value capture.

✗ B Internal audit — **Wrong.** Assurance, not ownership.

✗ C The vendor — **Wrong.** No.

✗ D The central hub — **Wrong.** The hub owns standards and platform.

Q10 **Answer: C** Review: M2 · 2.1–2.2

Which classification inputs place an initiative in a horizon?

✗ A Vendor preference — **Wrong.** Not relevant.

✗ B Calendar year only — **Wrong.** Calendar alone hides the reasons.

✓ C Proof-maturity, dependency position, risk, readiness and value timing — **Correct.** These justify placement.

✗ D Sponsor seniority and budget size — **Wrong.** Not placement criteria.

Q11 **Answer: D** Review: M2 · 2.3

Which balance test would reveal that the two biggest bets have no experiments?

✗ A Enabler share — **Wrong.** Concerns enablers.

✗ B Concentration — **Wrong.** Concerns vendors and platforms.

✗ C Value timing — **Wrong.** Related, but not this test.

✓ D Learning — does the portfolio buy information about the most uncertain bets? — **Correct.** This test targets uncertainty reduction.

Q12 **Answer: B** Review: M2 · 2.4

What is the main financial benefit of tranche-based funding?

✗ A Faster delivery — **Wrong.** Not necessarily.

✓ B The amount at risk is limited to the current tranche and money follows evidence — **Correct.** Staged release limits exposure.

✗ C No need for gates — **Wrong.** Gates are essential.

✗ D Lower total cost guaranteed — **Wrong.** Not guaranteed.

Q13 **Answer: A** Review: M2 · 2.5

Which is usually the binding constraint in AI portfolios?

✓ A Often a scarce resource such as data engineers, platform teams, legal reviewers or absorption capacity — **Correct.** Identify and sequence around it.

✗ B Always money — **Wrong.** Money is often not binding.

✗ C Model availability — **Wrong.** Rarely binding today.

✗ D Office space — **Wrong.** Not typically.

Q14 **Answer: C** Review: M2 · 2.6

Why fund an enabler at portfolio level?

✗ (A) Because enablers have no value — **Wrong.** They unlock value.

✗ (B) Because finance prefers it — **Wrong.** Not the reason.

✓ (C) Because they unlock several use cases but belong to none, so they risk being orphaned — **Correct.** Portfolio funding prevents orphaned enablers.

✗ (D) To hide costs — **Wrong.** Not the purpose.

Q15 **Answer: B** Review: M2 · 2.7

Which practice fits investment under high uncertainty?

✗ (A) Full funding to move fast — **Wrong.** Maximises exposure.

✓ (B) Small experiments with stop rules; reversible commitments until evidence arrives — **Correct.** Buys information and keeps options open.

✗ (C) Waiting for certainty — **Wrong.** Forgoes learning.

✗ (D) Long contracts for discounts — **Wrong.** Irreversible before evidence.

Q16 **Answer: D** Review: M2 · 2.8

Under what condition can H1 savings fund H2?

✗ (A) When the sponsor approves — **Wrong.** Approval does not create cash.

✗ (B) Always — **Wrong.** Not if uncaptured.

✗ (C) When they are measured — **Wrong.** Measurement is not capture.

✓ (D) When they are captured — cost removed or capacity redeployed — and verified — **Correct.** Only captured value funds anything.

Q17 **Answer: B** Review: M2 · 2.10

Which statement is an observable defer condition?

✗ (A) “Defer until the CEO decides.” — **Wrong.** Not a condition.

✓ (B) “Defer until master-data match rate ≥ 95% for the top-3 regions, reviewed monthly by the data owner.” — **Correct.** Metric, threshold, scope, cadence and owner.

✗ (C) “Defer until things improve.” — **Wrong.** Not observable.

✗ (D) “Defer to next year.” — **Wrong.** A date.

Q18 **Answer: A** Review: M3 · 3.1

What should happen before automating a workflow?

✓ (A) Redesign: remove, simplify, fix at source, parallelise — **Correct.** Automating waste makes it faster waste.

✗ (B) Write a prompt — **Wrong.** Build detail.

✗ (C) Train users — **Wrong.** Later.

✗ (D) Select an agent platform — **Wrong.** Premature.

Q19 **Answer: C** Review: M3 · 3.2

A task requires reading varied customer e-mails and drafting replies that an agent then edits and sends. Which technique fits?

- ✗ (A) Predictive AI — **Wrong.** Nothing to predict.
- ✗ (B) Fully autonomous agent — **Wrong.** Not needed; the human sends.
- ✓ (C) Assistant / generative drafting with human review — **Correct.** A person decides and acts with AI help.
- ✗ (D) Ordinary automation — **Wrong.** Inputs are unstructured and variable.

Q20 **Answer: D** Review: M3 · 3.4

Which statement about the agentic surface is correct?

- ✗ (A) It should cover the whole workflow. — **Wrong.** It should be narrow.
- ✗ (B) It is defined by the vendor. — **Wrong.** The organization defines it.
- ✗ (C) It covers only deterministic steps. — **Wrong.** The reverse.
- ✓ (D) It covers steps with variable inputs, multiple paths and judgment, where acting without waiting adds value and actions are controllable. — **Correct.** That is the definition and the test.

Q21 **Answer: B** Review: M3 · 3.6

Which is a citizen-developer governance element?

- ✗ (A) Unlimited agent creation — **Wrong.** Leads to sprawl.
- ✓ (B) An agent registry with owners, tiered permissions and review before promotion — **Correct.** Core governance elements.
- ✗ (C) Vendor-managed permissions only — **Wrong.** Insufficient.
- ✗ (D) No training needed — **Wrong.** Builders need training.

Q22 **Answer: A** Review: M3 · 3.7

Which permission design is best for an agent resolving IT access requests?

- ✓ (A) Least-privilege identity, scoped write actions, limits, approval for sensitive groups, logging and a kill switch — **Correct.** Least privilege with controls.
- ✗ (B) Read-only everywhere — **Wrong.** It could not act on its surface.
- ✗ (C) Domain-admin service account — **Wrong.** Far beyond need.
- ✗ (D) The requester's own credentials — **Wrong.** Inappropriate and insecure.

Q23 **Answer: D** Review: M3 · 3.8

Why must upkeep be in the opportunity frame?

- ✗ (A) Vendors pay for it. — **Wrong.** Not necessarily.
- ✗ (B) It is optional. — **Wrong.** Without it, agents degrade.
- ✗ (C) It is a one-off cost. — **Wrong.** It recurs.
- ✓ (D) Prompts, connectors, model updates, content and evaluation sets require recurring effort that reduces net value — **Correct.** Upkeep is a permanent run cost.

Q24 **Answer: C** Review: M3 · 3.9

Agent actions are contained to a few records and easily undone. What does this imply?

✗ (A) No logging needed — **Wrong.** Logging is always needed.

✗ (B) It should be removed from the roadmap — **Wrong.** No reason.

✓ (C) A good first agent candidate; autonomy can rise on evidence — **Correct.** Low reach and reversible.

✗ (D) Highest scrutiny — **Wrong.** That is wide reach and irreversible.

Q25 **Answer: C** Review: M3 · 3.9

Residual work for an agent is estimated at 30 escalations per day. What must the frame include?

✗ (A) Nothing — escalations are rare — **Wrong.** Thirty per day is material.

✗ (B) The vendor's support hours — **Wrong.** Escalations are business work.

✓ (C) Who handles them, at what capacity and cost — deducted from net value and staffed as a precondition — **Correct.** Residual work shapes cost and feasibility.

✗ (D) A promise to reduce them later — **Wrong.** Not a plan.

Q26 **Answer: D** Review: M4 · 4.1

What distinguishes testing from planning?

✗ (A) They are the same. — **Wrong.** They are not.

✗ (B) Testing uses spreadsheets. — **Wrong.** Irrelevant.

✗ (C) Testing happens after execution. — **Wrong.** It should happen before.

✓ (D) Planning builds a coherent case; testing looks for how it fails, ideally independently — **Correct.** Different purposes and mindsets.

Q27 **Answer: B** Review: M4 · 4.2–4.3

"Users will accept 70% of AI drafts" is…

✗ (A) A disposition — **Wrong.** No.

✓ (B) An assumption to test with evidence or a small experiment — **Correct.** Assumptions are tested and corrected.

✗ (C) A control — **Wrong.** No.

✗ (D) A fragile dependency — **Wrong.** It is a belief that may be false.

Q28 **Answer: A** Review: M4 · 4.4

Which pattern is a single point of failure?

✓ (A) One platform engineer maintaining all connectors for live agents — **Correct.** Many items depend on one person.

✗ (B) Two vendors for one capability — **Wrong.** Redundancy.

✗ (C) Parallel teams — **Wrong.** Not a SPOF.

✗ (D) Documented runbooks — **Wrong.** A mitigation.

Q29 **Answer: D** Review: M4 · 4.5

What is the purpose of silent individual writing in a premortem?

✗ (A) To create documentation — **Wrong.** A by-product.

✗ (B) To save time — **Wrong.** Not the main purpose.

✗ (C) To assign blame — **Wrong.** The opposite of the intent.

✓ (D) To avoid anchoring on the loudest or most senior voice so hidden risks surface — **Correct.** Independent idea generation.

Q30 **Answer: C** Review: M4 · 4.6

To whom should a red team report?

✗ (A) The roadmap owner — **Wrong.** This compromises independence.

✗ (B) The vendor — **Wrong.** No.

✓ (C) The decision-holder, with the roadmap team responding in writing — **Correct.** Preserves independence and accountability.

✗ (D) No one — **Wrong.** Findings must reach the decision.

Q31 **Answer: A** Review: M4 · 4.8

A control requires 800 reviewer hours a month at phase 3; 160 are staffed. What is the right action?

✓ (A) Record the gap as a failure mode and choose a disposition — staff up or redesign the sampling — as a precondition for phase 3 — **Correct.** Controls must hold at load.

✗ (B) Accept — reviewers will speed up — **Wrong.** Speeding up is how the control fails silently.

✗ (C) Remove the control — **Wrong.** Removing it increases risk.

✗ (D) Wait for incidents — **Wrong.** Too late.

Q32 **Answer: B** Review: M4 · 4.10

Which disposition fits a fragile external dependency that cannot be removed?

✗ (A) Silence — **Wrong.** Accepting risk by default.

✓ (B) Fallback design (and possibly resequencing) — **Correct.** Plan B for a dependency that may break.

✗ (C) Correction — **Wrong.** Correction fixes wrong assumptions.

✗ (D) Removal — **Wrong.** Too severe if the value is real.

Q33 **Answer: B** Review: M4 · 4.11

Which four criteria evaluate a roadmap in this course?

✗ (A) Data, technology, people, process — **Wrong.** Readiness-style categories.

✓ (B) Value, feasibility, resilience, execution risk — **Correct.** These are the four.

✗ (C) Strategy, structure, systems, style — **Wrong.** A generic framework.

✗ (D) Cost, speed, novelty, size — **Wrong.** Not the criteria.

Q34 **Answer: D** Review: M1 · 1.4, M3 · 3.9

Integrated: A sponsor wants an autonomous agent for a cross-functional workflow, starting in Q1. Readiness is "not yet" and there is no end-to-end owner. What belongs on the roadmap?

✗ (A) Nothing — drop it — **Wrong.** The value may be real.

✗ (B) The agent in Q1 as requested — **Wrong.** Ignores readiness and ownership.

✗ (C) A vendor pilot to show progress — **Wrong.** Likely to stall.

✓ (D) Redesign and ownership as preconditions, enablers in H1, the agent on the agentic surface in a later horizon, starting when preconditions are met — **Correct.** Combines Modules 1 and 3.

Q35 **Answer: C** Review: M2 · 2.4, 2.8; M4 · 4.8

Integrated: H1 shows strong pilot value, but captured savings are unverified and review capacity at scale is unknown. What should the H2 funding decision be?

✗ A Release the full H2 tranche — **Wrong.** Gate conditions are unmet.

✗ B Cancel H2 — **Wrong.** There is no reason to cancel.

✓ C Release on observable conditions: verified captured value and a passed control-load test for the scale step — **Correct.** Tranches follow evidence; controls must hold at load.

✗ D Fund H3 instead — **Wrong.** This skips proof.

Q36 **Answer: A** Review: M1 · 1.8, M4 · 4.4, 4.10

Integrated: A premortem reveals that three H1 initiatives depend on one platform team already at 120% load. Which combination fits?

✓ A Resequence under a WIP limit, add a fallback (contractor framework) and record the dependency as a fragility with an owner — **Correct.** Capacity, sequencing and fragility handled together.

✗ B Start all three and add a buffer of two weeks — **Wrong.** A buffer does not fix 120% load.

✗ C Ignore it — the team will manage — **Wrong.** Accepting risk by silence.

✗ D Remove all three — **Wrong.** Too severe.

Q37 **Answer: A** Review: M1–M4

Integrated: Which sequence reflects this course?

✓ A Design the roadmap with preconditions and an operating model → sequence investment across horizons with tranches → frame workflow and agent opportunities → pressure-test before funding release — **Correct.** This is the course arc.

✗ B Pick an agent platform → roadmap — **Wrong.** Technology-first.

✗ C Pressure-test → classify → design — **Wrong.** Design comes first.

✗ D Fund → build → plan — **Wrong.** Reversed.

R6 Further reading and next steps

Useful public companions to this course. Check for the latest versions.

Resource	Why it is useful
Literature on real options and staged investment	Background for tranche funding and sequencing under uncertainty
Premortem method (Gary Klein) and red-teaming practice	Techniques for independent pressure-testing
Reference-class forecasting and planning-fallacy research	Correcting optimistic schedules and benefits
NIST AI RMF 1.0 and ISO/IEC 42001	Governance expectations for the operating model and scaling
EU AI Act — deployer obligations (confirm current dates)	Preconditions for high-risk and agentic uses

Practice assignments

1. Take a portfolio you know; build a roadmap canvas (T1.1) and precondition cards (T1.2) for its top five initiatives.
2. Classify ten initiatives by horizon (T2.1), run the balance test (T2.2) and write tranche plans (T2.3) for two of them.
3. Map one workflow (T3.1), isolate its agentic surface (T3.2) and complete an agent opportunity frame (T3.5).
4. Run a 45-minute premortem with colleagues, complete a control load test (T4.4) and a failure mode log (T4.5), and score the roadmap (T4.6).

ENTERPRISE AI TRANSFORMATION AND CONSULTING PROFESSIONAL

COURSE 5 · STUDY BOOK

AI Business Cases and ROI Models

How to build AI investment cases that finance can verify, challenge and approve — or decline: traceable benefits, honest costs, a transparent ROI model, measurable value metrics and a decision-ready case that keeps a credible “no”.

4 MODULES · CASE STUDIES · TEMPLATES · CHEAT SHEETS · MOCK EXAM

Study edition · 2026

Includes full answer explanations

How to use this book

This study book accompanies **Course 5, AI Business Cases and ROI Models**, of the Enterprise AI Transformation and Consulting Professional certificate. It is written for consultants, advisors, management and finance analysts, and transformation leads who must justify, challenge or approve enterprise AI investments.

Each module follows the same rhythm so that you always know where you are:

Element	What you get	How to use it
Module opener	Learning objectives, section list, study time	Read first; return to it to self-test at the end
Overview & key takeaways	The module on one screen	Skim before, re-read after
Core sections	Concepts, tables, diagrams, worked examples	Study actively: annotate, restate in your own words
Callout boxes	Key ideas, examples, pitfalls, exam traps, consultant tips	Exam traps are the most frequently tested ideas
Case study	A realistic enterprise scenario with model answers	Answer the questions before reading the model answers
Templates	Fill-in consulting tools (T-codes), one per page or less	Copy into your own engagements
Do / Avoid	Best practices and pitfalls side by side	Use as a pre-meeting checklist
Cheat sheet	One-page module summary	Revise from these in the final week
Knowledge check	Scenario questions with explanations for **every** option	Understanding why the wrong answers are wrong matters as much as getting the right one

Callout key

KEY IDEA
The idea to remember if you remember nothing else.

EXAM TRAP
A frequently tested distinction or a tempting wrong answer.

EXAMPLE
A worked illustration of the concept.

CONSULTANT'S TIP
Practical wording or technique from the field.

PITFALL
A common mistake made in real engagements.

DEFINITION
A precise definition worth learning word for word.

> **NOTE · Builds on Courses 1–4**
> This course turns earlier work into money terms: value hypotheses and the effort ledger (Course 2), readiness findings (Course 3), roadmap horizons, tranches and gates (Course 4), and the evidence ladder and decision rights (Course 1). Financial concepts are taught at **advisory depth**: enough to build, read and challenge a model with finance — not to replace your finance team's methods, discount rates or accounting policies.

Suggested study plan (four weeks)

Week	Study	Practise
1	Module 1	Build a value driver tree and benefit traceability register for one AI initiative
2	Module 2	Rebuild the worked ROI model in a spreadsheet; add your own hidden costs and a tornado chart
3	Module 3	Write a measurement plan with baselines, paired metrics, comparison design and Y1/Y2 targets
4	Module 4, then revise all cheat sheets and the glossary	Draft a two-page investment committee memo; take the mock exam (60 min) and review every explanation

> **NOTE · About the examples**
> All organizations in case studies and examples are fictional, and their figures are illustrative. Regulatory and framework references are for orientation only; always check current law and guidance, and refer legal interpretation to qualified specialists.

Contents

FRONT **How to use this book** 2

MODULE 1 **Structuring the AI Business Case** 6
What a decision-ready business case is for 7
The structure of an AI business case 7
Value drivers and driver trees 8
Benefit traceability: landing line and accountable owner 8
P&L and cash impact at advisory depth 9
Hard ROI, soft ROI and strategic option value 10
Separating efficiency gains from new-value claims 10
Adjusting pilot results for production conditions 10
Scenarios, sensitivity and capital budgeting under uncertainty 11
Assumptions and decision quality 11
Executive summary logic and investment committee expectations 12
Case study: Hollin Water 13
Consulting templates 15
Best practices and pitfalls 17
Module 1 cheat sheet 18
Knowledge check 19
Knowledge check: answers and explanations 21

MODULE 2 **Building the ROI Model** 23
Model architecture 24
Total cost of ownership 24
The hidden costs of AI adoption 25
The cost register and AI spend governance 25
Benefit timing and adoption ramps 26
Payback, discounted cash flow, NPV and IRR 26
Staged investment 27
Sensitivity, scenarios and breakeven 28
Benefit realization 28
ROI assumptions finance will challenge 29
Distinguishing value created from effort relocated 29
Consulting templates 30
Best practices and pitfalls 33
Module 2 cheat sheet 34
Knowledge check 35
Knowledge check: answers and explanations 37

MODULE 3 **Defining Value Metrics** 39
The metric hierarchy 40
Adoption and usage vs business value — and vanity metrics 40
Connecting to indicators budget owners already review 41
Hard-dollar benefits vs softer productivity gains 41
Baselines and benchmarks 42
Paired quality metrics 42
Comparison designs 43
Attribution levels and claim strength 43

Leading and lagging indicators as adoption matures 44
Proxy measures — without overstating value 44
Setting year-1 and year-2 targets 44
The AI measurement plan 45
Case study: Arden Bank 46
Consulting templates 47
Best practices and pitfalls 49
Module 3 cheat sheet 50
Knowledge check 51
Knowledge check: answers and explanations 53

MODULE 4 **Making the Decision-Ready Case 55**
What decision-ready means 56
Formats: CFO-ready case, investment committee memo, board pre-read 56
The 60-second test 57
Separating strategic rationale from quantitative analysis 57
Connecting financial claims to evidence 58
Conservative scenarios 58
The cost and risk of inaction 59
Roadmap and readiness alignment 59
Clear decision requests 59
Preserving a credible null recommendation 60
Case study: Pellworth Hotels 61
Consulting templates 63
Best practices and pitfalls 65
Module 4 cheat sheet 66
Knowledge check 67
Knowledge check: answers and explanations 69

BACK **Course at a glance: the master cheat sheet 71**
Template index 71
Glossary 73
Mock exam 75
Mock exam: answer key and explanations 81
Further reading and next steps 88

MODULE 1

Structuring the AI Business Case

Building a credible case around value drivers, traceable benefits and real P&L impact — distinguishing hard ROI, soft ROI and option value, adjusting pilot results for production, and presenting ranges and assumptions finance can verify and challenge.

YOU WILL BE ABLE TO

- Structure an AI business case that enables an approve-or-decline decision
- Build value driver trees and trace every benefit to a landing line and accountable owner
- Explain P&L and cash impact at advisory depth
- Distinguish hard ROI, soft ROI and strategic option value
- Separate efficiency gains from new-value claims
- Adjust pilot results for production conditions
- Use scenarios, sensitivity and ranges for capital budgeting under uncertainty
- Write executive-summary logic that meets investment committee expectations

IN THIS MODULE

1. What a decision-ready case is for
2. The structure of an AI business case
3. Value drivers
4. Benefit traceability: landing line and owner
5. P&L and cash impact
6. Hard ROI, soft ROI, option value
7. Efficiency vs new value
8. Adjusting pilot results for production
9. Scenarios, sensitivity and ranges
10. Assumptions and decision quality
11. Executive summary and investment committee expectations
12. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4–5 hours

An AI business case exists to enable a decision: approve, approve in stages, or decline. Finance leaders will approve only what they can **verify** and **challenge**. This module teaches the structure that makes a case verifiable — value drivers, traceable benefits with a landing line and an owner, honest classification of hard, soft and option value, production-adjusted pilot results, and ranges rather than single numbers.

KEY TAKEAWAYS

1. A decision-ready case lets finance **verify, challenge and approve — or decline** — on transparent evidence.
2. Build value from **driver trees**: measurable drivers, each with a source and a conversion to money.
3. Trace every benefit to a **landing line** (where it appears in the P&L or budget) and an **accountable owner**.
4. Separate **hard ROI** (cash, P&L), **soft ROI** (real but not banked) and **strategic option value** — and never add them together silently.
5. Separate **efficiency gains** from **new-value claims**; they need different evidence.
6. **Adjust pilot results** for production: users, novelty, data mix, adoption and the ledger.
7. Present **ranges, scenarios and sensitivities**; the quality of assumptions drives the quality of the decision.

1.1 What a decision-ready business case is for

A business case is not a sales document. Its job is to give the decision-holder the information to choose well — including choosing **not** to invest. Cases written to persuade tend to overstate benefits, understate costs and hide assumptions; finance teams know this, and discount them accordingly. Cases written to inform earn trust and get decided faster.

KEY IDEA

The test of a good AI business case: **could a sceptical finance partner trace every number to its source, change any assumption, and see what happens to the answer?**

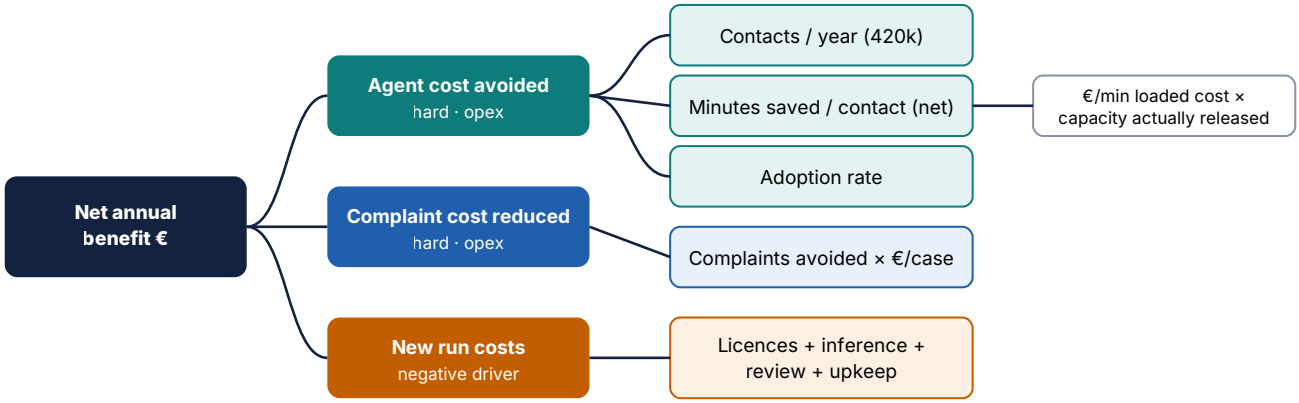
1.2 The structure of an AI business case

Section	Content	Typical length
1 · Executive summary	Recommendation, decision requested, return range, key assumptions, main risks	1 page
2 · Problem and strategic rationale	Business problem, why now, strategic fit, alternatives (incl. do nothing, non-AI)	1 page
3 · Solution and scope	What will be deployed, where, at what autonomy, roadmap horizon, readiness finding	½–1 page
4 · Benefits	Driver tree; traceability register (landing line, owner, timing, evidence)	1–2 pages
5 · Costs	One-off and recurring, TCO, hidden costs (Module 2)	1 page

Section	Content	Typical length
6 · Financial analysis	Cash flows, NPV, payback, scenarios, sensitivity	1–2 pages
7 · Risks and assumptions	Assumption register; risks; dependencies	1 page
8 · Measurement and verification	Metrics, baselines, targets, benefit-tracking plan (Module 3)	½–1 page
9 · Decision request	What is asked, how much, in which tranches, under what conditions	½ page

1.3 Value drivers and driver trees

A **value driver** is a measurable operational factor that causes value to change: volume, time per unit, error rate, conversion rate, price, loss rate. A **driver tree** decomposes the financial outcome into drivers, so each can be estimated, sourced and challenged separately.

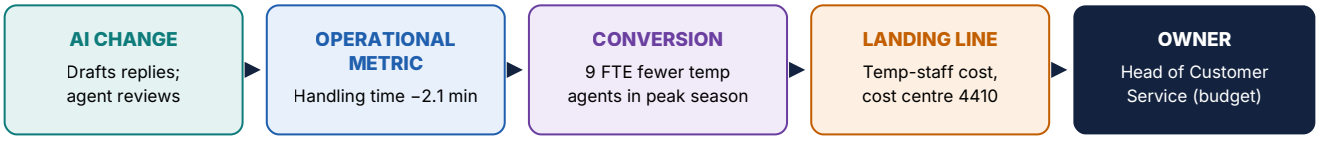


Every leaf is a measurable driver with a source; every branch states how it converts to money.

Figure 1.1 — Value driver tree for a customer-service drafting assistant

> **CONSULTANT'S TIP**
> Include **negative drivers** (new run costs, moved effort) in the same tree. Keeping them visible next to the benefits prevents the "gross saving" illusion from Course 2.

1.4 Benefit traceability: landing line and accountable owner



If any link is missing, the benefit cannot be verified — and finance will not count it.

Figure 1.2 — The traceability chain

DEFINITION · Landing line

The specific line in the P&L, budget or balance sheet where a benefit will appear — e.g., "temporary staff costs, cost centre 4410", "bad-debt write-offs", "revenue, product line X". If you cannot name it, finance cannot see the benefit arrive.

DEFINITION · Accountable benefit owner

The budget holder for the landing line, who agrees that the benefit is achievable and accepts accountability for realising it — usually by allowing the budget or target to change.

Weak benefit statement	Traceable benefit statement
"Saves 12,000 agent hours."	"Reduces peak-season temporary staffing by 9 FTE (€310–380k/yr) in temp-staff costs, CC 4410; owner: Head of Customer Service, who will reduce the 2027 temp budget accordingly."
"Improves customer experience."	"Cuts complaints about incorrect answers by 15–25% (baseline 6,200/yr × €45 handling) = €42–70k/yr in complaint-handling costs; owner: Head of Complaints."

EXAM TRAP

A benefit without a landing line and owner is, at best, **soft ROI**. An option that counts it as hard savings is usually wrong — even if the hours saved are real.

1.5 P&L and cash impact at advisory depth

Impact area	AI examples	Advisory note
Revenue	Higher conversion, retention, new products	Requires market evidence; watch cannibalisation
Cost of goods / services	Less waste, fewer service visits	Usually volume × unit-cost drivers
Operating expenses	Staff, contractors, outsourcing, complaints	Only counts if spend actually falls or growth is avoided
Capital expenditure / depreciation	Build costs that may be capitalised	Capitalisation follows accounting policy — confirm with finance
Working capital	Faster invoicing, better inventory	Cash benefit, often one-off
Risk / provisions	Lower fraud losses, fewer penalties	Use expected-loss logic; agree with risk and finance

NOTE

Cash vs P&L timing: a capitalised build cost hits cash now and the P&L over several years through depreciation; a working-capital improvement releases cash once without changing profit. Financial evaluation (NPV) uses cash flows; P&L impact matters for budget owners and targets. Show both when they differ materially.

1.6 Hard ROI, soft ROI and strategic option value

Category	Definition	Example	Treatment in the case
Hard ROI	Benefits that appear as cash or P&L changes on a named landing line, with an owner	Contractor spend removed; fraud losses reduced	In the NPV and payback
Soft ROI	Real improvements not (yet) converted into cash: capacity, satisfaction, quality, speed	Hours freed but not redeployed; higher employee satisfaction	Quantified separately; not in the core NPV unless a conversion mechanism is agreed
Strategic option value	The right, not the obligation, to pursue future opportunities created by this investment	Platform reused by later use cases; data asset; skills	Described qualitatively or valued with a stated method; never silently added to NPV

PITFALL · The stacked total

"Hard €0.4m + soft €1.1m + strategic €2m = €3.5m value" mixes categories with different certainty and different meanings. Finance will discount the whole figure — including the solid part. Present them in **separate columns**.

1.7 Separating efficiency gains from new-value claims

	Efficiency gains	New-value claims
What changes	Cost or time to do existing work	Revenue, new products, new capabilities, market share
Evidence base	Internal baselines, time studies, pilots	Customer and market evidence, experiments, comparable launches
Typical uncertainty	Lower (mainly adoption and the ledger)	Higher (customer behaviour, competition)
Common error	Counting capacity as cash	Assuming demand; ignoring cannibalisation

CONSULTANT'S TIP

Present efficiency and new-value benefits in separate sections with separate evidence. A case that is sound on efficiency should not be weakened by — or hide behind — a speculative new-revenue claim.

1.8 Adjusting pilot results for production conditions

Pilots systematically overstate production results: volunteers are enthusiastic, novelty raises effort, case mix is easier, the vendor helps, and review effort is often unmeasured. Adjust explicitly — and show the adjustment, so finance can challenge each factor.

Adjustment factor	Why	Illustrative factor	Running result
Pilot result	40% handling-time reduction among 15 volunteer agents	—	40.0%

Adjustment factor	Why	Illustrative factor	Running result
User selection	Volunteers more skilled and motivated than average	× 0.85	34.0%
Novelty effect	Extra attention during the pilot fades	× 0.90	30.6%
Case and data mix	Pilot excluded complex and non-German cases	× 0.85	26.0%
Adoption	80% of eligible contacts expected to use it	× 0.80	20.8%
Ledger	Review and correction effort (≈ 3 points) not measured in pilot	− 3 pts	≈ 18%

KEY IDEA

A 40% pilot result became ~18% for planning purposes. The factors are judgments — but **visible** judgments that finance can debate, rather than a hidden optimism that destroys trust when production disappoints.

1.9 Scenarios, sensitivity and capital budgeting under uncertainty

Tool	Question it answers	Good practice
Ranges	What is the plausible spread of each benefit and cost?	Asymmetric: benefits skew lower, costs higher
Scenarios	What happens if several assumptions move together?	Conservative, base and upside — defined by named assumptions
Sensitivity	Which single assumption moves the answer most?	One-way tests (tornado), breakeven values
Staged investment	Can we buy information before committing fully?	Tranches and gates (Course 4); value the option to stop

NOTE

Module 2 builds the cash-flow model, NPV and sensitivities step by step. Here, the structural point is that **uncertainty belongs in the case**, not behind it.

1.10 Assumptions and decision quality

A good decision is one made well with the information available — it can still turn out badly. Decision quality depends heavily on **assumption quality**: explicit, sourced, owned, tested where it matters most.

Assumption register field	Example
Assumption	Agents accept ≥ 60% of drafts with minor edits
Value used / range	60% (45–70%)
Source and evidence level	Pilot, 15 volunteers (moderate; selection bias)
Sensitivity	High — drives 55% of benefits
Owner	Head of Customer Service

Assumption register field	Example
Test before full commitment	Tranche 1 rollout to 60 representative agents, 6 weeks

1.11 Executive summary logic and investment committee expectations

Lead with the answer. A good executive summary follows a **pyramid** logic: recommendation first, then the three or four reasons, each supported by the analysis behind it.

Executive summary element	Example
Recommendation	Approve a staged investment of €0.40m, released in two tranches.
Decision requested	Investment committee approval of Tranche 1 (€0.12m) now; Tranche 2 at Gate 2 (Q3).
Return (range)	NPV €0.1–0.3m over 3 years in the base case; negative if adoption is 30% below plan.
Key assumptions	Draft acceptance ≥ 60%; 9 FTE temp reduction; inference ≤ €0.04/contact.
Main risks and mitigations	Adoption (tested in Tranche 1); review load (staffed); vendor pricing (cap negotiated).
What would change the recommendation	Tranche 1 acceptance below 45% → stop.

Investment committees typically expect…	…and react badly to…
A clear ask and alternatives, including do nothing	Only one option presented
Ranges and the main sensitivities	Single-point estimates to the euro
Named owners for benefits, with budget consequences	Benefits nobody will be held to
Full costs, including run and hidden costs	Licence cost presented as the cost
Evidence levels stated honestly	Vendor claims presented as facts
Staging and stop rules for uncertain investments	All-or-nothing commitments

1.12 Case study: Hollin Water

CASE STUDY **A business case finance could not verify — and its rebuild**

Hollin Water (fictional) · regional water utility · 1.1m customers

The original case

The digital team's case for a GenAI customer-service assistant claimed "€2.6m annual value": €0.9m "agent productivity" (from a 4-week pilot showing 45% faster handling), €0.7m "improved customer satisfaction", and €1.0m "strategic platform value". Costs: licences of €0.25m per year. The CFO returned it unread beyond page one.

What the advisor changed

- **Driver tree:** contacts × net minutes saved × adoption × €/minute — with review and correction time as a negative driver.
- **Pilot adjustment:** 45% pilot result → ≈ 20% planning value (selection, novelty, case mix, adoption, ledger), each factor shown.
- **Traceability:** the only hard landing line was outsourced overflow-call costs (€1.1m/yr budget), owner: Head of Customer Operations, who agreed to cut the 2027 overflow contract by 25–35% if Tranche 1 targets were met.
- **Classification:** hard ROI €0.28–0.39m/yr; soft ROI (in-house capacity, satisfaction) reported separately; platform value described qualitatively as an option for the leak-reporting assistant on the roadmap.
- **Costs:** added inference, human review, knowledge-content upkeep, integration and change management (Module 2) — total recurring cost €0.19m, one-off €0.35m.
- **Decision request:** staged: €0.12m Tranche 1, gate on acceptance rate and measured handling time.

Outcome

The smaller, verifiable case — hard benefit a seventh or less of the original headline — was approved at the next committee in Tranche form.

Discussion questions

1. Why did the CFO reject the €2.6m case?
2. Which part of the original value was hard ROI, and why only that part?
3. Why was the platform value not added to the NPV?
4. What made the rebuilt case more likely to be approved despite a lower number?

Model answers (brief)

1. Stacked categories, an unadjusted pilot, no landing lines or owners, and costs limited to licences — nothing could be verified.
2. Only the reduction in outsourced overflow costs, because it had a named landing line and an owner who committed to change the budget.
3. Option value is uncertain and depends on future decisions. Adding it silently would contaminate the hard figures; it was described as a separate option.

4. Traceable, owned benefits; full costs; visible adjustments; staging with a stop rule. Finance could verify and challenge it.

1.13 Consulting templates

T1.1 Business Case Skeleton

Purpose: Check that every section of a decision-ready case is present.

- [ ] Executive summary: recommendation, ask, return range, key assumptions, risks, what would change it
- [ ] Problem, strategic rationale and alternatives (incl. do nothing and non-AI)
- [ ] Solution, scope, autonomy, roadmap horizon and readiness finding
- [ ] Benefits: driver tree and traceability register
- [ ] Costs: one-off, recurring, hidden, TCO
- [ ] Financial analysis: cash flows, NPV, payback, scenarios, sensitivity
- [ ] Assumption register and risks
- [ ] Measurement and verification plan
- [ ] Decision request: amount, tranches, conditions, stop rules

Decision-holder and date: .....

Finance partner who reviewed the model: .....

T1.2 Value Driver Tree

Purpose: Decompose the financial outcome into measurable drivers.

Outcome (e.g., net annual benefit €): .....

Branch	Driver	Value (range)	Source	Conversion to €	Negative driver?
Agent cost avoided	*Contacts / year*	*420k*	*CRM 2025*	*× net min × €/min*	*No*

T1.3 Benefit Traceability Register

Purpose: Trace every benefit to a landing line and an accountable owner.

Benefit	AI change → operational metric	Conversion mechanism	Landing line	Owner (budget holder)	Type (hard / soft / option)	Timing	Evidence level

T1.4 Pilot-to-Production Adjustment Sheet

Purpose: Adjust pilot results visibly for production conditions.

Factor	Rationale	Factor applied	Running result	Agreed with
Pilot result		—		
User selection				
Novelty effect				
Case / data mix				
Vendor or expert support in pilot				
Adoption at scale				
Ledger (review, correction, escalation)				
Planning value				

T1.5 Assumption Register

Purpose: Make assumptions explicit, sourced, owned and tested.

#	Assumption	Value (range)	Source / evidence level	Sensitivity (H/M/L)	Owner	Test before full commitment

1.14 Best practices and pitfalls

✓ DO THIS

- Write to inform the decision, including the option to decline.
- Build driver trees with negative drivers included.
- Name a landing line and owner for every hard benefit.
- Keep hard, soft and option value in separate columns.
- Separate efficiency gains from new-value claims.
- Show every pilot-to-production adjustment.
- Present ranges, scenarios and the key sensitivity.

× AVOID THIS

- Persuasion documents with hidden assumptions.
- “Productivity” as a benefit line.
- Hours saved counted as cash without a budget change.
- Stacked totals mixing hard, soft and strategic value.
- Speculative revenue propping up an efficiency case.
- Pilot results used unadjusted.
- Single-point figures to the euro.

1.15 Module 1 cheat sheet

MODULE 1 · ONE-PAGE SUMMARY

Structuring the AI Business Case

PURPOSE

Enable approve / stage / decline on transparent evidence finance can verify and challenge.

STRUCTURE

Exec summary · problem & rationale · solution · benefits · costs · financials · assumptions & risks · measurement · decision request.

DRIVER TREE

Outcome → branches → measurable drivers with sources and €-conversion. Include negative drivers.

TRACEABILITY

AI change → operational metric → conversion → **landing line** → **accountable owner** (budget holder).

HARD · SOFT · OPTION

Hard: cash/P&L on a landing line with owner → in NPV. **Soft**: real, not banked → separate. **Option**: right to future opportunities → qualitative or stated method. Never stack.

EFFICIENCY VS NEW VALUE

Existing work cheaper/faster vs revenue/new capability. Different evidence, uncertainty, errors.

PILOT → PRODUCTION

Selection · novelty · case mix · vendor support · adoption · ledger. Visible factors (e.g., 40% → ~18%).

UNCERTAINTY

Asymmetric ranges · conservative/base/upside scenarios · one-way sensitivity · breakeven · staging.

ASSUMPTIONS

Value/range · source · evidence level · sensitivity · owner · test. Decision quality ≠ outcome.

EXECUTIVE SUMMARY & COMMITTEES

Pyramid: recommendation → ask → return range → key assumptions → risks → what would change it. Committees want alternatives, ranges, owners, full costs, honest evidence, staging and stop rules.

1.16 Knowledge check

Answer all questions before turning the page. Explanations for every option follow.

K1 **What is the primary purpose of an AI business case?**

- (A) To document the technical architecture
- (B) To persuade the committee to approve the project
- (C) To enable the decision-holder to approve, stage or decline on transparent evidence that finance can verify
- (D) To satisfy procurement

K2 **What is a benefit's "landing line"?**

- (A) The date the benefit starts
- (B) The line manager of the team using AI
- (C) The final line of the executive summary
- (D) The specific P&L, budget or balance-sheet line where the benefit will appear

K3 **An initiative frees 12,000 agent hours, but no budget or staffing plan will change. How should this be classified?**

- (A) Strategic option value
- (B) Soft ROI, reported separately unless a conversion mechanism and owner are agreed
- (C) Hard ROI at the loaded hourly cost
- (D) It should not be mentioned

K4 **A case totals "hard €0.4m + soft €1.1m + strategic €2m = €3.5m value". What is the main problem?**

- (A) It stacks categories with different certainty and meaning, so finance discounts the whole figure
- (B) Soft value should be larger.
- (C) Strategic value must always be zero.
- (D) The numbers are too large.

K5 **Which claim needs customer and market evidence rather than internal time studies?**

- (A) Reducing rework
- (B) Reducing invoice processing time
- (C) Increasing cross-sell revenue through AI recommendations
- (D) Cutting contractor spend

K6 **A pilot with volunteers showed a 40% time reduction. What is the best planning approach?**

- (A) Halve it without explanation.
- (B) Ignore the pilot.
- (C) Use 40% — it was measured.
- (D) Apply visible adjustments for selection, novelty, case mix, adoption and the ledger

K7 **Which executive summary opening best follows pyramid logic?**

- (A) "We recommend a staged €0.40m investment; approve Tranche 1 (€0.12m) now — base NPV €0.1–0.3m, negative if adoption is 30% below plan."
- (B) A history of the project
- (C) A description of AI technology trends
- (D) A list of all team members

K8 **Which practice most improves decision quality?**

(A) Copying a competitor's figures

(B) Explicit, sourced, owned assumptions with ranges, and tests of the most sensitive ones before full commitment

(C) Single-point estimates for clarity

(D) Removing risks to avoid alarming the committee

1.17 Knowledge check: answers and explanations

K1 **Answer: C** Review: 1.1

What is the primary purpose of an AI business case?

- ✗ (A) To document the technical architecture — **Wrong.** Architecture is not the purpose.
- ✗ (B) To persuade the committee to approve the project — **Wrong.** Persuasion-oriented cases hide assumptions and lose finance's trust.
- ✓ (C) To enable the decision-holder to approve, stage or decline on transparent evidence that finance can verify — **Correct.** The case exists to inform a decision, including a "no".
- ✗ (D) To satisfy procurement — **Wrong.** Procurement is a separate process.

K2 **Answer: D** Review: 1.4

What is a benefit's "landing line"?

- ✗ (A) The date the benefit starts — **Wrong.** That is timing.
- ✗ (B) The line manager of the team using AI — **Wrong.** That is a person, not a line.
- ✗ (C) The final line of the executive summary — **Wrong.** Not the meaning.
- ✓ (D) The specific P&L, budget or balance-sheet line where the benefit will appear — **Correct.** Naming it lets finance see the benefit arrive.

K3 **Answer: B** Review: 1.6

An initiative frees 12,000 agent hours, but no budget or staffing plan will change. How should this be classified?

- ✗ (A) Strategic option value — **Wrong.** Option value concerns future opportunities, not capacity freed now.
- ✓ (B) Soft ROI, reported separately unless a conversion mechanism and owner are agreed — **Correct.** Real capacity, not banked value.
- ✗ (C) Hard ROI at the loaded hourly cost — **Wrong.** Without a landing-line change and an owner, the hours are not cash.
- ✗ (D) It should not be mentioned — **Wrong.** Soft value is real and worth reporting separately.

K4 **Answer: A** Review: 1.6

A case totals "hard €0.4m + soft €1.1m + strategic €2m = €3.5m value". What is the main problem?

- ✓ (A) It stacks categories with different certainty and meaning, so finance discounts the whole figure — **Correct.** Present them in separate columns.
- ✗ (B) Soft value should be larger. — **Wrong.** Irrelevant.
- ✗ (C) Strategic value must always be zero. — **Wrong.** It can be described, just not silently added.
- ✗ (D) The numbers are too large. — **Wrong.** Size is not the issue.

K5 **Answer: C** Review: 1.7

Which claim needs customer and market evidence rather than internal time studies?

- ✗ (A) Reducing rework — **Wrong.** Efficiency.
- ✗ (B) Reducing invoice processing time — **Wrong.** Efficiency: internal baselines suffice.
- ✓ (C) Increasing cross-sell revenue through AI recommendations — **Correct.** A new-value claim depends on customer behaviour.
- ✗ (D) Cutting contractor spend — **Wrong.** Efficiency.

K6 **Answer: D** Review: 1.8

A pilot with volunteers showed a 40% time reduction. What is the best planning approach?

✗ (A) Halve it without explanation. — **Wrong.** An arbitrary, invisible haircut.

✗ (B) Ignore the pilot. — **Wrong.** It is useful evidence once adjusted.

✗ (C) Use 40% — it was measured. — **Wrong.** Pilot conditions overstate production results.

✓ (D) Apply visible adjustments for selection, novelty, case mix, adoption and the ledger — **Correct.** Explicit factors that finance can challenge.

K7 **Answer: A** Review: 1.11

Which executive summary opening best follows pyramid logic?

✓ (A) "We recommend a staged €0.40m investment; approve Tranche 1 (€0.12m) now — base NPV €0.1–0.3m, negative if adoption is 30% below plan." — **Correct.** Recommendation, ask and return range up front.

✗ (B) A history of the project — **Wrong.** The answer comes first.

✗ (C) A description of AI technology trends — **Wrong.** Background, not the answer.

✗ (D) A list of all team members — **Wrong.** Irrelevant.

K8 **Answer: B** Review: 1.10

Which practice most improves decision quality?

✗ (A) Copying a competitor's figures — **Wrong.** Not the client's evidence.

✓ (B) Explicit, sourced, owned assumptions with ranges, and tests of the most sensitive ones before full commitment — **Correct.** Assumption quality drives decision quality.

✗ (C) Single-point estimates for clarity — **Wrong.** These hide uncertainty.

✗ (D) Removing risks to avoid alarming the committee — **Wrong.** This degrades the decision.

MODULE 2

Building the ROI Model

An ROI model finance can trust: full and hidden AI costs, benefit timing, payback, discounted cash flow and NPV, staged investment, sensitivity analysis and benefit realization — and a clear line between value created and effort merely relocated.

YOU WILL BE ABLE TO

- Structure a transparent ROI model: inputs, calculations, outputs, scenarios
- Build total cost of ownership, including the hidden costs of AI adoption
- Maintain a cost register and apply AI spend governance
- Model benefit timing and adoption ramps
- Calculate payback, discounted cash flow, NPV and IRR at advisory depth
- Model staged investment and run sensitivity and breakeven analysis
- Anticipate the ROI assumptions finance will challenge
- Distinguish value created from effort relocated

IN THIS MODULE

1. Model architecture
2. Total cost of ownership
3. Hidden costs of AI adoption
4. Cost register and AI spend governance
5. Benefit timing and ramps
6. Payback, DCF, NPV and IRR
7. Staged investment
8. Sensitivity, scenarios and breakeven
9. Benefit realization
10. Assumptions finance will challenge
11. Created vs relocated value
12. Templates · cheat sheet · knowledge check

Suggested study time · 5 hours

Finance will trust an AI ROI model when they can see every input, follow every calculation and change any assumption. This module builds such a model at advisory depth: all the costs — especially the hidden ones AI creates — realistic benefit timing, the core investment measures (payback, NPV, IRR), staged investment, sensitivity and breakeven, and the discipline to separate value **created** from effort merely **relocated**. A worked example runs through the module (all figures in €k).

KEY TAKEAWAYS

1. Separate **inputs**, **calculations** and **outputs**; every input sourced and owned; scenarios switchable.
2. **TCO** includes one-off, recurring and exit costs — not just licences.
3. Hidden costs: **verification and rework**, **human review**, **change management**, **inference** (usage-based) and **cost transfers** to other teams or customers.
4. Model **adoption ramps**; benefits rarely arrive at full value on day one.
5. Use **NPV** as the core value measure; **payback** for liquidity and risk; **IRR** with care.
6. **Stage** investment to value the option to stop; run **sensitivity and breakeven** to find what really matters.
7. Prepare for the **assumptions finance will challenge**; count only **created** value.

2.1 Model architecture

Layer	Content	Rules
Inputs & assumptions	Volumes, rates, prices, adoption, costs, discount rate, scenario switches	One place only; source, owner, date and range for each; colour-coded as inputs
Calculations	Driver trees, cost build-up, ramps, cash flows	No hard-coded numbers; formulas reference inputs; readable steps
Outputs	NPV, payback, IRR, annual P&L view, cash view, charts	Derived only; never typed over
Scenarios & sensitivity	Conservative / base / upside sets; one-way tests; breakeven	Switchable; documented
Checks & log	Totals reconcile; version log; change log	Finance can audit changes

CONSULTANT'S TIP

Agree the **discount rate, time horizon, tax treatment and capitalisation approach** with finance before building. These are finance policy decisions, not modelling choices — and disagreements about them late in the process waste weeks.

2.2 Total cost of ownership

Cost category	One-off	Recurring	Exit / end of life
Technology	Integration, configuration, environments, security set-up	Licences, hosting, inference / API usage, monitoring tools	Data export, decommissioning

Cost category	One-off	Recurring	Exit / end of life
Data	Cleaning, labelling, pipelines, permissions review	Pipeline upkeep, content maintenance	Archiving, deletion
People & change	Training, process redesign, communications, works-council engagement	Human review, escalation handling, champions, refresher training	Re-training on replacement
Governance & risk	Risk assessment, DPIA, legal review, documentation	Monitoring, audits, re-validation after changes, incident handling	—
Programme	Project team, vendor services, contingency	Product ownership, platform support	Transition project

2.3 The hidden costs of AI adoption

Hidden cost	Why it is missed	How to estimate
Verification and rework	Pilots measure the AI step, not the checking and fixing around it	Edit rates × correction time; downstream error handling
Human review	"Human in the loop" assumed free	Volume × review rate × minutes × loaded cost (Course 4 control-load test)
Change management	Treated as a one-off training session	Training hours × staff, redesign effort, manager time, communications, adoption support
Inference cost	Pilot volumes are tiny; usage grows with adoption and longer prompts	Transactions × tokens (or calls) × unit price, with growth and price-change scenarios
Cost transfers	Effort moves to other teams, customers or partners rather than disappearing	Ledger: who verifies, corrects, escalates or waits (Course 2)
Upkeep	Prompts, connectors, content and evaluation sets need maintenance	Effort per period by item (Course 4, T3.4)

EXAMPLE · Inference cost scales with success

A pilot of 2,000 interactions/month cost €60 in inference (€0.03 each). At full adoption — 300,000 interactions/month, with prompts twice as long after adding retrieval context — the cost becomes 300,000 × €0.03 × 2 ≈ **€18,000/month**, about €216k a year. The cost line most likely to surprise finance is the one that grows *because* the project succeeds.

EXAM TRAP

"The vendor licence is €90k per year, so annual cost is €90k" is a classic wrong answer. Recurring cost includes inference, review, upkeep, monitoring and support — often several times the licence.

2.4 The cost register and AI spend governance

Cost register field	Purpose
Cost item and category	Completeness and classification
One-off / recurring / exit	Correct timing in cash flows

Cost register field	Purpose
Driver and unit cost	Scales correctly with volume (e.g., per transaction)
Range and basis	Honest uncertainty
Budget owner and cost centre	Who pays — and whether costs land on a different team than benefits
Capex / opex treatment (per finance policy)	P&L vs cash view

BEST PRACTICE · AI spend governance

- Tag AI spend (licences, inference, cloud) to initiatives and owners.
- Set usage budgets and alerts for inference and API costs.
- Review unit cost per transaction monthly against the model.
- Charge back or show back usage to consuming units where policy allows.
- Release spend through stage gates (Course 4).

2.5 Benefit timing and adoption ramps

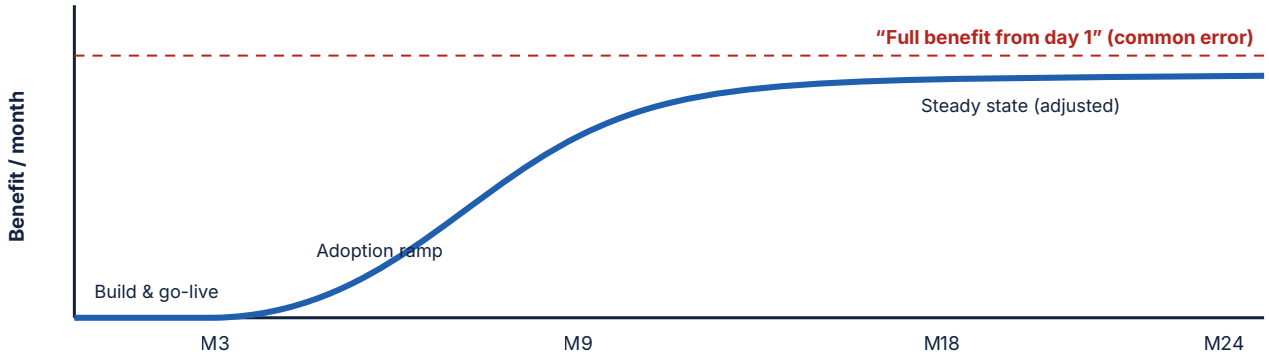


Figure 2.1 — Benefits follow build lag and an adoption ramp

Model benefits month by month or quarter by quarter: a build period with no benefit, an adoption ramp (often S-shaped), then steady state at the **production-adjusted** level (Module 1). Benefits from cost reduction also depend on **when budgets change** — e.g., when an outsourcing contract can be reduced.

2.6 Payback, discounted cash flow, NPV and IRR

Measure	Definition	Use	Watch-out
Simple ROI	(Total net benefit − investment) ÷ investment	Quick communication	Ignores timing; period-dependent
Payback period	Time until cumulative net cash flow turns positive	Liquidity and risk intuition	Ignores value after payback and time value
Discounted payback	Payback using discounted cash flows	More conservative payback	Same limits beyond payback
NPV	Sum of cash flows discounted at the required rate	Core value measure: positive NPV creates value at that rate	Sensitive to rate, horizon and terminal assumptions

Measure	Definition	Use	Watch-out
IRR	Discount rate at which NPV = 0	Comparison with hurdle rate	Misleading for unusual cash-flow patterns and when ranking projects of different size

$$NPV = \Sigma_{t=0..T}\ CF_t \div (1 + r)^t$$ · discount factor at 10%: year 1 = 0.909, year 2 = 0.826, year 3 = 0.751

Worked example: AI invoice processing (€k, discount rate 10%, 3-year horizon)

	Year 0	Year 1	Year 2	Year 3
Investment (build, integration, data, change)	−400			
Gross benefit (ramp: 50% in year 1)		300	600	600
Recurring costs (licences 90, inference 40, human review 60, upkeep & monitoring 50)		−240	−240	−240
Net cash flow	**−400**	**60**	**360**	**360**
Cumulative	−400	−340	+20	+380
Discounted cash flow	-400	54.5	297.5	270.5

Result	Value	Interpretation
NPV @ 10%	€223k	Creates value at the required return — in the base case
Payback	1.9 years	Cumulative cash turns positive late in year 2
Discounted payback	2.2 years	Positive early in year 3
IRR	33%	Above a 10% hurdle
Simple 3-year ROI	95%	(780 − 400) ÷ 400

> **NOTE**
> Figures computed exactly; rounding to €k. Your finance team's discount rate, horizon, tax and terminal-value conventions take precedence. At advisory depth you need to build, read and challenge — not to set finance policy.

2.7 Staged investment

Staging turns an all-or-nothing commitment into a sequence of smaller ones, each with an option to stop. Its value is clearest when success is uncertain.

EXAMPLE · Valuing the option to stop (simplified)

Suppose a 60% chance the invoice project succeeds (NPV €223k) and a 40% chance adoption halves benefits — in which case committing the full €400k up front yields NPV ≈ −€387k.

- **All-in now:** 0.6 × 223 + 0.4 × (−387) ≈ **−€21k** expected NPV.
- **Staged:** an €80k pilot first; continue only if it succeeds (spending the remaining €320k); stop otherwise: 0.6 × 223 + 0.4 × (−80) ≈ **+€102k**.

Simplifications: the pilot is assumed to reveal success or failure reliably and the delay is ignored. The lesson holds: when uncertainty is high, **the right to stop is worth money**.

2.8 Sensitivity, scenarios and breakeven

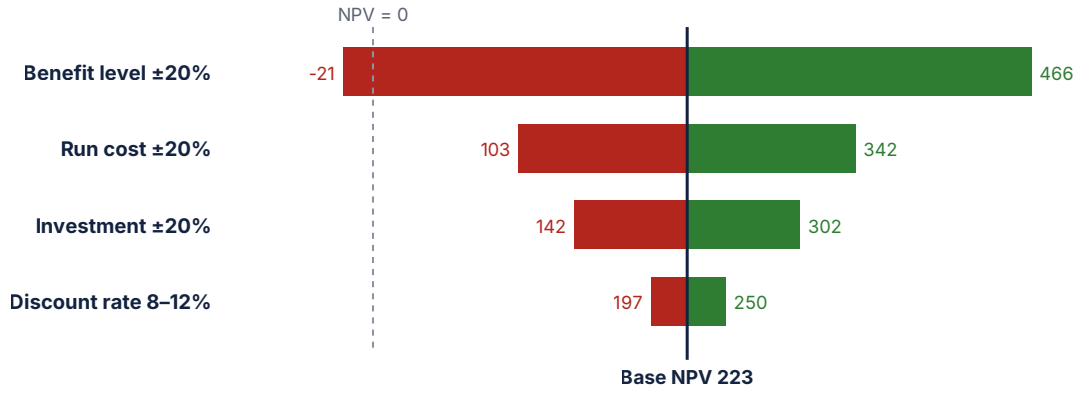


Figure 2.2 — Tornado: one-way sensitivity of NPV (€k) in the worked example

Scenario	Key assumptions	NPV (€k)	Payback
Conservative	Benefits 200 / 450 / 500; recurring 260; investment 450	−167	Not within 3 years
Base	As worked example	223	1.9 yrs
Upside	Benefits 350 / 700 / 750	463	1.6 yrs
Adoption −30%	Benefits × 0.7	−143	Not within 3 years
Inference cost doubles	Recurring 280	123	2.2 yrs

KEY IDEA · Breakeven

In the worked example, NPV reaches zero when benefits are about **82% of plan**. That single number tells the committee more than the base NPV: **if adoption or savings fall short by more than ~18%, the investment destroys value** — so the case should test benefits early (staging) before committing the full amount.

2.9 Benefit realization

Element	Content
Benefit profile	Expected benefit per period per landing line, with owner
Tracking	Actuals vs plan, from finance systems, not only operational dashboards

Element	Content
Budget mechanism	How budgets or targets change when benefits arrive
Review cadence	Quarterly benefits review with finance and owners
Escalation	What happens when benefits fall short — corrective action, re-forecast, stop

2.10 ROI assumptions finance will challenge

Likely challenge	Prepared response
“Where will the saving appear in my budget?”	Landing line, owner, budget change and timing (Module 1)
“Your pilot was with volunteers.”	Pilot-to-production adjustment sheet
“What happens to inference cost at full volume?”	Volume and price scenarios; usage caps; contract terms
“Who reviews the outputs and what does that cost?”	Review cost line from control-load test
“Isn’t this effort just moved to another team?”	Ledger and transfers shown as costs
“Why this discount rate / horizon?”	Finance-agreed policy
“What if adoption is lower?”	Sensitivity, breakeven, staged investment
“What is the cost of not doing it?”	Do-nothing baseline (Module 4)

2.11 Distinguishing value created from effort relocated

The final test before recommending investment: for each benefit, ask whether the effort **disappeared** or **moved**.

Situation	Created or relocated?
AI drafts letters; clerks save 6 minutes; supervisors add 2 minutes of checking	4 minutes created; 2 minutes relocated (cost to supervisors)
Self-service bot deflects calls; customers spend longer and call back	Partly relocated to customers; re-contacts offset savings
Automated extraction; finance team handles more exceptions	Net created only after exception cost
Agent resolves tickets; IT spends time maintaining connectors	Created, minus new upkeep cost

KEY IDEA

Count as benefit only the **net** effort removed from the organization (and its customers), after every transfer. Relocated effort is a cost line, not a saving.

2.12 Consulting templates

T2.1 ROI Model Input Sheet

Purpose: One place for every input, with source, owner and range.

Input	Value	Unit	Low	High	Source / evidence level	Owner	Date
Volume / year							
Net time saved per unit							
Adoption at steady state							
Ramp profile							
Loaded cost / hour							
Inference cost / transaction							
Review rate & minutes							
Discount rate (finance)							
Horizon (finance)							

T2.2 Cost Register & TCO

Purpose: Capture every cost, its driver and who pays.

Cost item	Category	One-off / recurring / exit	Driver × unit cost	Range	Budget owner / cost centre	Capex / opex

T2.3 Hidden Cost Checklist

Purpose: Make sure the costs AI adoption creates are in the model.

- [ ] Verification and rework (edit rates × correction time; downstream errors)
- [ ] Human review (volume × rate × minutes × loaded cost) at each phase's volume
- [ ] Change management (training, redesign, manager time, communications, adoption support)
- [ ] Inference / API cost at full volume, with growth and price scenarios
- [ ] Cost transfers to other teams, customers or partners (ledger)
- [ ] Upkeep: prompts, connectors, content, evaluation sets, permissions
- [ ] Governance: monitoring, audits, re-validation, incident handling
- [ ] Exit costs: data export, transition, decommissioning

Items found missing and added (with values): .....

T2.4 Cash Flow, NPV & Payback Worksheet

Purpose: Compute the core measures transparently.

Line	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5
Investment						
Gross benefit (hard only)						
Recurring costs						
Net cash flow						
Cumulative						
Discount factor	*1.000*					
Discounted cash flow						

NPV: Payback: Discounted payback: IRR: .....

T2.5 Sensitivity, Scenario & Breakeven Sheet

Purpose: Show what really drives the answer.

ONE-WAY SENSITIVITY

Assumption	Low value	NPV at low	High value	NPV at high	Rank

SCENARIOS

Scenario	Named assumptions	NPV	Payback
Conservative			
Base			
Upside			

Breakeven: benefits must reach ____% of plan (or adoption ____%) for NPV = 0

T2.6 Finance Challenge Preparation

Purpose: Prepare answers before finance asks.

Expected challenge	Prepared answer	Evidence / model reference	Owner
Where does the saving land?			
Pilot conditions vs production?			
Inference cost at full volume?			
Who reviews outputs, at what cost?			
Is effort just relocated?			
What if adoption is lower?			
Cost of not doing it?			

2.13 Best practices and pitfalls

✓ DO THIS

- Separate inputs, calculations and outputs.
- Agree discount rate, horizon and capitalisation with finance first.
- Model TCO including hidden and exit costs.
- Scale inference cost with volume and prompt length.
- Model ramps from production-adjusted values.
- Show tornado, scenarios and breakeven.
- Count only net created value; book transfers as costs.

× AVOID THIS

- Hard-coded numbers inside formulas.
- Licence cost presented as total cost.
- Assuming human review is free.
- Full benefits from go-live.
- IRR as the only measure.
- A single base case with no downside.
- Counting relocated effort as savings.

2.14 Module 2 cheat sheet

MODULE 2 · ONE-PAGE SUMMARY

Building the ROI Model

ARCHITECTURE

Inputs (sourced, owned, ranged) → calculations (no hard-codes) → outputs → scenarios → checks & log. Finance sets rate, horizon, tax, capitalisation.

TCO

One-off · recurring · exit across technology, data, people & change, governance, programme.

HIDDEN COSTS

Verification & rework · human review · change management · inference (scales with success) · cost transfers · upkeep · governance · exit.

MEASURES

NPV = Σ CF/$(1+r)^t$ (core) · payback (liquidity) · discounted payback · IRR (hurdle; careful) · simple ROI (communication).

WORKED EXAMPLE

−400; +60; +360; +360 @10% → NPV ≈ €223k · payback 1.9y · IRR ≈ 33% · breakeven at ~82% of benefits.

STAGING

Option to stop is worth money when success is uncertain (example: all-in ≈ −21 vs staged ≈ +102 expected NPV).

SENSITIVITY

Tornado (one-way) · scenarios (named assumptions) · breakeven. Benefit level usually dominates.

REALIZATION & GOVERNANCE

Benefit profile by landing line · tracked in finance systems · budget mechanism · quarterly review · tagged AI spend, usage caps, unit-cost reviews.

CREATED VS RELOCATED

Net effort removed after every transfer = benefit. Moved effort = cost line.

2.15 Knowledge check

Answer all questions before turning the page.

K1 **Which cost line is most likely to grow because the AI project succeeds?**

- (A) Legal review at approval
- (B) Initial training
- (C) One-off integration
- (D) Inference / API usage cost

K2 **A case lists annual cost as the €90k vendor licence only. What is missing?**

- (A) Nothing, if the vendor is reputable
- (B) Recurring costs such as inference, human review, upkeep, monitoring and support — and one-off and exit costs
- (C) Only the VAT
- (D) Only depreciation

K3 **Net cash flows are −400, +60, +360, +360 (€k) at 10%. Approximately what is the NPV?**

- (A) €223k
- (B) €95k
- (C) −€40k
- (D) €380k

K4 **What does a payback period of 1.9 years NOT tell you?**

- (A) Liquidity exposure
- (B) Whether money is recovered within the horizon
- (C) The value created after payback and the time value of money
- (D) When cumulative cash turns positive

K5 **NPV turns to zero when benefits are 82% of plan. What should the case do with this insight?**

- (A) Lower the discount rate
- (B) Ignore it — the base case is positive.
- (C) Raise the benefit estimate
- (D) Highlight it and stage the investment so benefit levels are tested before full commitment

K6 **Why can staging increase expected NPV when success is uncertain?**

- (A) It guarantees success.
- (B) It creates an option to stop after a smaller spend if early evidence is poor
- (C) It removes the need for a pilot.
- (D) It lowers the discount rate.

K7 **A bot deflects calls, but customers spend longer online and many call back later. How should the model treat this?**

- (A) Count net savings after re-contacts, and show customer effort as a transfer
- (B) Ignore customers — only internal costs matter
- (C) Double the savings for experience benefits
- (D) Count all deflected calls as savings

K8 **Which item is a finance policy decision rather than a modelling choice?**

(A) Whether to show a tornado chart
(B) Colour-coding inputs
(C) The discount rate, horizon, tax treatment and capitalisation approach
(D) Ramp curve shape

2.16 Knowledge check: answers and explanations

K1 **Answer: D** Review: 2.3

Which cost line is most likely to grow because the AI project succeeds?

- ✗ A — Legal review at approval — **Wrong.** One-off.
- ✗ B — Initial training — **Wrong.** Mostly one-off.
- ✗ C — One-off integration — **Wrong.** One-off costs do not scale with use.
- ✓ D — Inference / API usage cost — **Correct.** Usage-based cost scales with adoption, volume and prompt length.

K2 **Answer: B** Review: 2.2–2.3

A case lists annual cost as the €90k vendor licence only. What is missing?

- ✗ A — Nothing, if the vendor is reputable — **Wrong.** Licences are only part of recurring cost.
- ✓ B — Recurring costs such as inference, human review, upkeep, monitoring and support — and one-off and exit costs — **Correct.** TCO covers all of them.
- ✗ C — Only the VAT — **Wrong.** Far more is missing.
- ✗ D — Only depreciation — **Wrong.** Depreciation is an accounting view, not the missing costs.

K3 **Answer: A** Review: 2.6

Net cash flows are −400, +60, +360, +360 (€k) at 10%. Approximately what is the NPV?

- ✓ A — €223k — **Correct.** −400 + 54.5 + 297.5 + 270.5 ≈ 222.5.
- ✗ B — €95k — **Wrong.** Not correct.
- ✗ C — −€40k — **Wrong.** Not correct.
- ✗ D — €380k — **Wrong.** That is the undiscounted cumulative total.

K4 **Answer: C** Review: 2.6

What does a payback period of 1.9 years NOT tell you?

- ✗ A — Liquidity exposure — **Wrong.** Payback does inform this.
- ✗ B — Whether money is recovered within the horizon — **Wrong.** Payback does show this.
- ✓ C — The value created after payback and the time value of money — **Correct.** Payback ignores both.
- ✗ D — When cumulative cash turns positive — **Wrong.** That is exactly what payback tells you.

K5 **Answer: D** Review: 2.8

NPV turns to zero when benefits are 82% of plan. What should the case do with this insight?

- ✗ A — Lower the discount rate — **Wrong.** The rate is finance policy, not a lever.
- ✗ B — Ignore it — the base case is positive. — **Wrong.** Breakeven shows how little margin exists.
- ✗ C — Raise the benefit estimate — **Wrong.** Manipulates the model.
- ✓ D — Highlight it and stage the investment so benefit levels are tested before full commitment — **Correct.** Breakeven tells the committee what matters; staging buys information.

K6 **Answer: B** Review: 2.7

Why can staging increase expected NPV when success is uncertain?

✗ (A) It guarantees success. — **Wrong.** It does not.

✓ (B) It creates an option to stop after a smaller spend if early evidence is poor — **Correct.** Stopping limits losses in failure states.

✗ (C) It removes the need for a pilot. — **Wrong.** Staging usually begins with a pilot.

✗ (D) It lowers the discount rate. — **Wrong.** It does not.

K7 **Answer: A** Review: 2.11

A bot deflects calls, but customers spend longer online and many call back later. How should the model treat this?

✓ (A) Count net savings after re-contacts, and show customer effort as a transfer — **Correct.** Relocated effort is not created value.

✗ (B) Ignore customers — only internal costs matter — **Wrong.** Transfers to customers matter for value and experience.

✗ (C) Double the savings for experience benefits — **Wrong.** Unsupported.

✗ (D) Count all deflected calls as savings — **Wrong.** This ignores re-contacts and customer effort.

K8 **Answer: C** Review: 2.1

Which item is a finance policy decision rather than a modelling choice?

✗ (A) Whether to show a tornado chart — **Wrong.** A modelling choice.

✗ (B) Colour-coding inputs — **Wrong.** A modelling convention.

✓ (C) The discount rate, horizon, tax treatment and capitalisation approach — **Correct.** Agree these with finance first.

✗ (D) Ramp curve shape — **Wrong.** A modelling assumption, though it needs evidence.

MODULE 3

Defining Value Metrics

Making AI ROI measurable and accountable: separating adoption from business value, linking to indicators budget owners already review, setting baselines and Y1/Y2 targets, pairing speed with quality, choosing comparison designs and claiming only what the attribution level supports.

YOU WILL BE ABLE TO

- → Separate adoption and usage metrics from business value metrics
- → Connect AI success metrics to indicators budget owners already review
- → Distinguish hard-dollar benefits from softer productivity gains in metrics
- → Set pre-intervention baselines, benchmarks and first- and second-year targets
- → Pair productivity metrics with quality metrics
- → Choose comparison designs and attribution levels that support defensible claims
- → Use leading and lagging indicators as adoption matures, and proxies without overstating value
- → Build an AI measurement plan that avoids vanity metrics

IN THIS MODULE

1. The metric hierarchy
2. Adoption vs value; vanity metrics
3. Linking to indicators owners already review
4. Hard-dollar vs productivity metrics
5. Baselines and benchmarks
6. Paired quality metrics
7. Comparison designs
8. Attribution levels and claim strength
9. Leading and lagging indicators over time
10. Proxy measures
11. Year-1 and year-2 targets
12. The measurement plan
13. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4 hours

If value cannot be measured, it cannot be claimed — and finance will not fund the next phase. This module teaches you to define AI value metrics that are **accountable** (linked to indicators budget owners already review), **defensible** (baselines, paired quality metrics, comparison designs and stated attribution levels) and **honest** (no vanity metrics, no proxies dressed up as value).

KEY TAKEAWAYS

1. Use a **metric hierarchy**: adoption → usage quality → operational outcome → financial value. Only the top proves ROI.
2. **Adoption metrics** show use, not value; on their own they are **vanity metrics**.
3. Link AI metrics to **indicators budget owners already review**, so value shows up where decisions are made.
4. Measure against a **pre-intervention baseline**; **pair** every speed metric with a **quality** metric.
5. Choose a **comparison design** and state the **attribution level**; claim no more than it supports.
6. **Leading indicators** dominate early; **lagging** value indicators take over as adoption matures.
7. Set **year-1 and year-2 targets** and a **measurement plan** before go-live.

3.1 The metric hierarchy

Figure 3.1 — From adoption to financial value

Each level is necessary but not sufficient for the next. High adoption with poor usage quality (rubber-stamping, heavy edits) does not move outcomes; better outcomes that are not converted into a budget change do not produce financial value.

3.2 Adoption and usage vs business value — and vanity metrics

Metric	Type	What it shows	What it does not show
Active users per week	Adoption	People are using the tool	Whether use helps
Prompts / sessions	Activity	Volume of interaction	Value — may even indicate struggle
Draft acceptance rate	Usage quality	Outputs are good enough to use	Whether time or cost fell

Metric	Type	What it shows	What it does not show
Handling time per case	Operational outcome	Effort changed	Whether cost fell (capacity ≠ cash)
Temp-staff cost (CC 4410)	Financial value	Money saved on a landing line	—

DEFINITION · Vanity metric

A metric that looks impressive and moves in the right direction but does not connect to a decision or to value — e.g., "1.2 million prompts this quarter". It becomes harmful when presented as evidence of ROI.

EXAM TRAP

"Usage doubled, so ROI has doubled" is wrong. Adoption is a precondition for value, not a measure of it.

3.3 Connecting to indicators budget owners already review

Budget owners run their units on a small set of indicators: cost per contact, backlog, days sales outstanding, first-time-fix rate, write-offs. If the AI metrics are different, AI value lives in a separate dashboard nobody uses for decisions. Map each AI value metric to an existing management indicator — and report through the existing report wherever possible.

AI metric	Existing indicator the owner reviews	Report
Net handling-time reduction	Cost per contact; contacts per FTE	Monthly contact-centre performance pack
Extraction straight-through rate	Invoices processed per FTE; cost per invoice	Finance operations KPI report
Fraud model precision and recall	Fraud losses; false-positive investigations	Monthly risk report
Maintenance prediction lead time	Unplanned downtime hours; maintenance cost	Plant operations review

3.4 Hard-dollar benefits vs softer productivity gains

	Hard-dollar metric	Productivity metric
Measures	Money on a landing line	Time, throughput, capacity
Example	Outsourcing invoice reduced by €28k/month	Cases per agent-hour +18%
Source	Finance systems	Operational systems, time studies
Becomes hard when…	—	Capacity is removed or redeployed to valued work, by decision of the owner

CONSULTANT'S TIP

Report both — but in separate lines, and track the **conversion** from productivity to hard dollars explicitly (e.g., "18% productivity → 9 FTE reduction in temps → €310k"). That conversion is where value most often leaks.

3.5 Baselines and benchmarks

Baseline rule	Why
Measure before the intervention	Afterwards, memory and incentives distort the “before”
Use a long enough period	Captures normal variation; typically 8–12 weeks or more
Account for seasonality	Compare like with like (same months, or adjust)
Fix the definition	Same metric definition, data source and population before and after
Record the context	Staffing, volumes, process or policy changes in the period
Store it	Baseline record signed off by the metric owner

PITFALL · Benchmark borrowing

External benchmarks (“industry average handling time is 6 minutes”) are useful for context, not as a baseline. Your own pre-intervention measurement is what value is calculated against.

3.6 Paired quality metrics

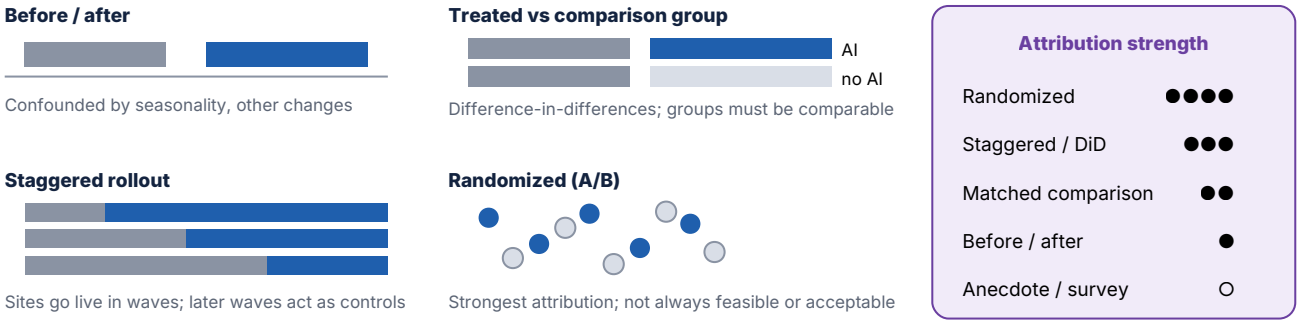
Every productivity or speed metric can be improved by lowering quality. Pair each with a quality metric that would reveal it.

Productivity metric	Paired quality metric
Handling time per contact	Re-contact rate within 7 days; complaint rate
Chatbot deflection rate	Re-contact via other channels; customer satisfaction on deflected contacts
Memos drafted per analyst	Error rate found in sample review; committee rework
Invoices processed straight-through	Payment errors; supplier queries
Tickets auto-resolved	Reopen rate; user satisfaction

KEY IDEA

A productivity gain counts only if its paired quality metric holds or improves. Report the pair together, always.

3.7 Comparison designs



Choose the strongest design the context allows — and make claims no stronger than the design supports.

Figure 3.2 — Comparison designs and attribution strength

Design	How	When to use	Main weakness
Before / after	Compare baseline with post-launch period	When nothing else is possible	Other changes and seasonality confound
Matched comparison	Compare users with similar non-users	Voluntary adoption	Self-selection: adopters differ
Difference-in-differences	Compare change in treated group vs change in comparison group	Units or sites comparable at baseline	Groups must follow similar trends
Staggered rollout	Units go live in waves; not-yet-live units are controls	Phased rollouts (natural fit with Course 4)	Waves must be comparable
Randomized (A/B)	Random assignment of cases, users or teams	When feasible and acceptable	Operational and ethical constraints

3.8 Attribution levels and claim strength

Attribution level	Evidence	Defensible claim wording
A0 · Anecdote	Testimonials, surveys of perceived time saved	"Users report…"
A1 · Before / after	Baseline vs post-launch, same definitions	"Since launch, X changed by…; other factors may contribute."
A2 · Comparison	Matched or difference-in-differences	"Compared with similar teams, X improved by … more."
A3 · Staggered / quasi-experimental	Waves with comparable controls	"The rollout is associated with … , consistent across waves."
A4 · Randomized	A/B test	"AI caused an improvement of … (± range)."

EXAM TRAP

Claiming causation ("AI saved €1.2m") from a before/after comparison during a period when volumes fell is an overstatement. Match the wording to the attribution level.

3.9 Leading and lagging indicators as adoption matures

Stage	Dominant indicators	Example	Decision they support
Launch (months 0–3)	Leading: adoption, usage quality	Active users; acceptance and edit rates	Fix adoption barriers; check oversight
Ramp (3–9)	Leading + early outcomes	Handling time; paired quality	Expand or adjust
Steady state (9–24)	Lagging: financial value	Cost per contact; temp-staff budget	Scale gate; next tranche; benefit realization

3.10 Proxy measures — without overstating value

When the real value metric arrives late or is hard to measure, a **proxy** stands in: “time saved per draft” as a proxy for cost reduction; “early warning lead time” as a proxy for avoided downtime. Proxies are legitimate if their limits are stated.

Rule for proxies	Example
Name it as a proxy	“Proxy for: contact-centre cost”
State the link and its uncertainty	“Converts to cost only if temp hours are reduced”
Plan when the real metric replaces it	“Replaced by budget actuals from month 9”
Never convert proxies to € without the conversion mechanism	No “€ value of time saved” unless budgets change

3.11 Setting year-1 and year-2 targets

Metric	Baseline	Year-1 target	Year-2 target	Owner
Active use (eligible agents)	0%	≥ 75%	≥ 85%	CS operations lead
Net handling time	7.4 min	−10% (6.7)	−18% (6.1)	Head of Customer Service
Re-contact within 7 days (paired)	11.2%	≤ 11.2%	≤ 10.5%	Head of Customer Service
Temp-staff cost CC 4410	€1.10m	−€0.15m	−€0.33m	Head of Customer Service / finance

BEST PRACTICE

Year-1 targets usually emphasise adoption and operational outcomes with partial financial value (ramp); year-2 targets carry the steady-state financial value. Align targets with the ROI model’s ramp (Module 2) so the two never disagree.

3.12 The AI measurement plan

Plan element	Content
Metric hierarchy	Metrics at each level with definitions
Baselines	Values, periods, sources, sign-off
Paired metrics	Quality pair for each productivity metric
Comparison design and attribution level	Design chosen; claim wording agreed
Targets	Year-1 and year-2, aligned with the ROI model
Reporting	Existing management reports used; cadence; audience
Ownership	Metric owners; finance validation of financial metrics
Decision links	Which gate or tranche each metric informs

3.13 Case study: Arden Bank

CASE STUDY **"Our copilot saved 140,000 hours" — or did it?**

Arden Bank (fictional) · retail bank · 4,500 staff

Situation

Six months after rolling out a copilot to 3,000 staff, the programme reports "140,000 hours saved, worth €7m". The figure comes from a monthly pop-up asking users how much time the tool saved them that week. The CFO asks why operating costs have not moved.

What the advisor found

- The headline was **A0** (self-reported, no baseline) converted to euros without a conversion mechanism.
- Adoption was high (78% weekly active) but usage quality was unmeasured.
- For one function — mortgage operations — a usable baseline existed: case-processing time and rework rate from the workflow system.

The redesigned measurement

- **Hierarchy:** adoption → usage quality (acceptance, edits) → operational outcomes (processing time, paired with rework rate) → financial value (overtime and contractor cost in mortgage ops).
- **Design:** the next three regional mortgage teams went live in staggered waves (A3), with not-yet-live teams as comparisons.
- **Result after 4 months:** processing time −11% relative to comparison teams; rework unchanged (pair held); contractor cost reduced by €0.36m/yr after the owner cut two contracts.
- **Reporting:** metrics added to the existing mortgage-operations performance pack.

Discussion questions

1. What attribution level did the original claim have, and why did costs not move?
2. Why was rework chosen as the paired metric?
3. How did the staggered rollout improve attribution?
4. Write a defensible one-sentence claim from the redesigned results.

Model answers (brief)

1. A0. Hours were self-reported and never converted into budget changes, so they stayed as soft capacity.
2. Faster processing could come from cutting corners; unchanged rework shows quality held.
3. Not-yet-live teams served as comparisons, so seasonal and market effects were separated from the copilot's effect.
4. "In staggered rollout across three regions, mortgage processing time fell 11% relative to comparison teams with no rise in rework, enabling a €0.36m annual contractor reduction."

3.14 Consulting templates

T3.1 Metric Hierarchy Map

Purpose: Define metrics at every level and link them to existing indicators.

Level	Metric	Definition & source	Existing indicator / report it links to	Owner
1 Adoption & activity				
2 Usage quality				
3 Operational outcome				
3 Paired quality metric				
4 Financial value				

T3.2 Baseline Record

Purpose: Fix the pre-intervention baseline before go-live. **Use when:** before any go-live

Metric and exact definition: .....

Data source and population: .....

Baseline period (dates) and length: .....

Seasonality treatment: .....

Baseline value and variation (range): .....

Context during period (volumes, staffing, changes): .....

Signed off by metric owner / date: .....

T3.3 Paired Metric & Comparison Design Sheet

Purpose: Pair productivity with quality and choose the comparison design.

PAIRS

Productivity metric	Paired quality metric	Threshold for "quality held"

DESIGN

- ☐ Before / after (A1) — confounders listed
- ☐ Matched comparison (A2) — matching criteria stated
- ☐ Difference-in-differences (A2–A3) — comparable trends checked
- ☐ Staggered rollout (A3) — wave plan aligned with roadmap
- ☐ Randomized (A4) — feasibility and acceptability confirmed

Agreed claim wording for this attribution level: .....

T3.4 Measurement Plan with Year-1 / Year-2 Targets

Purpose: One plan linking metrics, targets, reporting and decisions.

Metric	Level	Baseline	Y1 target	Y2 target	Leading / lagging	Proxy? (for what, until when)	Report & cadence	Decision it informs

3.15 Best practices and pitfalls

✓ DO THIS

- Measure at all four levels; claim ROI only from the top.
- Report through indicators owners already review.
- Baseline before go-live, with fixed definitions.
- Pair every productivity metric with a quality metric.
- Use staggered rollouts as natural comparison designs.
- Word claims to match the attribution level.
- Align targets with the ROI model's ramp.

× AVOID THIS

- Usage counts as evidence of ROI.
- Separate AI dashboards nobody uses for decisions.
- Reconstructing baselines after launch.
- Speed gains without quality checks.
- Self-reported time savings converted to euros.
- Causal claims from before/after data.
- Proxies presented as financial value.

3.16 Module 3 cheat sheet

MODULE 3 · ONE-PAGE SUMMARY

Defining Value Metrics

HIERARCHY

1 adoption & activity → 2 usage quality → 3 operational outcome → 4 financial value. Only 4 proves ROI.

VANITY METRICS

Look good, link to no decision or value (prompts, sessions). Adoption ≠ value.

OWNERS' INDICATORS

Map AI metrics to KPIs budget owners already review; report in existing packs.

HARD VS PRODUCTIVITY

€ on landing line vs time/throughput. Track the conversion explicitly.

BASELINES

Before go-live · long enough · seasonality · fixed definition · context · signed off. Benchmarks = context only.

PAIRED QUALITY

Every speed metric with a quality metric (re-contact, errors, reopen, complaints). Gain counts only if quality holds.

DESIGNS & ATTRIBUTION

Before/after (A1) · matched / DiD (A2) · staggered (A3) · randomized (A4); anecdote A0. Claim wording must match the level.

LEADING → LAGGING

Launch: adoption, usage quality · ramp: outcomes · steady state: financial value.

PROXIES & TARGETS

Name proxy, state link, plan replacement, no € without conversion. Y1 (adoption, outcomes, partial €) and Y2 (steady-state €), aligned with ROI ramp.

3.17 Knowledge check

Answer all questions before turning the page.

K1 **Weekly active users of a copilot doubled. What can you conclude about ROI?**

- (A) Nothing yet — value needs usage-quality, outcome and financial metrics
- (B) ROI doubled
- (C) ROI fell
- (D) The project should scale

K2 **Why link AI metrics to indicators budget owners already review?**

- (A) Because AI metrics are unreliable
- (B) So AI value appears where decisions and budgets are managed, making it accountable
- (C) To hide AI costs
- (D) To reduce reporting effort only

K3 **Which baseline practice is correct?**

- (A) Use the first week after launch
- (B) Reconstruct the baseline from memory after launch
- (C) Measure before go-live over a sufficient period with fixed definitions and a sign-off
- (D) Use an industry benchmark as the baseline

K4 **A chatbot raises deflection from 20% to 45%. Which paired metric is most important?**

- (A) Number of intents configured
- (B) Chatbot response speed
- (C) Number of chatbot sessions
- (D) Re-contact through other channels for deflected customers

K5 **Sites go live in three waves, and not-yet-live sites are used as comparisons. Which design and attribution level is this?**

- (A) Before / after, A1
- (B) Randomized, A4
- (C) Anecdote, A0
- (D) Staggered rollout, about A3

K6 **Which claim fits a before/after (A1) comparison during a period when volumes also fell?**

- (A) “Since launch, cost per contact fell 9%; lower volumes and other factors may contribute.”
- (B) “AI is transformational.”
- (C) “AI caused €1.2m savings.”
- (D) “No conclusions possible.”

K7 **“Minutes saved per draft” is used as a proxy for cost reduction. Which practice is correct?**

- (A) Report it as hard value
- (B) Convert minutes to euros at the salary rate
- (C) Label it a proxy, state the conversion condition and plan when budget actuals replace it
- (D) Drop it entirely

K8 **Which target set is best aligned with an adoption ramp?**

(A) Only usage targets

(B) Year 1: adoption, operational outcomes and partial financial value; year 2: steady-state financial value, aligned with the ROI model

(C) No targets until year 2

(D) Full financial value in year 1

3.18 Knowledge check: answers and explanations

K1 **Answer: A** Review: 3.1–3.2

Weekly active users of a copilot doubled. What can you conclude about ROI?

✓ **A** Nothing yet — value needs usage-quality, outcome and financial metrics — **Correct.** Only the top of the hierarchy proves ROI.

✗ B ROI doubled — **Wrong.** Adoption is a precondition, not a measure of value.

✗ C ROI fell — **Wrong.** There is no basis for that.

✗ D The project should scale — **Wrong.** Scale decisions need value evidence.

K2 **Answer: B** Review: 3.3

Why link AI metrics to indicators budget owners already review?

✗ A Because AI metrics are unreliable — **Wrong.** Not the reason.

✓ **B** So AI value appears where decisions and budgets are managed, making it accountable — **Correct.** Separate AI dashboards rarely drive decisions.

✗ C To hide AI costs — **Wrong.** No.

✗ D To reduce reporting effort only — **Wrong.** Effort is a side benefit.

K3 **Answer: C** Review: 3.5

Which baseline practice is correct?

✗ A Use the first week after launch — **Wrong.** That is already treated.

✗ B Reconstruct the baseline from memory after launch — **Wrong.** Memory and incentives distort it.

✓ **C** Measure before go-live over a sufficient period with fixed definitions and a sign-off — **Correct.** This is a defensible baseline.

✗ D Use an industry benchmark as the baseline — **Wrong.** Benchmarks are context, not baselines.

K4 **Answer: D** Review: 3.6

A chatbot raises deflection from 20% to 45%. Which paired metric is most important?

✗ A Number of intents configured — **Wrong.** Build metric.

✗ B Chatbot response speed — **Wrong.** Not a quality pair for deflection.

✗ C Number of chatbot sessions — **Wrong.** Activity, not quality.

✓ **D** Re-contact through other channels for deflected customers — **Correct.** It reveals whether deflection really solved the problem.

K5 **Answer: D** Review: 3.7–3.8

Sites go live in three waves, and not-yet-live sites are used as comparisons. Which design and attribution level is this?

✗ A Before / after, A1 — **Wrong.** Comparison groups are used.

✗ B Randomized, A4 — **Wrong.** Assignment is not random.

✗ C Anecdote, A0 — **Wrong.** Measured data is used.

✓ **D** Staggered rollout, about A3 — **Correct.** Waves with comparable controls.

K6 **Answer: A** Review: 3.8

Which claim fits a before/after (A1) comparison during a period when volumes also fell?

✓ **A** "Since launch, cost per contact fell 9%; lower volumes and other factors may contribute." — **Correct.** Wording matches the attribution level.

✗ B "AI is transformational." — **Wrong.** Not a claim about evidence.

✗ C "AI caused €1.2m savings." — **Wrong.** A causal claim overstates A1 evidence.

✗ D "No conclusions possible." — **Wrong.** A1 supports a qualified claim.

K7 **Answer: C** Review: 3.10

"Minutes saved per draft" is used as a proxy for cost reduction. Which practice is correct?

✗ A Report it as hard value — **Wrong.** Overstatement.

✗ B Convert minutes to euros at the salary rate — **Wrong.** Not without a conversion mechanism such as a budget change.

✓ **C** Label it a proxy, state the conversion condition and plan when budget actuals replace it — **Correct.** Proxies are legitimate with stated limits.

✗ D Drop it entirely — **Wrong.** Proxies are useful early.

K8 **Answer: B** Review: 3.11

Which target set is best aligned with an adoption ramp?

✗ A Only usage targets — **Wrong.** No value accountability.

✓ **B** Year 1: adoption, operational outcomes and partial financial value; year 2: steady-state financial value, aligned with the ROI model — **Correct.** Targets and model agree.

✗ C No targets until year 2 — **Wrong.** Year-1 targets drive early decisions.

✗ D Full financial value in year 1 — **Wrong.** This ignores the ramp.

MODULE 4

Making the Decision-Ready Case

Turning analysis into a CFO-ready business case, investment committee memo or board pre-read: return, timeline, assumptions, dependencies and verification findable in a minute — conservative, evidence-linked, aligned with roadmap and readiness, and able to say “no”.

YOU WILL BE ABLE TO

- Explain what makes a business case decision-ready
- Shape a CFO-ready case, investment committee memo and board pre-read
- Separate strategic rationale from quantitative analysis
- Connect every financial claim to evidence of stated maturity
- Use conservative scenarios and quantify the cost and risk of inaction
- Align the case with the roadmap and readiness findings
- Write clear decision requests
- Preserve and present a credible null recommendation

IN THIS MODULE

1. What decision-ready means
2. Formats: CFO case, IC memo, board pre-read
3. The 60-second test
4. Strategic rationale vs quantitative analysis
5. Connecting claims to evidence
6. Conservative scenarios
7. Cost and risk of inaction
8. Roadmap and readiness alignment
9. Clear decision requests
10. The credible null recommendation
11. Case study · templates · cheat sheet · knowledge check

Suggested study time · 4 hours

All the analysis of Modules 1–3 is wasted if executives cannot find the return, the timeline, the assumptions, the dependencies and the verification plan quickly — or if they cannot trust them. This module turns analysis into a **decision-ready** document: answer first, strategy and numbers kept separate, every claim tied to evidence, conservative by default, consistent with the roadmap and readiness findings, ending in a precise decision request. And it shows how to keep a **credible "no"** when the numbers do not support proceeding.

KEY TAKEAWAYS

1. Decision-ready = the decision-holder can **find, trust and act on** the return, timeline, assumptions, dependencies and verification plan.
2. Match the format to the forum: **CFO-ready case**, **investment committee memo**, **board pre-read**.
3. Pass the **60-second test**: the key facts are locatable on page one.
4. Keep **strategic rationale** and **quantitative analysis** separate — strategy must not inflate the numbers.
5. Tag every claim with its **evidence maturity**; lead with **conservative** scenarios.
6. Quantify the **cost and risk of inaction** honestly.
7. Align with the **roadmap horizon** and **readiness finding**; end with a **clear decision request**.
8. Preserve a **credible null recommendation** — "do not invest (now)" — when the numbers say so.

4.1 What decision-ready means

Decision-ready case…	Not decision-ready…
States the recommendation and the decision requested first	Builds up to a conclusion on page 12
Shows the return as a range with the main sensitivity	Shows one optimistic number
Makes assumptions, dependencies and evidence levels explicit	Buries them in appendices or omits them
Aligns with the roadmap and readiness findings	Contradicts them
Includes a verification plan and stop rules	Assumes success
Can recommend "no" or "not now"	Is written to secure approval

4.2 Formats: CFO-ready case, investment committee memo, board pre-read

Format	Audience and purpose	Length	Emphasis
CFO-ready business case	Finance leadership; verify and challenge before formal approval	8–15 pages + model	Model transparency, landing lines, costs, sensitivities
Investment committee memo	Formal approval forum; decide on a specific request	2–4 pages	Decision request, return range, risks, conditions, evidence levels

Format	Audience and purpose	Length	Emphasis
Board pre-read	Board oversight; strategic fit and material risk	1–2 pages	Strategic rationale, scale of commitment, risk, how value will be verified

Investment committee memo structure	Content
1 · Decision requested	What, how much, which tranche, by when, under what conditions
2 · Recommendation and rationale	Three to four sentences
3 · Return and timeline	NPV range, payback, breakeven, ramp; conservative scenario first
4 · Key assumptions and evidence	Top five assumptions with evidence maturity
5 · Dependencies and readiness	Roadmap horizon; readiness finding; preconditions
6 · Risks and inaction	Main risks and mitigations; cost and risk of not proceeding
7 · Verification plan	Metrics, baselines, gates, stop rules
Appendix	Model summary, driver tree, traceability register

4.3 The 60-second test

Hand page one to a colleague and ask them to find six things within a minute. If they cannot, restructure.

Must be findable on page one	Example
The decision requested	“Approve Tranche 1: €120k now.”
The return (range) and payback	“NPV €0.1–0.3m base; payback ~2 years; negative below 82% of planned benefits.”
The timeline	“Tranche 1 Jan–Jun; Gate 2 July; steady state month 18.”
The key assumptions	“Acceptance ≥ 60%; 9 FTE temp reduction; inference ≤ €0.04/contact.”
The dependencies	“Knowledge content clean-up (owner: CS ops); works-council agreement.”
The verification plan	“Staggered rollout; baseline signed; stop if acceptance < 45% at Gate 2.”

4.4 Separating strategic rationale from quantitative analysis

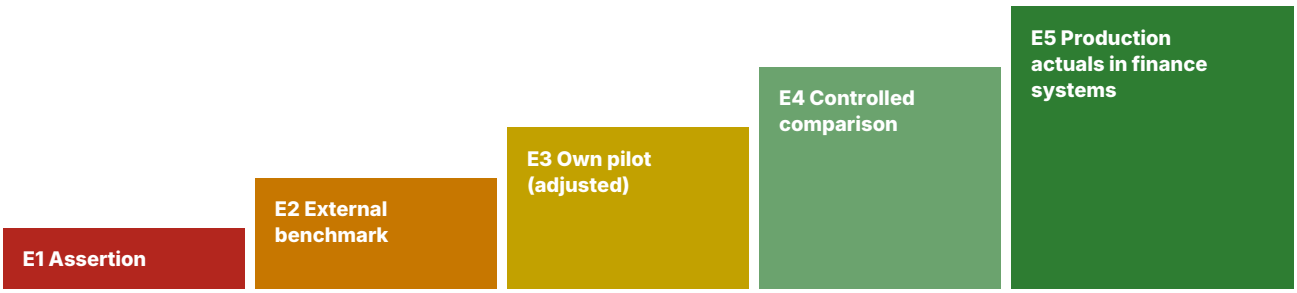
Strategic arguments — competitive position, capability building, customer expectations — are legitimate and often decisive. But they must be presented **as strategy**, separately from the financial analysis. When strategy is converted into unsupported numbers, the whole case loses credibility; when it is hidden, the committee lacks a valid reason to approve a modest-NPV investment.

Do	Don’t
“Financially, NPV is €0.1–0.3m. Strategically, this builds the retrieval platform that three roadmap use cases depend on (option value, described in section 5).”	“Total value €3.5m including €2m strategic platform value.”
“If the committee values the capability-building rationale, the case is acceptable even at the conservative NPV of −€0.05m.”	Adjusting assumptions until NPV is positive to avoid a strategic discussion.

KEY IDEA

Let executives make strategic judgments **explicitly**. A case can legitimately say: "the numbers alone do not justify this; the strategic rationale might — that is your call."

4.5 Connecting financial claims to evidence



Every financial claim is tagged with its evidence maturity; committees weight claims accordingly.

Figure 4.1 — Evidence maturity for financial claims

Claim	Evidence	Maturity
Handling time −18% (planning value)	Own pilot, adjusted (Module 1)	E3
Temp-staff reduction of 9 FTE	Owner commitment + budget mechanism; no actuals yet	E3
Inference ≤ €0.04/contact	Vendor price list + own pilot token counts	E2–E3
Complaint cost −15%	Vendor case study	E1–E2

CONSULTANT'S TIP

Put the evidence-maturity tag next to every number in the summary table. Committees then weight claims themselves — and trust the document more.

4.6 Conservative scenarios

Lead with the **conservative** scenario, then the base, then the upside. A conservative scenario is not a pessimistic fantasy; it uses the lower end of defensible ranges for the assumptions that matter most (usually adoption and benefit level), and the higher end for costs.

Scenario (worked example, €k)	NPV	Payback	What it assumes
Conservative	−167	Not within 3 years	Lower benefits (200/450/500), higher run cost (260), higher investment (450)
Base	223	1.9 years	Adjusted pilot values
Upside	463	1.6 years	Higher benefits (350/700/750)

EXAMPLE · What the conservative scenario changes

The worked example from Module 2 is negative in the conservative case. That does not mean "reject" — it means the committee should not commit the full €400k up front. The decision-ready response is a **staged request**: fund the tranche that tests the critical assumptions, with a stop rule.

4.7 The cost and risk of inaction

"Do nothing" is rarely "nothing changes". Model the do-nothing baseline honestly: volumes may grow, costs rise, competitors move, shadow AI spreads, skills age. But inaction arguments are easy to exaggerate; they need the same evidence discipline as benefits.

Inaction factor	Honest treatment	Overstatement to avoid
Volume growth	"Contacts grow 6%/yr (3-year trend); without change, 5 extra temp FTE by 2028."	"Costs will explode."
Competitive position	"Two competitors have launched agent assist; no evidence yet of customer switching."	"We will lose market share."
Shadow AI risk	"1,100 staff use unapproved tools with customer data (survey); governed alternative reduces exposure."	"A data breach is inevitable."
Option value of waiting	"Waiting 6 months for vendor price reductions saves ~€40k but delays benefits ~€150k."	Ignoring that waiting can be rational

NOTE

Waiting is sometimes the **right** answer — when prices are falling quickly, evidence is about to arrive, or readiness gaps make early investment wasteful. A good inaction analysis states both sides.

4.8 Roadmap and readiness alignment

Check	Inconsistency to avoid
Horizon matches the roadmap (Course 4)	Case requests full-scale funding for an H2 item whose preconditions are not met
Readiness finding reflected (Course 3)	"Not yet" finding, but the case assumes go-live in month 2
Remediation costs included	Readiness remediation (data, oversight, training) missing from costs
Tranches match roadmap gates	Funding schedule ignores gate criteria
Dependencies shared with other cases	Enabler costs counted in several cases — or in none

4.9 Clear decision requests

Element	Example
Decision	Approve Tranche 1 of the customer-service drafting assistant
Amount	€120k (of a €400k envelope)
Scope	60 representative agents in two sites; 6 months

Element	Example
Conditions	Baseline signed; review staffing in place; works-council agreement
Gate and criteria for next tranche	Gate 2 (July): acceptance ≥ 60%, net handling time −12% vs comparison sites, re-contact not worse
Stop rule	Acceptance < 45% or re-contact up > 1 pt → stop and report
Alternatives considered	Do nothing; template library (non-AI); full rollout now
Decision needed by	Investment committee, 14 January

PITFALL · The vague ask

"We seek the committee's support for the AI programme" cannot be approved, declined or held to account. A decision request names an amount, a scope, conditions, a next gate and a stop rule.

4.10 Preserving a credible null recommendation

Sometimes the honest conclusion is **do not invest**, or **not now**. An advisor who can say this credibly is more valuable — and more trusted on the next case — than one whose cases always conclude "proceed".

Element of a credible null recommendation	Example
State it plainly, up front	"We recommend not investing in this initiative now."
Show the numbers that drive it	"Conservative NPV −€0.3m; base NPV −€0.05m; breakeven requires 140% of the pilot result."
Acknowledge what is valuable	"The pilot confirmed staff demand and a 9% time saving."
Say what would change the answer	"Inference price below €0.01/contact, or volume above 600k/yr."
Offer the better alternative	"Invest €40k in a template library (non-AI) capturing ~60% of the benefit."
Set a revisit trigger	"Revisit when vendor pricing changes or volumes pass 500k."

EXAM TRAP

When asked how to respond to a sponsor's pressure to "make the numbers work", the correct answer preserves the analysis and the null recommendation — and offers alternatives (staging, scope change, non-AI options) — rather than adjusting assumptions until NPV turns positive.

4.11 Case study: Pellworth Hotels

CASE STUDY **A guest-messaging assistant the numbers did not support**

Pellworth Hotels (fictional) · 38 hotels in Germany and Austria

Situation

The commercial director wants a GenAI assistant to answer guest messages (booking changes, amenities, local tips) across all hotels. The vendor's case claims €1.4m annual value. The CFO asks the advisor for an investment committee memo.

Analysis

- **Baseline:** 310,000 guest messages/year; average 3.2 minutes each, handled by front-desk staff between other tasks — not a separate cost line.
- **Landing line:** no front-desk staffing reduction was possible (minimum staffing rules), so time saved was **soft ROI**. Hard benefit: reduced overtime (€90–140k/yr) and fewer OTA booking-change fees through faster responses (€30–60k/yr, E2 evidence).
- **Costs:** licences €50k, inference €15k, content upkeep across 38 hotels €25k, review of booking changes €15k per year (€105k); one-off €120k. Year 1 benefits at 60% (ramp).
- **NPV (3 years, 10%):** conservative −€171k (lower benefits, higher costs); base −€41k; upside +€56k.
- **Strategic rationale:** guest-experience leadership; described separately.

The memo's recommendation

Recommendation: do not invest in the full assistant now. The financial case is negative in conservative and base scenarios; the value is mainly soft capacity at the front desk, which minimum staffing rules prevent from becoming savings.

Alternative: approve €45k for (1) templated quick replies in the existing messaging tool (non-AI) and (2) a 3-hotel trial of AI drafting for booking changes only, where OTA fees create hard value.

What would change our view: trial shows ≥ 40% fewer OTA change fees, or the vendor offers usage-based pricing below €0.15 per message.

Strategic note: if the committee values guest-experience positioning above the financial result, note that NPV turns positive only in the upside scenario, where both adoption and fee reduction exceed plan.

Discussion questions

1. Why was most of the vendor's €1.4m soft ROI for Pellworth?
2. How does the memo keep strategy separate from the numbers?
3. Why is the alternative proposal important to the credibility of the null recommendation?
4. What revisit triggers does the memo set?

Model answers (brief)

1. Time saved could not reduce front-desk staffing because of minimum staffing rules, so there was no landing line for most of it.
2. Financial scenarios come first; guest-experience positioning is a separate note the committee can weigh explicitly.
3. It shows the advisor is solving the business problem, not opposing AI, and captures the hard value that does exist.
4. A ≥ 40% reduction in OTA change fees in the trial, or vendor pricing below €0.15 per message.

4.12 Consulting templates

T4.1 Investment Committee Memo

Purpose: Two to four pages that let the committee decide.

1 Decision requested (what · amount · tranche · conditions · by when):

2 Recommendation and rationale (3–4 sentences):

3 Return and timeline — conservative / base / upside NPV; payback; breakeven:

4 Top five assumptions with evidence maturity (E1–E5):

5 Dependencies, roadmap horizon, readiness finding:

6 Risks, mitigations, cost and risk of inaction:

7 Verification plan: metrics, baselines, gates, stop rules:

Strategic rationale (kept separate from the numbers):

T4.2 Board Pre-Read (one page)

Purpose: Strategic fit, scale of commitment and verification for board oversight.

Purpose of the investment and strategic fit:

Scale of commitment (envelope, tranches) and timeline:

Financial summary (range, conservative case first):

Material risks and how they are controlled:

How value will be verified and reported to the board:

Decision or noting requested:

T4.3 Evidence Maturity & Alignment Table

Purpose: Tag every claim and check alignment with roadmap and readiness.

CLAIMS

Financial claim	Value / range	Evidence	Maturity (E1–E5)

ALIGNMENT

- ☐ Roadmap horizon and preconditions consistent with requested funding
- ☐ Readiness finding reflected in timing and scope
- ☐ Readiness remediation costs included
- ☐ Tranches match roadmap gates
- ☐ Shared enabler costs allocated once

T4.4 Inaction Analysis & Decision Request Box

Purpose: State the do-nothing baseline honestly and make the ask precise.

INACTION

Factor	Evidence-based effect	Value / range	Case for waiting?

DECISION REQUEST

Decision: .....

Amount (of envelope): .....

Scope: .....

Conditions: .....

Next gate and criteria: .....

Stop rule: .....

Alternatives considered: .....

Decision-holder and date: .....

T4.5 Null Recommendation & 60-Second Test

Purpose: Keep a credible “no” and check page-one findability.

NULL RECOMMENDATION

Recommendation stated plainly: .....

Numbers driving it: .....

What is valuable anyway: .....

What would change the answer: .....

Better alternative: .....

Revisit trigger: .....

60-SECOND TEST (PAGE ONE)

- [ ] Decision requested
- [ ] Return range and payback
- [ ] Timeline
- [ ] Key assumptions
- [ ] Dependencies
- [ ] Verification plan

4.13 Best practices and pitfalls

✓ DO THIS

- Put the decision and recommendation on page one.
- Match format and length to the forum.
- Keep strategic rationale separate from the numbers.
- Tag claims with evidence maturity.
- Lead with the conservative scenario.
- Model inaction honestly, including when waiting is rational.
- End with a precise, staged decision request and stop rule.

× AVOID THIS

- Conclusions on page twelve.
- One document for every audience.
- Strategic value folded into NPV.
- Vendor claims presented as facts.
- Optimistic base case only.
- Scare-story inaction arguments.
- Adjusting assumptions to make NPV positive.

4.14 Module 4 cheat sheet

MODULE 4 · ONE-PAGE SUMMARY

Making the Decision-Ready Case

DECISION-READY

Find, trust, act: return, timeline, assumptions, dependencies, verification. Can recommend “no”.

FORMATS

CFO case (8–15 pp + model) · **IC memo** (2–4 pp, decision) · **board pre-read** (1–2 pp, strategy, risk, verification).

60-SECOND TEST

Page one: decision · return range & payback · timeline · key assumptions · dependencies · verification.

STRATEGY VS NUMBERS

Separate sections. Let executives judge strategy explicitly; never fold it into NPV.

EVIDENCE MATURITY

E1 assertion · E2 external benchmark · E3 own pilot (adjusted) · E4 controlled comparison · E5 production actuals.

CONSERVATIVE FIRST

Conservative → base → upside. Negative conservative case → staged request, not automatic rejection.

INACTION

Volume trends, competition, shadow AI, skills — evidenced. Waiting can be rational.

ALIGNMENT

Roadmap horizon · readiness finding · remediation costs · tranches = gates · shared enablers once.

DECISION REQUEST & NULL RECOMMENDATION

Request: decision · amount · scope · conditions · next gate & criteria · stop rule · alternatives · by when. Null: plain statement · driving numbers · what is valuable · what would change it · better alternative · revisit trigger.

4.15 Knowledge check

Answer all questions before turning the page.

K1 **Which format suits a formal approval forum deciding on a specific funding request?**

- (A) A 40-page strategy report
- (B) A vendor proposal
- (C) An investment committee memo of 2–4 pages, built around the decision request
- (D) A board pre-read

K2 **What does the 60-second test check?**

- (A) Whether the presenter can pitch in a minute
- (B) Whether the model calculates in under a minute
- (C) Whether the NPV is positive
- (D) Whether a reader can locate the decision, return, timeline, assumptions, dependencies and verification plan on page one

K3 **A case includes €2m "strategic platform value" in its NPV. What should be done?**

- (A) Remove it from the NPV and present the strategic rationale separately, so executives can judge it explicitly
- (B) Delete all mention of strategy
- (C) Keep it — strategy matters.
- (D) Double it to reflect future use cases

K4 **A claim is supported only by a vendor case study. What evidence maturity should be shown?**

- (A) No tag needed
- (B) E1–E2 (assertion / external benchmark)
- (C) E4
- (D) E5

K5 **The conservative scenario is negative while the base case is positive. What is the decision-ready response?**

- (A) Request a staged tranche that tests the critical assumptions, with a stop rule
- (B) Raise the base assumptions.
- (C) Hide the conservative scenario.
- (D) Reject automatically.

K6 **Which inaction statement is most defensible?**

- (A) "Doing nothing costs nothing."
- (B) "A breach is inevitable."
- (C) "Contacts grow 6% a year on a 3-year trend; without change we need ~5 more temp FTE by 2028."
- (D) "Competitors will destroy us if we wait."

K7 **A readiness assessment says “not yet”, but the case assumes go-live in month 2 with no remediation costs. What is the issue?**

(A) The readiness assessment is wrong.
(B) Misalignment: timing should follow readiness preconditions, and remediation costs belong in the case
(C) The case should not mention readiness.
(D) None — the case and readiness are separate.

K8 **A sponsor asks the advisor to “make the numbers work” for a case that is negative in all scenarios. What should the advisor do?**

(A) Increase adoption assumptions until NPV is positive
(B) Refuse to write anything
(C) Remove the costs that make it negative
(D) Present a credible null recommendation with what would change it, a better alternative and a revisit trigger

4.16 Knowledge check: answers and explanations

K1 **Answer: C** Review: 4.2

Which format suits a formal approval forum deciding on a specific funding request?

- ✗ (A) A 40-page strategy report — **Wrong.** Too long, and not decision-focused.
- ✗ (B) A vendor proposal — **Wrong.** Not the client's document.
- ✓ (C) An investment committee memo of 2–4 pages, built around the decision request — **Correct.** Designed for formal approval.
- ✗ (D) A board pre-read — **Wrong.** Boards oversee; committees approve specific requests.

K2 **Answer: D** Review: 4.3

What does the 60-second test check?

- ✗ (A) Whether the presenter can pitch in a minute — **Wrong.** Not the test.
- ✗ (B) Whether the model calculates in under a minute — **Wrong.** Not the test.
- ✗ (C) Whether the NPV is positive — **Wrong.** Not the test.
- ✓ (D) Whether a reader can locate the decision, return, timeline, assumptions, dependencies and verification plan on page one — **Correct.** Findability of key facts.

K3 **Answer: A** Review: 4.4

A case includes €2m "strategic platform value" in its NPV. What should be done?

- ✓ (A) Remove it from the NPV and present the strategic rationale separately, so executives can judge it explicitly — **Correct.** Separation preserves credibility and lets strategy be weighed openly.
- ✗ (B) Delete all mention of strategy — **Wrong.** Strategy is a legitimate consideration.
- ✗ (C) Keep it — strategy matters. — **Wrong.** Strategy matters but must not inflate the numbers.
- ✗ (D) Double it to reflect future use cases — **Wrong.** Further inflation.

K4 **Answer: B** Review: 4.5

A claim is supported only by a vendor case study. What evidence maturity should be shown?

- ✗ (A) No tag needed — **Wrong.** Every claim should be tagged.
- ✓ (B) E1–E2 (assertion / external benchmark) — **Correct.** Vendor material is weak, external evidence.
- ✗ (C) E4 — **Wrong.** E4 needs a controlled comparison.
- ✗ (D) E5 — **Wrong.** E5 means production actuals.

K5 **Answer: A** Review: 4.6

The conservative scenario is negative while the base case is positive. What is the decision-ready response?

- ✓ (A) Request a staged tranche that tests the critical assumptions, with a stop rule — **Correct.** Staging addresses the downside.
- ✗ (B) Raise the base assumptions. — **Wrong.** Manipulation.
- ✗ (C) Hide the conservative scenario. — **Wrong.** This destroys trust.
- ✗ (D) Reject automatically. — **Wrong.** A negative conservative case does not mean automatic rejection.

K6 **Answer: C** Review: 4.7

Which inaction statement is most defensible?

✗ (A) "Doing nothing costs nothing." — **Wrong.** Ignores real trends.

✗ (B) "A breach is inevitable." — **Wrong.** Overstated.

✓ (C) "Contacts grow 6% a year on a 3-year trend; without change we need ~5 more temp FTE by 2028." — **Correct.** Evidence-based and quantified.

✗ (D) "Competitors will destroy us if we wait." — **Wrong.** An overstatement with no evidence.

K7 **Answer: B** Review: 4.8

A readiness assessment says "not yet", but the case assumes go-live in month 2 with no remediation costs. What is the issue?

✗ (A) The readiness assessment is wrong. — **Wrong.** No basis for that.

✓ (B) Misalignment: timing should follow readiness preconditions, and remediation costs belong in the case — **Correct.** Alignment is a decision-readiness requirement.

✗ (C) The case should not mention readiness. — **Wrong.** It must.

✗ (D) None — the case and readiness are separate. — **Wrong.** They must be consistent.

K8 **Answer: D** Review: 4.10

A sponsor asks the advisor to "make the numbers work" for a case that is negative in all scenarios. What should the advisor do?

✗ (A) Increase adoption assumptions until NPV is positive — **Wrong.** This corrupts the analysis.

✗ (B) Refuse to write anything — **Wrong.** Unhelpful.

✗ (C) Remove the costs that make it negative — **Wrong.** Dishonest.

✓ (D) Present a credible null recommendation with what would change it, a better alternative and a revisit trigger — **Correct.** This preserves credibility and still helps the business.

R1 Course at a glance: the master cheat sheet

ALL FOUR MODULES ON ONE PAGE

AI Business Cases and ROI Models

M1 · STRUCTURING THE CASE

Inform the decision, including "no". Driver trees (with negative drivers). Traceability: AI change → metric → conversion → **landing line** → **owner**. **Hard · soft · option** in separate columns. Efficiency vs new value. Visible **pilot-to-production** adjustments. Ranges, scenarios, sensitivity, staging. Assumption register. Pyramid executive summary; committee expectations.

M2 · ROI MODEL

Inputs → calculations → outputs → scenarios → checks. TCO: one-off, recurring, exit. **Hidden costs**: verification & rework, review, change, inference (scales with success), transfers, upkeep. Ramps. **NPV** core; payback; IRR with care. Worked example: NPV ≈ €223k, payback 1.9y, IRR ≈ 33%, breakeven at ~82% of benefits. Staging values the option to stop. Created vs relocated.

M3 · VALUE METRICS

Hierarchy: adoption → usage quality → outcome → €. No vanity metrics. Link to owners' KPIs. Baselines before go-live. **Paired quality** metrics. Designs: before/after · matched · DiD · staggered · randomized; attribution A0–A4 and matching claim wording. Leading → lagging. Honest proxies. Y1/Y2 targets aligned with ramp.

M4 · DECISION-READY CASE

CFO case · IC memo · board pre-read. **60-second test**. Strategy separate from numbers. Evidence maturity E1–E5. **Conservative first**. Honest inaction analysis. Align with roadmap and readiness. Precise decision request with stop rule. **Credible null recommendation**.

R2 Template index

Code	Template	Use when
T1.1	Business Case Skeleton	Structuring
T1.2	Value Driver Tree	Benefits
T1.3	Benefit Traceability Register	Benefits
T1.4	Pilot-to-Production Adjustment Sheet	Using pilot results
T1.5	Assumption Register	Throughout
T2.1	ROI Model Input Sheet	Modelling
T2.2	Cost Register & TCO	Costs
T2.3	Hidden Cost Checklist	Costs
T2.4	Cash Flow, NPV & Payback Worksheet	Financials
T2.5	Sensitivity, Scenario & Breakeven Sheet	Financials
T2.6	Finance Challenge Preparation	Before finance review
T3.1	Metric Hierarchy Map	Measurement
T3.2	Baseline Record	Before go-live
T3.3	Paired Metric & Comparison Design Sheet	Measurement design

Code	Template	Use when
T3.4	Measurement Plan with Y1/Y2 Targets	Measurement
T4.1	Investment Committee Memo	Approval
T4.2	Board Pre-Read	Board oversight
T4.3	Evidence Maturity & Alignment Table	Approval
T4.4	Inaction Analysis & Decision Request Box	Approval
T4.5	Null Recommendation & 60-Second Test	Approval

R3 Glossary

A

Accountable benefit owner
The budget holder for a landing line who accepts accountability for realising a benefit.

Adoption metric
A measure of use (active users, sessions); necessary but not evidence of value.

Adoption ramp
The gradual increase in usage and benefit after go-live, often S-shaped.

Attribution level
Strength of evidence that AI caused an observed change (A0 anecdote to A4 randomized).

B

Baseline
Pre-intervention measurement of a metric with fixed definitions, period and source.

Benchmark
External or internal comparison point; context, not a substitute for a baseline.

Benefit realization
Tracking and banking benefits against plan through finance systems and budget changes.

Benefit traceability
Chain from AI change → operational metric → conversion → landing line → owner.

Board pre-read
One- to two-page document for board oversight: strategic fit, commitment, risk, verification.

Breakeven
The value of an assumption (e.g., benefit level) at which NPV equals zero.

C

Capital budgeting
Evaluating investments by their expected cash flows, timing and risk.

Capitalisation
Accounting treatment recording certain costs as assets depreciated over time; follows finance policy.

Comparison design
Method for isolating AI's effect: before/after, matched, difference-in-differences, staggered, randomized.

Conservative scenario
Scenario using the lower end of defensible benefit ranges and higher end of cost ranges.

Cost register
List of all cost items with category, timing, driver, range, owner and accounting treatment.

Cost transfer
Effort or cost moved to another team, customer or partner rather than removed.

D

Decision request
Precise ask: decision, amount, scope, conditions, next gate, stop rule, alternatives, date.

Decision-ready case
A case in which the decision-holder can find, trust and act on the key facts quickly.

Difference-in-differences
Compares the change in a treated group with the change in a comparison group.

Discount rate
Required rate of return used to convert future cash flows to present value; set by finance.

Discounted payback
Time until cumulative discounted cash flow turns positive.

Driver tree
Decomposition of a financial outcome into measurable value drivers.

E

Efficiency gain
Benefit from doing existing work at lower cost or time.

Evidence maturity
Strength of evidence for a financial claim (E1 assertion to E5 production actuals).

Executive summary
First-page synthesis: recommendation, ask, return range, assumptions, risks, what would change it.

H

Hard ROI
Benefits appearing as cash or P&L changes on a named landing line, with an owner.

Hidden costs
AI costs often omitted: verification, rework, human review, change management, inference, transfers, upkeep.

Human review cost
Cost of people checking AI outputs: volume × rate × time × loaded cost.

I

Inaction analysis
Evidence-based assessment of the cost and risk of not proceeding, including when waiting is rational.

Inference cost
Usage-based cost of running AI models; scales with volume and prompt length.

Investment committee memo
Two- to four-page document for a formal approval decision.

IRR
Internal rate of return: the discount rate at which NPV equals zero.

L

Lagging indicator
Outcome measure that appears later, e.g., financial value.

Landing line
The P&L, budget or balance-sheet line where a benefit appears.

Leading indicator
Early signal predictive of later value, e.g., adoption, acceptance rate.

N

Net present value (NPV)
Sum of cash flows discounted at the required rate; positive NPV creates value at that rate.

New-value claim
Benefit from new revenue or capability; needs market or customer evidence.

Null recommendation
A recommendation not to invest, or not now, presented with reasons, alternatives and revisit triggers.

O

Option value (strategic)
The value of the right, not obligation, to pursue future opportunities created by an investment.

P

Paired quality metric
Quality measure reported with a productivity metric to reveal quality trade-offs.

Payback period
Time until cumulative net cash flow turns positive.

Pilot-to-production adjustment
Visible factors reducing pilot results for selection, novelty, case mix, adoption and the ledger.

Proxy measure
A stand-in metric for value that is hard or slow to measure; must state its limits.

R

Randomized comparison (A/B)
Random assignment to AI and non-AI groups; strongest attribution.

Relocated effort
Effort moved rather than removed; a cost line, not a saving.

S

Scenario analysis
Evaluating outcomes when several named assumptions change together.

Sensitivity analysis
Testing how outcomes change when one assumption changes (e.g., tornado chart).

Soft ROI
Real benefits not converted into cash or P&L changes, e.g., capacity, satisfaction.

Staged investment
Committing funds in tranches with gates and options to stop.

Staggered rollout
Deploying in waves so not-yet-live units serve as comparisons.

Stop rule
Predefined result that ends further investment.

T

Total cost of ownership (TCO)
All one-off, recurring and exit costs over the life of a solution.

V

Value driver
Measurable operational factor that causes financial value to change.

Vanity metric
Impressive-looking metric with no link to decisions or value.

Verification and rework cost
Effort to check and correct AI outputs and downstream errors.

R4 Mock exam

Format: 38 multiple-choice questions · one best answer each · **Time:** 60 minutes · **Suggested pass mark:** 70% (27 correct)

Questions 1–9 cover Module 1, 10–18 Module 2, 19–26 Module 3, 27–34 Module 4 and 35–38 are integrated scenarios. Answer every question before checking the answer key, which explains why each option is right or wrong.

Q1 **What makes a business case decision-ready for finance?**

- A Transparent evidence that finance can verify and challenge, enabling approve, stage or decline
- B Vendor endorsement
- C Persuasive language
- D A large headline number

Q2 **What is missing from “saves 12,000 hours” as a hard benefit?**

- A A chart
- B A larger number
- C A vendor quote
- D A conversion mechanism, landing line and accountable owner

Q3 **Which belongs in the NPV?**

- A Strategic platform value
- B Contractor spend reduction with landing line and owner
- C Hours freed without budget change
- D Employee satisfaction gains

Q4 **Which claim is a new-value claim?**

- A Fewer contractor hours
- B Lower invoice processing cost
- C Higher conversion from AI-personalised offers
- D Less rework

Q5 **A pilot with volunteers and vendor support showed 40% savings. Which factor list should adjust it?**

- A None — pilots are reliable
- B Exchange rates
- C Inflation and tax
- D User selection, novelty, case mix, vendor support, adoption and the ledger

Q6 **Which statement about ranges is best practice?**

- A Ranges only for costs
- B Use symmetric ± ranges
- C Use asymmetric ranges: benefits skewed lower, costs higher
- D Avoid ranges for clarity

Q7 **What does an assumption register add to decision quality?**

- A) It guarantees good outcomes.
- B) It makes assumptions explicit, sourced, owned and tested, focusing attention on the most sensitive ones
- C) It replaces the model.
- D) It is only for auditors.

Q8 **Which executive summary element comes first?**

- A) The recommendation and decision requested
- B) Technology description
- C) Team biographies
- D) Methodology

Q9 **Investment committees typically react badly to…**

- A) Stop rules
- B) Ranges with stated sensitivities
- C) Single-point estimates and licence-only costs
- D) Named benefit owners

Q10 **Where should every input in an ROI model live?**

- A) On one input sheet with source, owner, date and range
- B) In e-mails
- C) In the vendor's tool
- D) Inside the formulas

Q11 **Which is a hidden cost of AI adoption?**

- A) Office rent
- B) Human review of outputs
- C) CEO salary
- D) Vendor licence

Q12 **A pilot's inference cost was €0.03 per interaction. At 300,000 interactions a month with prompts twice as long, what is the approximate monthly cost?**

- A) €900
- B) €60
- C) €9,000
- D) €18,000

Q13 **Which measure is the core value measure at advisory depth?**

- A) Payback only
- B) Number of users
- C) Simple ROI
- D) NPV at the finance-agreed discount rate

Q14 **Net cash flows −400, +60, +360, +360 (€k). When does cumulative cash turn positive?**

- A In year 3
- B Late in year 2 (≈1.9 years)
- C Never
- D During year 1

Q15 **What does a breakeven at 82% of planned benefits imply?**

- A Benefit assumptions are critical and should be tested early, for example through staging
- B The discount rate is wrong.
- C Costs are too low.
- D The case is safe.

Q16 **Why can staging raise expected NPV?**

- A It increases benefits.
- B It removes risk.
- C It limits spending in failure states through the option to stop
- D It reduces tax.

Q17 **Clerks save 6 minutes while supervisors add 2 minutes of checking. What is the created value?**

- A 6 minutes
- B 4 minutes, with 2 minutes recorded as a relocated cost
- C 8 minutes
- D 0 minutes

Q18 **Which decision belongs to finance policy rather than the advisor's model design?**

- A Tornado format
- B Scenario names
- C Input colour coding
- D Discount rate and capitalisation approach

Q19 **Which metric proves ROI?**

- A Weekly active users
- B Draft acceptance rate
- C Contractor cost reduction on the landing line
- D Handling time

Q20 **"1.2 million prompts this quarter" is best described as…**

- A A vanity metric unless linked to decisions and value
- B A quality metric
- C A baseline
- D A financial metric

Q21 **Why report AI value through existing management packs?**

- A) So value appears in indicators budget owners already use for decisions
- B) To save printing
- C) To hide AI
- D) Because AI metrics are illegal

Q22 **Which baseline is defensible?**

- A) Industry average
- B) Twelve weeks of pre-launch data with fixed definitions, seasonality noted and signed off
- C) The first post-launch month
- D) Users' recollection

Q23 **Which pair is correct?**

- A) Memos drafted with active users
- B) Deflection with sessions
- C) Handling time with re-contact rate
- D) Handling time with number of prompts

Q24 **Which design gives the strongest attribution?**

- A) Before / after
- B) User survey
- C) Matched comparison
- D) Randomized (A/B)

Q25 **Early after launch, which indicators dominate?**

- A) None
- B) Lagging financial value
- C) Annual P&L only
- D) Leading: adoption and usage quality

Q26 **How should a proxy be presented?**

- A) As financial value
- B) Labelled as a proxy, with its link to value, its limits and when it will be replaced
- C) Hidden
- D) As a baseline

Q27 **Which document suits board oversight of a large AI commitment?**

- A) A 1–2-page pre-read on strategic fit, commitment, risk and verification
- B) Raw data
- C) A 15-page model walkthrough
- D) A vendor brochure

Q28 **Which item must be findable on page one?**

- A Model formulas
- B Vendor history
- C The decision requested and return range
- D Team photos

Q29 **How should strategic rationale be treated?**

- A Converted to NPV
- B Omitted
- C Presented separately, so executives judge it explicitly
- D Placed in an appendix only

Q30 **A claim rests on own production actuals in finance systems. Which maturity?**

- A E1
- B E3
- C E2
- D E5

Q31 **Which scenario should lead a decision-ready case?**

- A Conservative
- B Upside
- C Vendor scenario
- D Any

Q32 **When can waiting be rational in an inaction analysis?**

- A Always
- B When prices are falling quickly, evidence is imminent or readiness gaps would waste early investment
- C Only when competitors wait
- D Never

Q33 **Which is a clear decision request?**

- A “Support the AI programme.”
- B “Approve Tranche 1, €120k of €400k, 60 agents for 6 months; Gate 2 criteria …; stop if acceptance < 45%; decide by 14 January.”
- C “Approve AI.”
- D “Note the update.”

Q34 **What makes a null recommendation credible?**

- A Saying no without explanation
- B Delaying the memo
- C Blaming the vendor
- D Driving numbers, what is valuable, what would change the answer, a better alternative and a revisit trigger

Q35 **Integrated: A pilot shows 40% savings; the hard landing line covers only overflow outsourcing; inference cost is unmodelled at scale. What should the case do?**

(A) Adjust the pilot, count only traceable hard benefit, add inference and other hidden costs, and present ranges with staging
(B) Use 40% across all staff costs
(C) Drop the case
(D) Add strategic value to close the gap

Q36 **Integrated: Year-1 targets in the measurement plan show full financial value, while the ROI model ramps to steady state in month 18. What is wrong?**

(A) Nothing
(B) The model should drop the ramp
(C) Targets and model disagree; year-1 targets should reflect the ramp
(D) Year-2 targets should be removed

Q37 **Integrated: The conservative NPV is negative, the base positive, the readiness finding is "phase", and the roadmap places scale in H2. What decision request fits?**

(A) A Tranche 1 for the ready phase, with gate criteria aligned to the roadmap and a stop rule
(B) Defer indefinitely
(C) Full funding now
(D) Reject

Q38 **Integrated: Which sequence reflects this course?**

(A) Write memo → build model → define metrics
(B) Choose vendor → justify
(C) Structure the case (drivers, traceability) → build the ROI model (TCO, NPV, sensitivity) → define value metrics and measurement → shape the decision-ready case
(D) Metrics → vendor → memo

R5 Mock exam: answer key and explanations

Quick answer grid

1 · A	**2** · D	**3** · B	**4** · C	**5** · D	**6** · C	**7** · B	**8** · A	**9** · C	**10** · A
11 · B	**12** · D	**13** · D	**14** · B	**15** · A	**16** · C	**17** · B	**18** · D	**19** · C	**20** · A
21 · A	**22** · B	**23** · C	**24** · D	**25** · D	**26** · B	**27** · A	**28** · C	**29** · C	**30** · D
31 · A	**32** · B	**33** · B	**34** · D	**35** · A	**36** · C	**37** · A	**38** · C		

Score	Interpretation	Next step
32–38	Exam-ready	Review missed items; run a real assessment with the templates
27–31	Pass level	Re-read cheat sheets for modules where you missed two or more
below 27	Not yet	Re-study weak modules using the “Review” references; retake in one week

Full explanations

Q1 **Answer: A** — Review: M1 · 1.1

What makes a business case decision-ready for finance?

✓ **A** Transparent evidence that finance can verify and challenge, enabling approve, stage or decline — **Correct.** Verification and challenge are the test.

✗ **B** Vendor endorsement — **Wrong.** Not independent evidence.

✗ **C** Persuasive language — **Wrong.** Persuasion reduces trust.

✗ **D** A large headline number — **Wrong.** Size does not create trust.

Q2 **Answer: D** — Review: M1 · 1.4

What is missing from “saves 12,000 hours” as a hard benefit?

✗ **A** A chart — **Wrong.** Presentation is not the issue.

✗ **B** A larger number — **Wrong.** Irrelevant.

✗ **C** A vendor quote — **Wrong.** Irrelevant.

✓ **D** A conversion mechanism, landing line and accountable owner — **Correct.** Without these it is soft ROI.

Q3 **Answer: B** — Review: M1 · 1.6

Which belongs in the NPV?

✗ **A** Strategic platform value — **Wrong.** Described separately.

✓ **B** Contractor spend reduction with landing line and owner — **Correct.** Hard ROI.

✗ **C** Hours freed without budget change — **Wrong.** Soft ROI.

✗ **D** Employee satisfaction gains — **Wrong.** Soft ROI.

Q4 **Answer: C** Review: M1 · 1.7

Which claim is a new-value claim?

✗ (A) Fewer contractor hours — **Wrong.** Efficiency.

✗ (B) Lower invoice processing cost — **Wrong.** Efficiency.

✓ (C) Higher conversion from AI-personalised offers — **Correct.** Revenue: new value.

✗ (D) Less rework — **Wrong.** Efficiency.

Q5 **Answer: D** Review: M1 · 1.8

A pilot with volunteers and vendor support showed 40% savings. Which factor list should adjust it?

✗ (A) None — pilots are reliable — **Wrong.** They overstate production results.

✗ (B) Exchange rates — **Wrong.** Irrelevant.

✗ (C) Inflation and tax — **Wrong.** Not pilot effects.

✓ (D) User selection, novelty, case mix, vendor support, adoption and the ledger — **Correct.** These are pilot-to-production factors.

Q6 **Answer: C** Review: M1 · 1.9

Which statement about ranges is best practice?

✗ (A) Ranges only for costs — **Wrong.** Benefits need ranges too.

✗ (B) Use symmetric ± ranges — **Wrong.** Benefits usually skew lower and costs higher.

✓ (C) Use asymmetric ranges: benefits skewed lower, costs higher — **Correct.** Reflects how estimates tend to err.

✗ (D) Avoid ranges for clarity — **Wrong.** Hides uncertainty.

Q7 **Answer: B** Review: M1 · 1.10

What does an assumption register add to decision quality?

✗ (A) It guarantees good outcomes. — **Wrong.** Good decisions can still turn out badly.

✓ (B) It makes assumptions explicit, sourced, owned and tested, focusing attention on the most sensitive ones — **Correct.** Assumption quality drives decision quality.

✗ (C) It replaces the model. — **Wrong.** It complements the model.

✗ (D) It is only for auditors. — **Wrong.** It serves decision-makers.

Q8 **Answer: A** Review: M1 · 1.11

Which executive summary element comes first?

✓ (A) The recommendation and decision requested — **Correct.** Pyramid logic: answer first.

✗ (B) Technology description — **Wrong.** Supporting detail.

✗ (C) Team biographies — **Wrong.** Irrelevant.

✗ (D) Methodology — **Wrong.** Supporting detail.

Q9 **Answer: C** Review: M1 · 1.11

Investment committees typically react badly to…

✗ (A) Stop rules — **Wrong.** Committees expect these.

✗ (B) Ranges with stated sensitivities — **Wrong.** Committees welcome these.

✓ (C) Single-point estimates and licence-only costs — **Correct.** These signal weak analysis.

✗ (D) Named benefit owners — **Wrong.** Committees expect these.

Q10 **Answer: A** Review: M2 · 2.1

Where should every input in an ROI model live?

✓ **A** On one input sheet with source, owner, date and range — **Correct.** Transparency and challengeability.

✗ **B** In e-mails — **Wrong.** Not auditable.

✗ **C** In the vendor's tool — **Wrong.** Not controlled by the client.

✗ **D** Inside the formulas — **Wrong.** Hard-coding makes the model unauditable.

Q11 **Answer: B** Review: M2 · 2.3

Which is a hidden cost of AI adoption?

✗ **A** Office rent — **Wrong.** Unrelated.

✓ **B** Human review of outputs — **Correct.** Often omitted, yet material.

✗ **C** CEO salary — **Wrong.** Unrelated.

✗ **D** Vendor licence — **Wrong.** Visible, if incomplete.

Q12 **Answer: D** Review: M2 · 2.3

A pilot's inference cost was €0.03 per interaction. At 300,000 interactions a month with prompts twice as long, what is the approximate monthly cost?

✗ **A** €900 — **Wrong.** Wrong order of magnitude.

✗ **B** €60 — **Wrong.** That was the pilot cost.

✗ **C** €9,000 — **Wrong.** This ignores the longer prompts.

✓ **D** €18,000 — **Correct.** 300,000 × €0.03 × 2.

Q13 **Answer: D** Review: M2 · 2.6

Which measure is the core value measure at advisory depth?

✗ **A** Payback only — **Wrong.** Ignores value after payback.

✗ **B** Number of users — **Wrong.** Not a financial measure.

✗ **C** Simple ROI — **Wrong.** Ignores timing.

✓ **D** NPV at the finance-agreed discount rate — **Correct.** Positive NPV creates value at the required return.

Q14 **Answer: B** Review: M2 · 2.6

Net cash flows −400, +60, +360, +360 (€k). When does cumulative cash turn positive?

✗ **A** In year 3 — **Wrong.** It turns positive in year 2.

✓ **B** Late in year 2 (≈1.9 years) — **Correct.** −340 + 360 turns positive near the end of year 2.

✗ **C** Never — **Wrong.** It does.

✗ **D** During year 1 — **Wrong.** Cumulative is −340 after year 1.

Q15 **Answer: A** Review: M2 · 2.8

What does a breakeven at 82% of planned benefits imply?

✓ **A** Benefit assumptions are critical and should be tested early, for example through staging — **Correct.** Small margin for error.

✗ **B** The discount rate is wrong. — **Wrong.** Unrelated.

✗ **C** Costs are too low. — **Wrong.** Not implied.

✗ **D** The case is safe. — **Wrong.** An 18% shortfall would destroy value.

Q16 **Answer: C** Review: M2 · 2.7

Why can staging raise expected NPV?

✗ A It increases benefits. — **Wrong.** Not directly.

✗ B It removes risk. — **Wrong.** It reduces exposure, not all risk.

✓ C It limits spending in failure states through the option to stop — **Correct.** The option to stop has value.

✗ D It reduces tax. — **Wrong.** No.

Q17 **Answer: B** Review: M2 · 2.11

Clerks save 6 minutes while supervisors add 2 minutes of checking. What is the created value?

✗ A 6 minutes — **Wrong.** This ignores the transfer.

✓ B 4 minutes, with 2 minutes recorded as a relocated cost — **Correct.** Count net effort removed.

✗ C 8 minutes — **Wrong.** Incorrect.

✗ D 0 minutes — **Wrong.** Some value is created.

Q18 **Answer: D** Review: M2 · 2.1

Which decision belongs to finance policy rather than the advisor's model design?

✗ A Tornado format — **Wrong.** A modelling choice.

✗ B Scenario names — **Wrong.** A modelling choice.

✗ C Input colour coding — **Wrong.** A convention.

✓ D Discount rate and capitalisation approach — **Correct.** Finance policy.

Q19 **Answer: C** Review: M3 · 3.1

Which metric proves ROI?

✗ A Weekly active users — **Wrong.** Adoption.

✗ B Draft acceptance rate — **Wrong.** Usage quality.

✓ C Contractor cost reduction on the landing line — **Correct.** Financial value — the top of the hierarchy.

✗ D Handling time — **Wrong.** Operational outcome; not yet money.

Q20 **Answer: A** Review: M3 · 3.2

"1.2 million prompts this quarter" is best described as…

✓ A A vanity metric unless linked to decisions and value — **Correct.** Activity without value linkage.

✗ B A quality metric — **Wrong.** No.

✗ C A baseline — **Wrong.** No.

✗ D A financial metric — **Wrong.** No.

Q21 **Answer: A** Review: M3 · 3.3

Why report AI value through existing management packs?

✓ A So value appears in indicators budget owners already use for decisions — **Correct.** Accountability.

✗ B To save printing — **Wrong.** Trivial.

✗ C To hide AI — **Wrong.** No.

✗ D Because AI metrics are illegal — **Wrong.** No.

Q22 **Answer: B** Review: M3 · 3.5

Which baseline is defensible?

✗ (A) Industry average — **Wrong.** Context only.

✓ (B) Twelve weeks of pre-launch data with fixed definitions, seasonality noted and signed off — **Correct.** Meets baseline rules.

✗ (C) The first post-launch month — **Wrong.** Already treated.

✗ (D) Users' recollection — **Wrong.** Unreliable.

Q23 **Answer: C** Review: M3 · 3.6

Which pair is correct?

✗ (A) Memos drafted with active users — **Wrong.** Not a quality pair.

✗ (B) Deflection with sessions — **Wrong.** Not a quality pair.

✓ (C) Handling time with re-contact rate — **Correct.** Quality guard against cutting corners.

✗ (D) Handling time with number of prompts — **Wrong.** Not a quality pair.

Q24 **Answer: D** Review: M3 · 3.7–3.8

Which design gives the strongest attribution?

✗ (A) Before / after — **Wrong.** Weakest measured design.

✗ (B) User survey — **Wrong.** Anecdotal.

✗ (C) Matched comparison — **Wrong.** Moderate.

✓ (D) Randomized (A/B) — **Correct.** Strongest attribution.

Q25 **Answer: D** Review: M3 · 3.9

Early after launch, which indicators dominate?

✗ (A) None — **Wrong.** Measurement starts at launch.

✗ (B) Lagging financial value — **Wrong.** Too early.

✗ (C) Annual P&L only — **Wrong.** Too late.

✓ (D) Leading: adoption and usage quality — **Correct.** They inform early decisions.

Q26 **Answer: B** Review: M3 · 3.10

How should a proxy be presented?

✗ (A) As financial value — **Wrong.** Overstatement.

✓ (B) Labelled as a proxy, with its link to value, its limits and when it will be replaced — **Correct.** Honest proxy use.

✗ (C) Hidden — **Wrong.** No.

✗ (D) As a baseline — **Wrong.** Incorrect.

Q27 **Answer: A** Review: M4 · 4.2

Which document suits board oversight of a large AI commitment?

✓ (A) A 1–2-page pre-read on strategic fit, commitment, risk and verification — **Correct.** Board format.

✗ (B) Raw data — **Wrong.** No.

✗ (C) A 15-page model walkthrough — **Wrong.** Too detailed.

✗ (D) A vendor brochure — **Wrong.** No.

Q28 **Answer: C** Review: M4 · 4.3

Which item must be findable on page one?

✗ A Model formulas — **Wrong.** Appendix.

✗ B Vendor history — **Wrong.** Irrelevant.

✓ C The decision requested and return range — **Correct.** Part of the 60-second test.

✗ D Team photos — **Wrong.** Irrelevant.

Q29 **Answer: C** Review: M4 · 4.4

How should strategic rationale be treated?

✗ A Converted to NPV — **Wrong.** This inflates the numbers.

✗ B Omitted — **Wrong.** It can be decisive.

✓ C Presented separately, so executives judge it explicitly — **Correct.** Separation preserves credibility.

✗ D Placed in an appendix only — **Wrong.** It may deserve prominence, but kept separate.

Q30 **Answer: D** Review: M4 · 4.5

A claim rests on own production actuals in finance systems. Which maturity?

✗ A E1 — **Wrong.** Assertion.

✗ B E3 — **Wrong.** Own pilot.

✗ C E2 — **Wrong.** External.

✓ D E5 — **Correct.** Production actuals.

Q31 **Answer: A** Review: M4 · 4.6

Which scenario should lead a decision-ready case?

✓ A Conservative — **Correct.** Lead with conservative, then base and upside.

✗ B Upside — **Wrong.** Optimistic.

✗ C Vendor scenario — **Wrong.** Not independent.

✗ D Any — **Wrong.** Order matters.

Q32 **Answer: B** Review: M4 · 4.7

When can waiting be rational in an inaction analysis?

✗ A Always — **Wrong.** No.

✓ B When prices are falling quickly, evidence is imminent or readiness gaps would waste early investment — **Correct.** A balanced inaction analysis.

✗ C Only when competitors wait — **Wrong.** Not the criterion.

✗ D Never — **Wrong.** Sometimes it is right.

Q33 **Answer: B** Review: M4 · 4.9

Which is a clear decision request?

✗ A "Support the AI programme." — **Wrong.** Vague.

✓ B "Approve Tranche 1, €120k of €400k, 60 agents for 6 months; Gate 2 criteria …; stop if acceptance < 45%; decide by 14 January." — **Correct.** Amount, scope, gate, stop rule, date.

✗ C "Approve AI." — **Wrong.** Vague.

✗ D "Note the update." — **Wrong.** Not a decision.

Q34 **Answer: D** Review: M4 · 4.10

What makes a null recommendation credible?

✗ (A) Saying no without explanation — **Wrong.** Not credible.

✗ (B) Delaying the memo — **Wrong.** Avoidance.

✗ (C) Blaming the vendor — **Wrong.** Unhelpful.

✓ (D) Driving numbers, what is valuable, what would change the answer, a better alternative and a revisit trigger — **Correct.** These elements preserve credibility.

Q35 **Answer: A** Review: M1–M2

Integrated: A pilot shows 40% savings; the hard landing line covers only overflow outsourcing; inference cost is unmodelled at scale. What should the case do?

✓ (A) Adjust the pilot, count only traceable hard benefit, add inference and other hidden costs, and present ranges with staging — **Correct.** Combines Modules 1 and 2.

✗ (B) Use 40% across all staff costs — **Wrong.** Unadjusted and untraceable.

✗ (C) Drop the case — **Wrong.** There may be value.

✗ (D) Add strategic value to close the gap — **Wrong.** Inflation.

Q36 **Answer: C** Review: M2 · 2.5, M3 · 3.11

Integrated: Year-1 targets in the measurement plan show full financial value, while the ROI model ramps to steady state in month 18. What is wrong?

✗ (A) Nothing — **Wrong.** The two must agree.

✗ (B) The model should drop the ramp — **Wrong.** The ramp reflects reality.

✓ (C) Targets and model disagree; year-1 targets should reflect the ramp — **Correct.** Alignment between Modules 2 and 3.

✗ (D) Year-2 targets should be removed — **Wrong.** No.

Q37 **Answer: A** Review: M4 · 4.6–4.9

Integrated: The conservative NPV is negative, the base positive, the readiness finding is "phase", and the roadmap places scale in H2. What decision request fits?

✓ (A) A Tranche 1 for the ready phase, with gate criteria aligned to the roadmap and a stop rule — **Correct.** Aligns finance, readiness and roadmap.

✗ (B) Defer indefinitely — **Wrong.** Not a decision.

✗ (C) Full funding now — **Wrong.** Ignores conservative case, readiness and roadmap.

✗ (D) Reject — **Wrong.** There is a viable staged path.

Q38 **Answer: C** Review: M1–M4

Integrated: Which sequence reflects this course?

✗ (A) Write memo → build model → define metrics — **Wrong.** Analysis comes before the memo.

✗ (B) Choose vendor → justify — **Wrong.** Solution-first.

✓ (C) Structure the case (drivers, traceability) → build the ROI model (TCO, NPV, sensitivity) → define value metrics and measurement → shape the decision-ready case — **Correct.** The course arc.

✗ (D) Metrics → vendor → memo — **Wrong.** Out of order.

R6 Further reading and next steps

Useful public companions to this course. Check for the latest versions.

Resource	Why it is useful
Corporate finance texts on NPV, IRR and capital budgeting	Background for the ROI model at advisory depth
Real-options literature applied to IT and R&D investment	Valuing staged investment and the option to stop
Benefits-realization management guidance (public-sector and project-management bodies)	Tracking and banking benefits
Introductory texts on quasi-experimental and A/B evaluation	Comparison designs and attribution
Your organization's finance policies (discount rate, capitalisation, investment governance)	The rules your case must follow

Practice assignments

1. Pick an AI initiative; build its driver tree (T1.2) and traceability register (T1.3); classify each benefit as hard, soft or option.
2. Rebuild the worked ROI example in a spreadsheet, then add three hidden costs (T2.3) and redo the tornado and breakeven (T2.5).
3. Write a measurement plan (T3.4) with baselines, paired metrics and a staggered-rollout design.
4. Draft a two-page IC memo (T4.1), apply the 60-second test, and write the null-recommendation version (T4.5) to compare.

ENTERPRISE AI TRANSFORMATION AND CONSULTING PROFESSIONAL

COURSE 6 · STUDY BOOK

AI Governance, Risk, and Executive Communication

How to govern enterprise AI so decisions are traceable and defensible: frameworks, standards and law; classification tied to concrete obligations; lifecycle risk management including agentic AI; controls a team can build — and human oversight that works at real volume.

2 MODULES · CASE STUDIES · TEMPLATES · CHEAT SHEETS · MOCK EXAM

Study edition · 2026

Includes full answer explanations

How to use this book

This study book accompanies **Course 6, AI Governance, Risk, and Executive Communication**, of the Enterprise AI Transformation and Consulting Professional certificate. It is written for consultants, risk and compliance professionals, internal auditors, governance leads and executives who must govern, assess and report on enterprise AI.

Each module follows the same rhythm so that you always know where you are:

Element	What you get	How to use it
Module opener	Learning objectives, section list, study time	Read first; return to it to self-test at the end
Overview & key takeaways	The module on one screen	Skim before, re-read after
Core sections	Concepts, tables, diagrams, worked examples	Study actively: annotate, restate in your own words
Callout boxes	Key ideas, examples, pitfalls, exam traps, consultant tips	Exam traps are the most frequently tested ideas
Case study	A realistic enterprise scenario with model answers	Answer the questions before reading the model answers
Templates	Fill-in consulting tools (T-codes), one per page or less	Copy into your own engagements
Do / Avoid	Best practices and pitfalls side by side	Use as a pre-meeting checklist
Cheat sheet	One-page module summary	Revise from these in the final week
Knowledge check	Scenario questions with explanations for **every** option	Understanding why the wrong answers are wrong matters as much as getting the right one

Callout key

KEY IDEA
The idea to remember if you remember nothing else.

EXAM TRAP
A frequently tested distinction or a tempting wrong answer.

EXAMPLE
A worked illustration of the concept.

CONSULTANT'S TIP
Practical wording or technique from the field.

PITFALL
A common mistake made in real engagements.

DEFINITION
A precise definition worth learning word for word.

> **NOTE · Builds on Courses 1–5**
> Earlier courses touched governance at many points: decision rights and scope of authority (Course 1), risk by use and portfolio risk classification (Course 2), oversight design and evidence grading (Course 3), agent permissions and control-load tests (Course 4), and verification plans (Course 5). This course brings them together into a governance and risk discipline. Regulatory content is stated as of **November 2026** and is **orientation, not legal advice**: formal determinations belong to qualified counsel.

Suggested study plan (four weeks)

Week	Study	Practise
1	Module 1 (sections 1.1–1.7)	Map the instruments that apply to one AI system; classify it and list its obligations (T1.1–T1.3)
2	Module 1 (sections 1.8–1.13)	Draft a data boundary matrix, a bias evaluation plan and a governance decision record
3	Module 2	Turn three risk categories into specific failure statements; specify controls; run the HITL volume test
4	Revise both cheat sheets and the glossary	Assemble a governance and risk plan; take the mock exam (60 min) and review every explanation

> **NOTE · About the examples**
> All organizations in case studies and examples are fictional, and their figures are illustrative. Regulatory and framework references are for orientation only; always check current law and guidance, and refer legal interpretation to qualified specialists.

Contents

FRONT **How to use this book** 2

MODULE 1 **AI Governance and Ethics** 6
What AI governance is 7
Frameworks, standards and law — by the questions they answer 8
Responsible AI policy 9
AI ethics for business leaders 9
AI risk classification 10
High-risk AI systems: obligations and timing 10
Connecting classification to concrete obligations 11
Human oversight controls — and whether they can operate 11
AI monitoring and escalation 12
Data boundary policy 12
Bias evaluation 13
Accountability and governance decision records 13
Communicating governance to executives 14
Case study: Rheinmark Kredit 15
Consulting templates 16
Best practices and pitfalls 19
Module 1 cheat sheet 20
Knowledge check 21
Knowledge check: answers and explanations 23

MODULE 2 **AI Risk Management** 25
Risk across the deployment lifecycle 26
The AI risk categories 27
Classifying under the EU AI Act and the NIST AI RMF 27
From broad categories to specific failures 28
Model risk and data risk 28
Regulatory risk 29
Third-party and AI vendor risk 29
Agentic and autonomous AI risk 30
Designing guardrails, approval gates and escalation paths a team can build 30
Testing human-in-the-loop at real volume 31
Audit readiness and monitoring 32
Reporting AI risk to executives 32
The AI governance and risk plan 32
Case study: Halvorsen Distribution 34
Consulting templates 36
Best practices and pitfalls 39
Module 2 cheat sheet 40
Knowledge check 41
Knowledge check: answers and explanations 43

BACK **Course at a glance: the master cheat sheet** 45
Template index 45
Glossary 46
Mock exam 48

Mock exam: answer key and explanations 54
Further reading and next steps 61

MODULE 1

AI Governance and Ethics

Practical governance for enterprise AI: telling frameworks, standards and law apart by the questions they answer; turning ethics into policy; tying risk classification to concrete obligations; testing oversight at real volume; and documenting decisions so they stay traceable and defensible.

YOU WILL BE ABLE TO

- Distinguish voluntary risk frameworks, certifiable management-system standards and binding law by the questions they answer
- Design a responsible AI policy and translate ethics principles into operating rules
- Classify AI systems by risk and connect each class to concrete governance obligations
- Describe high-risk AI system obligations for providers and deployers (EU AI Act, as amended)
- Specify human oversight controls and test whether they can operate at real volume
- Design monitoring, escalation and data boundary policies
- Plan a bias evaluation appropriate to the use
- Document governance decisions so reasoning, ownership, controls, monitoring, escalation and gaps are traceable

IN THIS MODULE

1. What AI governance is
2. Frameworks, standards and law
3. Responsible AI policy
4. Ethics for business leaders
5. Risk classification
6. High-risk AI systems
7. From classification to obligations
8. Human oversight controls
9. Monitoring and escalation
10. Data boundary policy
11. Bias evaluation
12. Accountability and governance decision records
13. Communicating governance to executives
14. Case study · templates · cheat sheet · knowledge check

Suggested study time · 6 hours

AI governance is the system of decision rights, policies, controls and evidence through which an organization makes sure its AI is used as intended, within the law and its own values — and can show that it is. This module teaches governance as a **practical discipline**: knowing which instrument answers which question, turning principles into rules people can follow, tying risk classes to concrete obligations, checking that human oversight can actually operate, and writing decisions down so they can be traced and defended later.

KEY TAKEAWAYS

1. Tell instruments apart by the question they answer: **voluntary frameworks** (how to think about risk), **certifiable management-system standards** (is there an auditable system?), **binding law** (what must we do?).
2. A **responsible AI policy** turns ethics principles into scope, roles, classification, permitted uses, data boundaries, oversight, incidents and exceptions.
3. **Risk classification** is only useful when it triggers **concrete obligations** with owners and evidence.
4. Under the EU AI Act as amended by the AI Omnibus (Regulation (EU) 2026/1744), stand-alone high-risk obligations apply from **2 December 2027** and product-embedded ones from **2 August 2028**; prohibitions and AI literacy already apply.
5. **Human oversight** must be tested at **real volume**; monitoring needs thresholds and a rehearsed **escalation ladder**.
6. **Data boundary policies** and **bias evaluations** must fit the use.
7. Every material governance decision needs a **decision record**: reasoning, owner, controls, monitoring, escalation, known gaps.

1.1 What AI governance is

Element	Question it answers	Example
Decision rights	Who may approve, change, pause or retire an AI system?	AI risk committee approves high-risk uses; system owner may pause
Policies and standards	What are the rules?	Responsible AI policy; data boundary standard
Risk classification	How much scrutiny does this use need?	Internal tiers mapped to legal classes
Controls	What prevents or detects harm?	Human review, access limits, monitoring thresholds
Oversight and assurance	Is it working?	Second-line review, internal audit, external certification
Evidence and records	Can we show it?	Inventory, decision records, logs, test results
Culture and literacy	Do people understand and follow it?	Role-based training, clear escalation

KEY IDEA

Governance is not a committee or a policy document. It is a working system: **rules → controls → evidence → decisions**, with named people at each step.

1.2 Frameworks, standards and law — by the questions they answer

They are complementary: law sets the floor, standards prove a system exists, frameworks guide how to think.

Figure 1.1 — Three kinds of governance instrument

Instrument	Type	What it gives you	Typical use
NIST AI Risk Management Framework (AI RMF 1.0) and Generative AI Profile	Voluntary framework	Functions Govern, Map, Measure, Manage; trustworthiness characteristics; generative-AI risk catalogue	Structuring risk thinking and programmes, inside or outside the US
OECD AI Principles	Intergovernmental principles	High-level values and a widely used AI-system definition	Policy anchoring; definitions
ISO/IEC 42001:2023	Certifiable management-system standard	Requirements for an AI management system (policy, roles, risk and impact assessment, controls, improvement)	Demonstrating a governed system to customers, regulators, auditors
ISO/IEC 23894:2023	Guidance standard (not certifiable)	How to apply risk management to AI, building on ISO 31000	Designing the risk process
EU AI Act — Regulation (EU) 2024/1689, amended by Regulation (EU) 2026/1744	Binding law	Prohibited practices, high-risk obligations, transparency duties, general-purpose model rules, AI literacy	Legal compliance in the EU market
GDPR, sector and labour law	Binding law	Data protection, automated decisions, sector rules, works-council rights	Constraints on data and use

EXAM TRAP

"We are ISO/IEC 42001 certified, so we comply with the EU AI Act" is wrong. Certification shows a management system operates; it does not by itself establish that a specific system meets specific legal obligations. Equally, "NIST AI RMF compliance" is a misnomer — the framework is voluntary and has no conformity regime.

1.3 Responsible AI policy

Policy component	Content	Operability test
Scope and definitions	What counts as AI; which systems and users are covered (incl. embedded and personal tools)	Can a manager tell whether a tool is in scope?
Principles	Short list of values with what each means here	Is each principle linked to at least one rule?
Roles and decision rights	Accountable executive, committee, system owners, overseers, second and third lines	Is there one named role per decision?
Risk classification	Tiers and triggers; mapping to legal classes	Can two people classify the same use the same way?
Permitted, restricted and prohibited uses	Examples per category	Can staff answer scenario questions correctly?
Data boundaries	Which data may go into which tools (Section 1.10)	Is it enforced technically where possible?
Human oversight and transparency	When review is required; disclosure to affected people	Is review capacity planned?
Third-party AI	Vendor assessment, contract terms, embedded features	Do procurement and IT apply it?
Monitoring, incidents and escalation	Thresholds, escalation ladder, reporting	Has it been rehearsed?
Training and literacy	Role-based measures	Are records kept?
Exceptions and review	Who may grant exceptions, how recorded; policy review cycle	Are exceptions logged and time-limited?

1.4 AI ethics for business leaders

Ethics becomes useful in business when principles turn into **questions people ask before deciding**, and into **rules** where the answer should not depend on who is asking.

Principle	Operational question for leaders
Fairness	Could this system treat groups of people differently in ways we could not justify to them?
Transparency	Will people affected know AI is involved, and can we explain outcomes to the degree the decision requires?
Accountability	Who answers for this system's outcomes — by name?
Privacy	Are we using personal data only as people would reasonably expect and the law permits?
Safety and reliability	What is the worst plausible failure, and what stops it?
Human agency	Can people contest, override or opt out where it matters?
Proportionality	Is AI necessary for this purpose, or are we using it because we can?

NOTE · Ethics and compliance

Compliance asks “is it allowed?”; ethics asks “is it right, and would we be comfortable defending it publicly?” Many reputational failures were legal. Use an **ethics review trigger** for novel uses, uses affecting vulnerable people, or uses where the organization would hesitate to explain the system openly.

1.5 AI risk classification

EU AI Act class	Meaning (orientation)	Examples
Prohibited practices	Uses banned outright	E.g., certain manipulative techniques, social scoring, emotion recognition in workplaces and education (with narrow exceptions)
High-risk	Annex III uses, or AI that is a safety component of products under listed EU product law (Annex I)	Recruitment and worker management; creditworthiness; risk assessment and pricing in life and health insurance; access to education; essential public services; certain critical infrastructure
Transparency obligations	Users or affected people must be informed	Chatbots interacting with people; AI-generated or manipulated content
Minimal risk	No specific obligations beyond general law (and AI literacy)	Spam filters, internal drafting aids
General-purpose AI models	Obligations on model providers	Foundation model documentation, copyright policy; additional duties for systemic-risk models

BEST PRACTICE · Internal tiers mapped to legal classes

Most organizations use an internal scale (e.g., Tier 1 low · Tier 2 moderate · Tier 3 high · Tier 4 prohibited/stop) that combines legal class with use-based risk factors (Course 2): who is affected, consequence, reversibility, autonomy, scale. The legal class sets the **minimum**; internal tiers can be stricter.

1.6 High-risk AI systems: obligations and timing

Role	Key obligations (orientation)
Provider (develops / places on market)	Risk management system across the lifecycle · data and data governance · technical documentation · automatic logging · transparency and instructions for use · human-oversight design · accuracy, robustness and cybersecurity · quality management system · conformity assessment, registration · post-market monitoring and serious-incident reporting
Deployer (uses under own authority)	Use according to instructions · assign human oversight to people with competence, training and authority · ensure input data under its control is relevant and sufficiently representative · monitor operation and inform provider of risks or serious incidents · keep automatically generated logs under its control (at least six months unless other law says otherwise) · inform workers’ representatives and affected workers before workplace use · inform people subject to certain decisions · fundamental-rights impact assessment for public bodies and certain private deployers (e.g., credit scoring, life and health insurance pricing)

Timeline (as amended by Regulation (EU) 2026/1744)	Date
Prohibited practices; AI literacy (Article 4)	Applying since 2 February 2025

Timeline (as amended by Regulation (EU) 2026/1744)	Date
General-purpose AI model obligations	Applying since 2 August 2025
Article 50 transparency obligations	2 August 2026 (grace period to 2 December 2026 for marking content from generative systems already on the market)
High-risk — stand-alone Annex III systems	2 December 2027 (deferred from 2 August 2026)
High-risk — AI in products under Annex I	2 August 2028 (deferred from 2 August 2027)

NOTE · AI literacy after the Omnibus

The amended Article 4 requires providers and deployers to **take measures to support** AI literacy among staff and others operating AI on their behalf; it no longer requires them to guarantee a specific level for each individual. It remains an obligation — an obligation of effort, not of result. Keep training records as evidence of the measures taken.

PITFALL · Reading "deferred" as "paused"

The deferral of high-risk obligations gives time to build governance properly; it does not pause prohibitions, transparency duties, AI literacy, GDPR or sector rules. Organizations that stop preparing will face the same work under more time pressure in 2027.

1.7 Connecting classification to concrete obligations

A classification that does not change what anyone does is decoration. For each class, define the obligations it triggers, who owns each, and the evidence that shows it is met.

Internal tier	Triggered obligations	Owner	Evidence
Tier 1 (minimal)	Inventory entry; acceptable-use rules; data boundary check	System owner	Register entry
Tier 2 (moderate / transparency)	+ risk assessment; disclosure to users; basic testing; monitoring	System owner + risk	Assessment record; disclosure text; test results
Tier 3 (high)	+ oversight specification; staffing and competence; logging and retention; bias evaluation; DPIA / FRIA where required; incident plan; committee approval	System owner, risk, DPO, legal, committee	Spec, rosters, training records, logs, evaluation, assessments, minutes
Tier 4 (stop)	No deployment; legal review	Legal / accountable executive	Decision record

1.8 Human oversight controls — and whether they can operate

Oversight design choice	Use when
Review every output before it takes effect	High consequence, irreversible, affects individuals' rights
Risk-based review (flagged cases only)	High volume; reliable flags for uncertainty or risk

Oversight design choice	Use when
Sampling review	Lower consequence; quality monitoring
Approval thresholds for actions	Agents acting in systems (Course 4)
Ability to stop	Always — named person, tested procedure

Review hours needed = volume × review rate × minutes per review ÷ 60 · compare with hours actually available to qualified reviewers

EXAMPLE

A credit-decision assistant handles 40,000 applications a month; policy requires review of every adverse recommendation (30% = 12,000) at 4 minutes = **800 hours** ≈ 5 FTE. If two reviewers are staffed (≈ 320 hours), each review gets ~1.6 minutes — the control exists on paper only. Module 2 adds seeded-error testing to measure whether reviewers actually catch errors.

1.9 AI monitoring and escalation

Monitor	Example metric	Threshold example
Performance	Accuracy vs reference sample; acceptance rate	Drop > 5 pts vs baseline over 2 weeks
Drift	Input distribution shift; output mix shift	Population stability beyond agreed limit
Fairness	Outcome / error-rate gaps across groups	Gap above agreed tolerance
Oversight health	Review time, override and edit rates	Median review time below minimum
Incidents and complaints	Count, severity	Any serious incident
Usage and scope	Use outside intended purpose; shadow use	Any out-of-scope use
Cost and capacity	Inference cost, queue lengths	Above budget or SLA

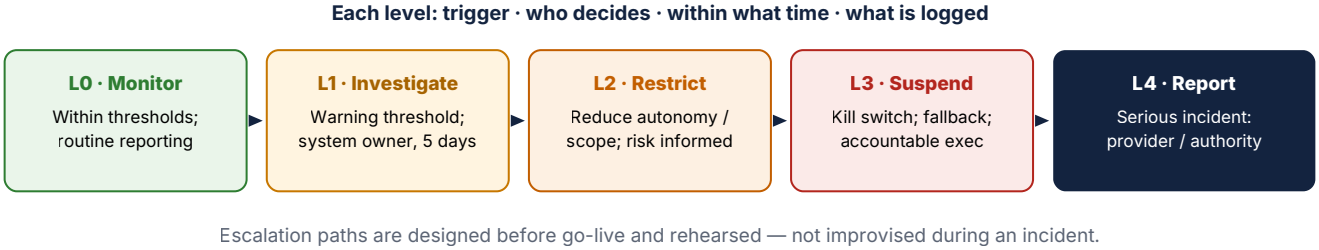


Figure 1.2 — An escalation ladder

1.10 Data boundary policy

A data boundary policy states which categories of data may be used with which kinds of AI tool, for what purposes, under what controls. It is the most frequently consulted part of any AI policy — and most effective when enforced technically (data-loss prevention, approved tools, permission-aware retrieval).

Data class ↓ / Tool type →	Public AI tools (consumer)	Enterprise-approved assistant	Approved AI system for a specific use
Public	Allowed	Allowed	Allowed
Internal	Not allowed	Allowed	Allowed
Confidential (customer, contracts, financials)	Not allowed	Allowed with conditions (no training on data; EU hosting; logging)	Allowed per approval
Restricted / special category (health, biometrics, HR cases)	Not allowed	Not allowed by default	Only with DPIA, explicit approval and enhanced controls

CONSULTANT'S TIP

Cover **outputs** too: AI outputs derived from confidential inputs inherit their classification; retention of prompts and outputs follows the strictest input class.

1.11 Bias evaluation

Step	Content
1 · Define the decision and harm	What outcome matters to people, and what harm would unfair treatment cause?
2 · Define groups	Relevant protected or vulnerable groups; legal basis for using group data; proxies where data is unavailable
3 · Choose metrics that fit the use	Selection-rate differences; error-rate parity (false positives / negatives by group); calibration; outcome quality for generative outputs
4 · Set tolerances in advance	Agreed thresholds, with reasoning
5 · Test on representative data	Including intersectional and small-group analysis where possible
6 · Mitigate and re-test	Data, model, threshold or process changes; human review for affected groups
7 · Document and monitor	Record choices, results, residual gaps; monitor in production

KEY IDEA

Fairness metrics can conflict — a system cannot generally satisfy all of them at once. The choice of metric is a **governance decision** tied to the use and its legal context, and should be recorded with its reasoning and reviewed with legal and domain experts.

1.12 Accountability and governance decision records

Months later, an auditor, regulator, court or journalist may ask: *why did you allow this system to do this?* A governance decision record answers that question with what was known and decided at the time.

Record field	Content
Decision	What was decided (approve, approve with conditions, reject, change, retire)
System and use	Inventory ID; intended purpose; scope

Record field	Content
Classification and basis	Legal class and internal tier, with reasoning and standard/date (Course 3)
Reasoning	Options considered; why this one; trade-offs accepted
Ownership	Accountable owner; decision-holder; consulted parties
Controls	Oversight design, limits, data boundaries, approvals
Monitoring and escalation	Metrics, thresholds, escalation ladder
Known gaps and residual risk	What is not yet controlled; who accepted it; until when
Review date	When the decision will be revisited

PITFALL · Minutes are not records

Committee minutes stating “approved after discussion” do not show reasoning, controls or accepted gaps. Attach a decision record to every material approval.

1.13 Communicating governance to executives

Executives need to know...	Format that works
Where AI is used and how much of it is high-risk	Inventory summary by tier and unit
Whether controls work	A few oversight-health and incident indicators against thresholds
What decisions need them	Decision requests with records attached
What gaps they are accepting	Explicit residual-risk list with owners and dates
What is changing externally	Short regulatory outlook with dates (e.g., Dec 2027 high-risk deadline)

CONSULTANT'S TIP

Lead with decisions and accepted gaps, not with framework diagrams. An executive who reads “we are accepting X until Y, owned by Z” understands governance better than one who reads a maturity score.

1.14 Case study: Rheinmark Kredit

CASE STUDY **Governing a credit-decision assistant**

Rheinmark Kredit (fictional) · consumer lender in Germany and Austria

Situation

Rheinmark wants to deploy a vendor AI system that scores consumer-loan applications and drafts recommendations for credit officers. The CRO asks for a governance design before the board approves.

Governance design

- **Instruments:** EU AI Act (creditworthiness evaluation of natural persons is an Annex III high-risk use; Rheinmark is a deployer, the vendor the provider); GDPR (automated decision-making and profiling questions referred to the DPO and counsel); banking supervisory expectations on model risk; ISO/IEC 42001 chosen as the management-system frame; NIST AI RMF used to structure risk identification.
- **Classification:** Tier 3 (high). Obligations: oversight specification, competent overseers, input-data checks, logging and retention, monitoring, incident reporting to the vendor, information to applicants, fundamental-rights impact assessment, bias evaluation, committee approval. Legal applicability date for the high-risk regime noted (2 December 2027), but the board decided to meet the obligations at go-live.
- **Oversight at volume:** 40,000 applications/month; review of all adverse recommendations needs ~800 hours/month; plan revised from 2 to 5 FTE trained overseers.
- **Data boundary:** applicant data restricted to the approved system; no use in general assistants.
- **Bias evaluation:** approval-rate and error-rate comparisons across age and gender groups; postcode tested as a potential proxy.
- **Decision record:** approved with conditions; residual gap — vendor explanation quality for declined applicants — accepted by the CRO until the vendor's next release, review in 6 months.

Discussion questions

1. Which question does each instrument answer for Rheinmark?
2. Why did the board meet high-risk obligations at go-live although they apply later?
3. What would have happened to oversight with two reviewers?
4. Why is the accepted residual gap recorded with an owner and date?

Model answers (brief)

1. The AI Act and GDPR: what Rheinmark must do as deployer. ISO/IEC 42001: whether it runs an auditable management system. NIST AI RMF: how to structure risk thinking.
2. Credit decisions affect people now. Building controls once avoids retrofitting, and supervisory and reputational expectations apply regardless of the legal date.
3. About 1.6 minutes per adverse case — rubber-stamping, so the control would exist only on paper.
4. So the decision stays traceable and defensible: someone accepted a known gap, for a reason, until a defined review.

1.15 Consulting templates

T1.1 Governance Instrument Map

Purpose: Identify which instruments apply and which question each answers.

AI system / use: .....

Instrument	Type (framework / standard / law)	Question it answers here	Applies? (basis)	Owner	Specialist confirmation
EU AI Act (as amended)					
GDPR					
Sector law / supervisory expectations					
Labour law / works council					
ISO/IEC 42001					
NIST AI RMF					
Internal AI policy					

T1.2 Responsible AI Policy Checklist

Purpose: Check that the policy is complete and operable.

- ☐ Scope and definitions cover built, bought, embedded and personal tools
- ☐ Principles each linked to at least one rule
- ☐ One named role per decision right
- ☐ Risk tiers with triggers and mapping to legal classes
- ☐ Permitted, restricted and prohibited uses with examples
- ☐ Data boundary matrix (incl. outputs and retention)
- ☐ Human oversight and transparency rules, with capacity planning
- ☐ Third-party AI rules applied by procurement and IT
- ☐ Monitoring thresholds, escalation ladder, incident reporting — rehearsed
- ☐ Role-based AI literacy measures and records
- ☐ Exception process: who, how logged, time-limited; policy review cycle

T1.3 Risk Classification & Obligations Map

Purpose: Tie classification to concrete obligations, owners and evidence.

System / intended purpose / role (provider or deployer): .....

Legal class and basis: .....

Internal tier and use-based factors: .....

Obligation	Applies from	Owner	Evidence	Status

T1.4 Data Boundary Matrix

Purpose: State which data may be used with which AI tools.

Data class	Public AI tools	Enterprise assistant	Approved specific system	Conditions / technical enforcement
Public				
Internal				
Confidential				
Restricted / special category				

Rules for AI outputs and retention of prompts / outputs: .....

Owner and review date: .....

T1.5 Bias Evaluation Plan

Purpose: Plan a fairness evaluation that fits the use.

Decision and potential harm:

Groups (and legal basis for group data / proxies):

Metrics chosen and why:

Tolerances agreed (before testing):

Test data and representativeness:

Mitigation options:

Production monitoring:

Reviewed by (legal, domain, risk):

T1.6 Governance Decision Record

Purpose: Keep every material decision traceable and defensible.

Decision and date:

System, inventory ID, intended purpose:

Classification (legal / internal) and basis, standard and date:

Options considered and reasoning:

Accountable owner / decision-holder / consulted:

Controls (oversight, limits, data boundaries, approvals):

Monitoring metrics, thresholds and escalation ladder:

Known gaps and residual risk — accepted by whom, until when:

Review date:

T1.7 Monitoring & Escalation Plan

Purpose: Define what is watched and what happens when thresholds are crossed.

MONITORING

Area	Metric	Warning threshold	Action threshold	Owner	Frequency
Performance					
Drift					
Fairness					
Oversight health					
Incidents / complaints					
Scope / usage					

ESCALATION LADDER

Level	Trigger	Who decides	Time limit	Logged where
L1 Investigate				
L2 Restrict				
L3 Suspend				
L4 Report				

1.16 Best practices and pitfalls

✓ DO THIS

- Name the question each instrument answers.
- Write policies people can apply to scenarios.
- Map internal tiers to legal classes as a minimum.
- Attach owners and evidence to every obligation.
- Test oversight capacity at real volume.
- Rehearse escalation before go-live.
- Record decisions with reasoning and accepted gaps.

× AVOID THIS

- Treating a certificate as legal compliance.
- Principles with no rules attached.
- Classification that changes nothing.
- Reading “deferred” as “paused”.
- Assuming review is free.
- Improvising escalation during incidents.
- “Approved after discussion” as the only record.

1.17 Module 1 cheat sheet

MODULE 1 · ONE-PAGE SUMMARY

AI Governance and Ethics

GOVERNANCE =

Decision rights · policies · classification · controls · oversight & assurance · evidence · literacy. Rules → controls → evidence → decisions.

INSTRUMENTS BY QUESTION

Framework (NIST AI RMF): how to think? · **Standard** (ISO/IEC 42001 certifiable; 23894 guidance): auditable system? · **Law** (EU AI Act, GDPR): what must we do?

POLICY COMPONENTS

Scope · principles→rules · roles · tiers · uses · data boundaries · oversight & transparency · third-party · monitoring & incidents · literacy · exceptions.

ETHICS → QUESTIONS

Fairness · transparency · accountability · privacy · safety · human agency · proportionality. Ethics review triggers.

EU AI ACT (AS AMENDED BY REG. (EU) 2026/1744)

Prohibitions & AI literacy since 2 Feb 2025 (literacy now "take measures to support") · GPAI since 2 Aug 2025 · transparency 2 Aug 2026 · high-risk Annex III **2 Dec 2027** · Annex I **2 Aug 2028**. Deployer duties: use per instructions, competent oversight, input data, monitoring, logs ≥ 6 months, inform workers and affected people, FRIA for certain deployers.

CLASSIFICATION → OBLIGATIONS

Each tier triggers named obligations with owner and evidence.

OVERSIGHT AT VOLUME

Hours = volume × rate × minutes ÷ 60 vs qualified hours available.

MONITORING & ESCALATION

Performance · drift · fairness · oversight health · incidents · scope · cost. L1 investigate → L2 restrict → L3 suspend → L4 report.

DATA BOUNDARIES & BIAS

Data class × tool type, incl. outputs. Bias: harm · groups · fit metrics · tolerances first · test · mitigate · document.

DECISION RECORDS & EXECUTIVES

Decision · system · classification & basis · reasoning · owners · controls · monitoring & escalation · **known gaps accepted by whom until when** · review date. Executives: tier summary, control health, decisions, accepted gaps, regulatory outlook.

1.18 Knowledge check

Answer all questions before turning the page. Explanations for every option follow.

K1 **Which question does ISO/IEC 42001 primarily answer?**

- (A) What are we legally required to do in the EU?
- (B) Do we operate an AI management system that an accredited auditor can certify?
- (C) How should we conceptually frame AI risk?
- (D) Which models are most accurate?

K2 **A vendor claims "we are NIST AI RMF compliant". How should this be read?**

- (A) As a loosely used phrase: the AI RMF is a voluntary framework without a conformity regime, so ask what practices are actually in place
- (B) As proof of fairness
- (C) As ISO certification
- (D) As legal compliance in the EU

K3 **After the AI Omnibus (Regulation (EU) 2026/1744), when do obligations for stand-alone Annex III high-risk systems apply?**

- (A) 2 August 2028
- (B) 2 August 2026
- (C) They were repealed
- (D) 2 December 2027

K4 **What does the amended AI literacy duty (Article 4) require?**

- (A) Every employee must pass a certified exam
- (B) Only providers must train their own staff
- (C) Providers and deployers must take measures to support AI literacy, without guaranteeing a specific level for each individual
- (D) Nothing — it was removed

K5 **Which use is listed among Annex III high-risk areas?**

- (A) Internal meeting summaries
- (B) Evaluating the creditworthiness of natural persons
- (C) Product-description drafting
- (D) Spam filtering

K6 **A classification scheme labels systems but triggers no different controls. What is wrong?**

- (A) Classification must trigger concrete obligations with owners and evidence
- (B) It needs more tiers
- (C) Nothing — labels aid awareness
- (D) It should use colours

K7 **40,000 decisions a month; 30% require review at 4 minutes; 320 reviewer hours are available. What is the conclusion?**

- (A) Review should be removed
- (B) Oversight is adequate
- (C) Review time should be cut to fit
- (D) Oversight cannot operate as designed; reviewers would have ~1.6 minutes per case

K8 **Which belongs in a governance decision record but is often missing from committee minutes?**

- (A) Agenda order
- (B) Meeting room
- (C) Known gaps and residual risk, who accepted them and until when
- (D) Attendees

1.19 Knowledge check: answers and explanations

K1 **Answer: B** Review: 1.2

Which question does ISO/IEC 42001 primarily answer?

- ✗ (A) What are we legally required to do in the EU? — **Wrong.** That is the role of binding law.
- ✓ (B) Do we operate an AI management system that an accredited auditor can certify? — **Correct.** 42001 is a certifiable management-system standard.
- ✗ (C) How should we conceptually frame AI risk? — **Wrong.** That is closer to a voluntary framework such as NIST AI RMF.
- ✗ (D) Which models are most accurate? — **Wrong.** Not addressed by the standard.

K2 **Answer: A** Review: 1.2

A vendor claims "we are NIST AI RMF compliant". How should this be read?

- ✓ (A) As a loosely used phrase: the AI RMF is a voluntary framework without a conformity regime, so ask what practices are actually in place — **Correct.** There is no formal NIST AI RMF compliance.
- ✗ (B) As proof of fairness — **Wrong.** No.
- ✗ (C) As ISO certification — **Wrong.** A different instrument.
- ✗ (D) As legal compliance in the EU — **Wrong.** NIST is not EU law.

K3 **Answer: D** Review: 1.6

After the AI Omnibus (Regulation (EU) 2026/1744), when do obligations for stand-alone Annex III high-risk systems apply?

- ✗ (A) 2 August 2028 — **Wrong.** That date applies to AI in products under Annex I.
- ✗ (B) 2 August 2026 — **Wrong.** This was the original date, now deferred.
- ✗ (C) They were repealed — **Wrong.** They were deferred, not repealed.
- ✓ (D) 2 December 2027 — **Correct.** Deferred by the Omnibus.

K4 **Answer: C** Review: 1.6

What does the amended AI literacy duty (Article 4) require?

- ✗ (A) Every employee must pass a certified exam — **Wrong.** Not required.
- ✗ (B) Only providers must train their own staff — **Wrong.** Deployers are covered too.
- ✓ (C) Providers and deployers must take measures to support AI literacy, without guaranteeing a specific level for each individual — **Correct.** An obligation of effort, not of result.
- ✗ (D) Nothing — it was removed — **Wrong.** It remains.

K5 **Answer: B** Review: 1.5

Which use is listed among Annex III high-risk areas?

- ✗ (A) Internal meeting summaries — **Wrong.** Minimal risk.
- ✓ (B) Evaluating the creditworthiness of natural persons — **Correct.** An Annex III use.
- ✗ (C) Product-description drafting — **Wrong.** Minimal risk.
- ✗ (D) Spam filtering — **Wrong.** Minimal risk.

K6 **Answer: A** Review: 1.7

A classification scheme labels systems but triggers no different controls. What is wrong?

✓ A Classification must trigger concrete obligations with owners and evidence — **Correct.** Otherwise it is decoration.

✗ B It needs more tiers — **Wrong.** More tiers do not fix the issue.

✗ C Nothing — labels aid awareness — **Wrong.** Classification is useful only when it changes obligations.

✗ D It should use colours — **Wrong.** Cosmetic.

K7 **Answer: D** Review: 1.8

40,000 decisions a month; 30% require review at 4 minutes; 320 reviewer hours are available. What is the conclusion?

✗ A Review should be removed — **Wrong.** Removal increases risk.

✗ B Oversight is adequate — **Wrong.** 800 hours are needed.

✗ C Review time should be cut to fit — **Wrong.** That produces rubber-stamping.

✓ D Oversight cannot operate as designed; reviewers would have ~1.6 minutes per case — **Correct.** Capacity falls far short.

K8 **Answer: C** Review: 1.12

Which belongs in a governance decision record but is often missing from committee minutes?

✗ A Agenda order — **Wrong.** Irrelevant.

✗ B Meeting room — **Wrong.** Irrelevant.

✓ C Known gaps and residual risk, who accepted them and until when — **Correct.** Key to traceability and defensibility.

✗ D Attendees — **Wrong.** Usually in minutes.

MODULE 2

AI Risk Management

Identifying and mitigating AI risk across the lifecycle — model, data, regulatory, third-party and agentic risk — turning broad categories into specific failures, designing controls a team can build, and testing whether human-in-the-loop oversight works at real volume.

YOU WILL BE ABLE TO

- → Manage AI risk across the full deployment lifecycle
- → Assess model, data, regulatory, third-party and vendor risk
- → Identify the distinctive risks of agentic and autonomous systems that act without review
- → Classify systems under the EU AI Act and structure risk work with the NIST AI RMF
- → Turn broad risk categories into specific failure statements
- → Specify guardrails, approval gates and escalation paths a team can build
- → Test whether human-in-the-loop oversight works at real volume, including seeded-error tests
- → Assemble an AI governance and risk plan that supports audit readiness and monitoring

IN THIS MODULE

1. Risk across the lifecycle
2. The AI risk categories
3. Classifying under the EU AI Act and NIST AI RMF
4. From categories to specific failures
5. Model and data risk
6. Regulatory risk
7. Third-party and vendor risk
8. Agentic and autonomous AI risk
9. Guardrails, gates and escalation a team can build
10. Testing human-in-the-loop at real volume
11. Audit readiness and monitoring
12. Reporting risk to executives
13. The AI governance and risk plan
14. Case study · templates · cheat sheet · knowledge check

Suggested study time · 6 hours

AI risk management is the work of finding how AI systems can fail or cause harm, deciding which risks to reduce, and building controls that actually work — then proving they work. This module covers the lifecycle and the major risk categories, gives agentic systems the special attention they need, insists that risks be written as **specific failures**, shows how to specify controls a team can build, and tests the control most plans leave unexamined: **whether a human reviewer can actually perform oversight at real volume**.

KEY TAKEAWAYS

1. Manage risk across the **whole lifecycle**: design, data, build/buy, test, deploy, operate, change, retire.
2. Cover **model, data, regulatory, third-party/vendor, security, operational** and **agentic** risk.
3. Use the **EU AI Act** for legal classification and the **NIST AI RMF** (and its Generative AI Profile) to structure identification and measurement.
4. Write risks as **specific failure statements**: system · does what wrong · under which condition · to whom · consequence.
5. Agentic systems that **act without review** need least privilege, action limits, approval gates, kill switches and logs.
6. Specify controls as **buildable**: trigger, check, action, owner, log, test.
7. Test human-in-the-loop for **capacity** and **effectiveness** (seeded-error tests) before relying on it.
8. Assemble an **AI governance and risk plan** that is audit-ready and monitored.

2.1 Risk across the deployment lifecycle

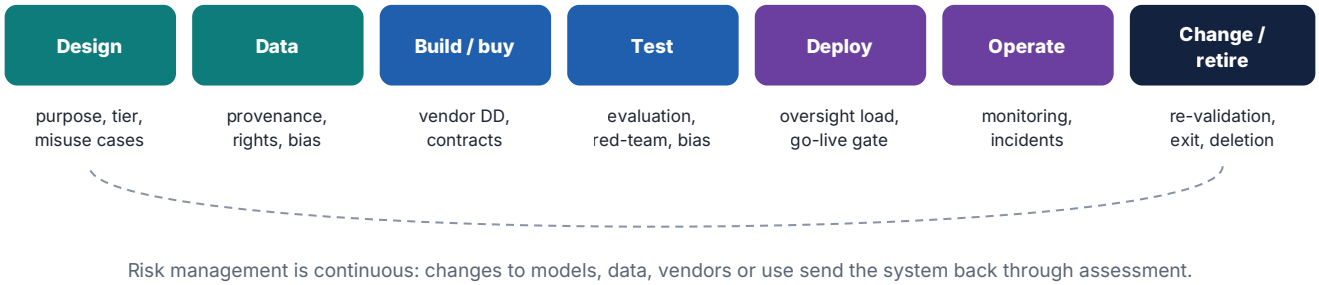


Figure 2.1 — Risk activities across the lifecycle

Stage	Key risk activities	Gate evidence
Design	Intended purpose, foreseeable misuse, classification, affected people	Classification record; risk assessment
Data	Provenance and rights, quality, representativeness, privacy, bias sources	Data readiness and lineage records (Course 3)
Build / buy	Vendor due diligence, contract terms, architecture and security	DD file; contract clauses
Test	Evaluation on representative data, robustness, red-teaming, bias evaluation	Test report with thresholds

Stage	Key risk activities	Gate evidence
Deploy	Oversight capacity and effectiveness, go-live criteria, rollback	HITL tests; go-live decision record
Operate	Monitoring, incidents, complaints, drift, usage outside purpose	Monitoring reports; incident log
Change / retire	Re-assessment after material change; exit and data deletion	Change records; decommission record

2.2 The AI risk categories

Category	What can go wrong	Typical controls
Model risk	Inaccuracy, poor generalisation, drift, hallucination, opaque reasoning, instability after updates	Validation, thresholds, monitoring, re-validation, human review
Data risk	Poor quality, unrepresentative data, missing rights, leakage of personal or confidential data, poisoning	Data governance, lineage, access control, DLP, provenance checks
Regulatory risk	Prohibited use, unmet high-risk or transparency duties, GDPR breaches, sector-rule breaches	Classification, obligation mapping, legal review, records
Third-party / vendor risk	Opaque models, silent model changes, sub-processors, lock-in, vendor failure	Due diligence, contract clauses, change notification, exit plans
Security risk	Prompt injection, data exfiltration, model abuse, insecure integrations	Threat modelling, red-teaming, input/output filtering, least privilege
Operational / people risk	Rubber-stamping, misuse, skills gaps, unmanaged shadow AI	Oversight design, training, policy, discovery
Agentic risk	Unreviewed actions, compounding errors, tool misuse, runaway costs, irreversible effects	Section 2.8

2.3 Classifying under the EU AI Act and the NIST AI RMF

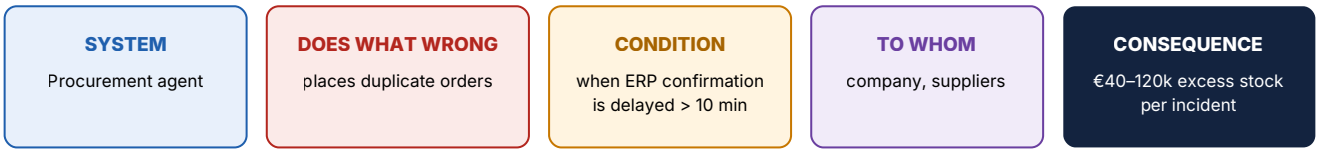
	EU AI Act	NIST AI RMF 1.0
Nature	Binding law (EU market)	Voluntary framework
Classification logic	Legal categories by intended purpose: prohibited, high-risk (Annex I / III), transparency, minimal; general-purpose model rules	Context-based: the Map function establishes purpose, context, impacts and risk tolerance; no fixed legal tiers
Output	Which legal obligations apply and from when	A structured view of risks and their measurement and management
Generative AI	Transparency duties; GPAI-model obligations on providers	Generative AI Profile (NIST AI 600-1) lists risks such as confabulation, data privacy, information integrity, harmful bias, IP and information security
Use together	Legal floor and deadlines	Method to identify, measure and manage risks beyond the legal minimum

KEY IDEA

Use the Act to answer **"what must we do and by when?"** and the RMF to answer **"what could go wrong here, and how do we measure and manage it?"**. A system can be legally minimal-risk and still carry material business risk that the RMF will surface.

2.4 From broad categories to specific failures

"Bias risk", "hallucination risk" and "agentic risk" cannot be controlled, tested or owned. A **specific failure statement** can.



"Agentic risk" cannot be controlled. "Duplicate orders when confirmation is delayed" can — with a specific control.

Figure 2.2 — Anatomy of a failure statement

Broad category	Specific failure statement
Bias risk	"The credit model declines applicants aged over 65 at twice the rate of comparable younger applicants because income from pensions is under-represented in training data — affecting ~800 applicants a year."
Hallucination risk	"The policy assistant cites a superseded cancellation rule when retrieval returns an archived document; customers receive wrong advice; complaint and redress cost."
Vendor risk	"The vendor updates its underlying model without notice; extraction accuracy on German invoices falls below 90%, undetected for weeks because we monitor volume, not accuracy."
Oversight risk	"At peak season, reviewers approve adverse recommendations in under a minute; unsupported declines reach applicants."

CONSULTANT'S TIP

Run a short workshop per system: for each category, ask "how exactly would this fail here?" until each answer names a condition and a consequence. Then rate likelihood and impact on the specific failure, not the category.

2.5 Model risk and data risk

Model-risk practice	What it involves
Independent validation	Someone other than the builders tests performance, assumptions and limits before use
Performance thresholds	Agreed minimum metrics on representative data; go/no-go
Robustness and red-teaming	Unusual, adversarial and edge-case inputs
Change control and re-validation	Material changes (incl. vendor model updates, prompt changes) trigger re-testing
Monitoring	Accuracy samples, drift, override rates in production

Model-risk practice	What it involves
Documentation	Purpose, limits, data, tests, known weaknesses

Data-risk practice	What it involves
Provenance and rights	Origin, licences, consent and purpose compatibility
Quality and representativeness	Profiles, fidelity samples, group coverage (Course 3)
Privacy and leakage	Data boundaries, DLP, retention, no vendor training on data
Integrity	Protection against poisoning of training or retrieval content

> **NOTE**
> Regulated sectors often already have model-risk management rules (for example, supervisory expectations for banks). Extend them to AI rather than creating a parallel process.

2.6 Regulatory risk

Regulatory risk source	Practical control
Misclassification (e.g., missing an Annex III use)	Classification record with legal review for borderline cases
Unmet deployer obligations for high-risk systems	Obligation map with owners and evidence (Module 1)
Transparency duties (chatbots, generated content)	Disclosure standards; content marking checks
Prohibited practices	Stop tier; screening at intake
GDPR (legal basis, automated decisions, DPIA)	DPO involvement; DPIA triggers
Changing timelines and guidance	Regulatory watch with dated baseline (e.g., Omnibus deferrals to Dec 2027 / Aug 2028)

2.7 Third-party and AI vendor risk

Risk	Due-diligence question	Contract / control
Opaque model and data	What model, trained on what, with what limits?	Documentation and instructions for use
Silent model changes	How are changes notified and tested?	Change-notification duty; regression testing rights; version pinning where possible
Data use	Is our data used for training? Where is it processed?	No-training clause; data residency; sub-processor list
Security	Testing, certifications, incident handling?	Security annex; incident notification
Fourth parties	Which upstream model providers?	Disclosure and flow-down terms
Viability and lock-in	Runway, ownership, exit options? (Course 1)	Exit assistance, data export, termination rights
Regulatory roles	Who is provider, who deployer, for which obligations?	Allocation of AI Act responsibilities; cooperation clauses

2.8 Agentic and autonomous AI risk

Agentic systems plan and take actions through tools, sometimes without a human reviewing each step. The risk profile changes: errors become **actions**, and actions can compound, spread and be hard to undo.

Distinctive agentic risk	Example	Control
Compounding errors	A wrong first step leads to a chain of wrong actions	Step limits; checkpoints; validation between steps
Tool misuse and permission creep	Agent uses a write permission it did not need	Least privilege; scoped credentials; allow-lists of actions
Prompt injection via content	Instructions hidden in an e-mail or web page redirect the agent	Treat content as data; input filtering; restrict actions after reading untrusted content
Irreversible actions	Payments, deletions, external messages	Approval gates for irreversible or high-value actions
Runaway loops and costs	Agent retries endlessly, consuming budget	Rate and spend limits; timeouts
Accountability gaps	Unclear who answers for an agent's action	Named owner; full action logs; attribution of actions to the agent identity
Silent scope expansion	Agent handles cases outside its intended purpose	Scope checks; escalation on out-of-scope inputs

EXAM TRAP

"A human can always check the agent's log afterwards" is not oversight for irreversible actions. If the action cannot be undone, the control must act **before** it — an approval gate or a hard limit.

2.9 Designing guardrails, approval gates and escalation paths a team can build

A control a team can build is **specific** (what exactly is checked), **triggered** (when), **actionable** (what happens), **owned** (by whom), **logged** and **testable**.

Vague control	Buildable specification
"Human oversight of payments."	Trigger: any agent-initiated payment > €500 or to a new supplier. Check: finance approver sees invoice, PO and supplier record. Action: approve / reject within 4 h; else auto-reject. Owner: AP lead. Log: decision, approver, time. Test: monthly sample of 20; seeded errors quarterly.
"Prevent hallucinations."	Trigger: every customer answer. Check: answer must cite a retrieved approved document; confidence below threshold → no answer. Action: hand-off to agent with context. Owner: content owner. Log: citations, hand-offs. Test: weekly 50-question regression set.
"Monitor bias."	Trigger: monthly. Check: approval-rate ratio by age band vs tolerance 0.8–1.25. Action: > tolerance → L1 investigation; persists 2 months → L2 restrict. Owner: model-risk lead. Log: report. Test: reconciliation by second line.

Control type	Purpose	Examples
Guardrail (preventive)	Stops a harmful action or output automatically	Action allow-lists, spend caps, content filters, blocked data classes

Control type	Purpose	Examples
Approval gate	Requires a human decision before something takes effect	High-value actions; adverse decisions about people; first use in a new context
Detective control	Finds problems after the fact	Sampling, reconciliations, drift monitors
Escalation path	Routes problems to the right authority in time	L1–L4 ladder with time limits (Module 1)
Kill switch and fallback	Stops the system and keeps the business running	Named owner, tested procedure, manual process

2.10 Testing human-in-the-loop at real volume

Most AI risk plans rely on human review as the decisive control — and almost none test it. Two tests are needed: **can reviewers do it** (capacity) and **do they catch errors when they do it** (effectiveness).

TEST 1 · CAPACITY

Volume: 12,000 reviews / month
Needed: 4 min each → 800 h
Available: 2 FTE × 160 h = 320 h

Effective time: 1.6 min per review
→ below the minimum needed to check sources; review degrades silently

Question: can it be done?

TEST 2 · EFFECTIVENESS

Seed 50 known errors into the normal review queue (unannounced timing, agreed with works council)

Caught: 31 of 50 = 62%
Target was ≥ 90% for material errors
→ control not effective at this load

Question: does it catch errors?

Figure 2.3 — Capacity and effectiveness tests for human review

Test	Method	Pass criterion (example)
Capacity	Volume × review rate × minutes needed vs qualified hours available, per phase and at peak	Available ≥ needed at peak, with buffer
Time-on-task	Review-time distribution from logs	Median ≥ minimum time needed to check sources
Seeded-error test	Insert known, realistic errors into the live queue; measure catch rate	≥ 90% of material seeded errors caught
Override health	Edit, override and rejection rates over time	Stable, non-zero rates; no collapse over weeks
Information sufficiency	Can reviewers see sources and reasons?	Reviewers can verify each material claim
Authority	Can reviewers reject without penalty or delay targets?	Rejections are not counted against productivity

EXAMPLE · How large a sample?

To have 95% confidence of finding at least one error when the true error rate is 2%, a sampling reviewer needs about **149** items ($n \geq \ln 0.05 \div \ln 0.98$). Smaller samples will often show "no errors" even when 2% of outputs are wrong — which is why "we sampled 20 and found nothing" is weak evidence.

NOTE · Doing seeded-error tests responsibly

Agree the design with HR and, where applicable, the works council; ensure seeded items cannot reach customers; use results to fix the **system** (staffing, tools, information), not to sanction individuals.

2.11 Audit readiness and monitoring

Audit file component	Content
Inventory and classification	Entry, classification record, standard and date
Risk register	Specific failure statements, ratings, owners, controls, residual risk
Control specifications and tests	Buildable specs; test results incl. HITL capacity and seeded-error tests
Decision records	Approvals, conditions, accepted gaps (Module 1)
Vendor file	Due diligence, contract clauses, change notices
Monitoring and incidents	Reports against thresholds; incident log and actions
Training and literacy	Role-based measures and records
Change log	Model, prompt, data, configuration and vendor changes with re-validation

BEST PRACTICE · Key risk indicators (KRIs)

Select a handful per system: oversight health (median review time, override rate), performance on reference sample, fairness ratio, incident count and severity, out-of-scope use, vendor change notices received vs tested. Each KRI has a threshold and an owner.

2.12 Reporting AI risk to executives

Instead of...	Report...
A heat map of categories (“model risk: amber”)	Top specific failures with consequence, control status and owner
“All controls in place”	Control test results: capacity, seeded-error catch rate, monitoring thresholds met
Long lists of open actions	Accepted residual risks and decisions requested
Generic regulatory updates	Dated obligations affecting our systems (e.g., Annex III deadline 2 Dec 2027)

CONSULTANT'S TIP

One page per material system: what it does, tier, top three failures and their control status, KRIs vs thresholds, decisions needed. Executives can then govern by exception.

2.13 The AI governance and risk plan

Plan section	Content
1 · Scope and classification	Systems, intended purposes, legal class and internal tier, standard and date

Plan section	Content
2 · Instruments and obligations	Applicable law, standards, frameworks; obligation map with owners
3 · Risk register	Specific failure statements across categories and lifecycle
4 · Controls	Guardrails, approval gates, detective controls, escalation, kill switch — buildable specs
5 · Human oversight	Design, staffing, capacity and effectiveness tests
6 · Third parties	Vendor DD, contract terms, change management, exit
7 · Monitoring and KRIs	Metrics, thresholds, owners, reporting
8 · Incidents and escalation	Ladder, reporting duties, rehearsal plan
9 · Audit readiness	Audit file index; assurance plan (second and third line)
10 · Gaps and decisions	Residual risks, acceptances, review dates

2.14 Case study: Halvorsen Distribution

CASE STUDY **An agent that orders from suppliers**

Halvorsen Distribution (fictional) · industrial parts distributor · 4 warehouses

Situation

Halvorsen plans a procurement agent that reads demand forecasts and stock levels, selects suppliers, places purchase orders through the ERP and answers supplier e-mails. The draft risk plan lists “agentic risk — high — mitigated by human oversight”.

What the risk review did

- **Classification:** not an Annex III use; transparency duty applies to supplier-facing e-mails generated by AI. Internal tier 3 because of financial exposure and autonomy.
- **Specific failures (excerpt):** duplicate orders when ERP confirmation is delayed (€40–120k excess stock each); new-supplier orders to a fraudulent bank account after a spoofed e-mail (prompt injection / social engineering); agent commits to delivery dates it cannot meet in supplier e-mails; silent vendor model update degrades supplier selection.
- **Controls specified:** idempotency check before each order (guardrail); approval gate for orders > €25k, new suppliers and any bank-detail change; agent treats e-mail content as data and cannot change master data; daily spend cap; kill switch owned by the Head of Procurement with manual fallback; vendor change-notification clause and weekly regression test.
- **HITL test:** projected 900 approvals/month at 6 minutes = 90 hours; procurement could provide 40. Disposition: raise the threshold to €40k for known suppliers (reducing approvals to ~380 = 38 hours) and add a second approver at month end. Seeded-error test: 20 seeded anomalous orders, 18 caught (90%) — passed.
- **Plan:** audit file opened; KRIs: duplicate-order count, approval turnaround, spend vs cap, out-of-scope e-mails.

Discussion questions

1. Why was “agentic risk — mitigated by human oversight” inadequate?
2. Which failures needed preventive controls rather than detective ones?
3. How did the HITL capacity test change the design?
4. What makes the approval-gate control “buildable”?

Model answers (brief)

1. It named a category, not failures. The oversight it relied on was untested and, at the planned volume, could not have operated.
2. Payments to changed bank details and duplicate orders are irreversible or costly, so they need guardrails and gates before the action, not log review afterwards.
3. Needed hours exceeded available hours. The threshold and approver design were changed until capacity matched, then effectiveness was tested.

4. Defined trigger (value, new supplier, bank change), the information shown, a time limit and default, a named owner, logging, and a test regime.

2.15 Consulting templates

T2.1 Lifecycle Risk Register

Purpose: Record specific failures across the lifecycle.

ID	Lifecycle stage	Category	Specific failure (system · wrong · condition · whom · consequence)	Likelihood	Impact	Control(s)	Owner	Residual

T2.2 Failure Statement Builder

Purpose: Turn a broad category into a controllable failure.

Broad category: .....

System / component: .....

Does what wrong: .....

Under which condition: .....

To whom: .....

Consequence (what, how much, how often): .....

How we would detect it: .....

Control that would prevent or limit it: .....

T2.3 AI Vendor Risk Checklist

Purpose: Assess third-party AI risk and contract protections.

- [ ] Model, training data and limits documented (instructions for use)
- [ ] Change notification for model updates; regression-testing rights; version pinning
- [ ] No training on our data; data residency; sub-processor list
- [ ] Security testing, certifications and incident notification
- [ ] Upstream (fourth-party) providers disclosed; flow-down terms
- [ ] Viability reviewed; exit assistance and data export agreed
- [ ] AI Act roles allocated (provider / deployer) with cooperation clauses
- [ ] Audit and information rights

Gaps and decisions: .....

T2.4 Control Specification (Guardrail / Gate / Escalation)

Purpose: Specify controls a team can build and test.

Control	Type	Trigger	Check performed	Action / default	Owner	Log	Test & frequency
Order approval gate	*Gate*	*> €25k, new supplier, bank change*	*PO, invoice, supplier record*	*Approve / reject in 4 h; else reject*	*Head of Procurement*	*ERP approval log*	*Monthly sample; quarterly seeded*

T2.5 Agentic Control Checklist

Purpose: Check controls for systems that act.

- [ ] Least-privilege identity; allow-list of actions
- [ ] Approval gates before irreversible or high-value actions
- [ ] Untrusted content treated as data; restricted actions after reading it
- [ ] Step, rate, time and spend limits
- [ ] Scope checks and escalation on out-of-scope inputs
- [ ] Full action logs attributable to the agent identity
- [ ] Kill switch owner named; fallback process tested
- [ ] Staged autonomy plan with evidence criteria

T2.6 Human-in-the-Loop Volume & Effectiveness Test

Purpose: Prove that human review can operate and catches errors.

CAPACITY

Phase / peak	Volume	Review rate	Minutes needed	Hours needed	Qualified hours available	Effective min / item	Pass?

EFFECTIVENESS

Seeded-error design (number, types, safeguards, agreed with): .....

Catch rate (caught / seeded) and target: .....

Median review time vs minimum: .....

Override / edit rate trend: .....

Sampling size for detecting error rate p ($n \geq \ln 0.05 \div \ln (1 - p)$): .....

Disposition if failed: .....

T2.7 AI Governance & Risk Plan Outline

Purpose: Assemble the plan and audit file index.

- [ ] 1 Scope and classification (standard and date)
- [ ] 2 Instruments and obligation map with owners
- [ ] 3 Risk register with specific failures
- [ ] 4 Control specifications (guardrails, gates, detective, escalation, kill switch)
- [ ] 5 Human oversight design and test results
- [ ] 6 Third-party file
- [ ] 7 Monitoring and KRIs with thresholds
- [ ] 8 Incident and escalation plan, rehearsal date
- [ ] 9 Audit file index and assurance plan
- [ ] 10 Residual risks, acceptances and review dates

2.16 Best practices and pitfalls

✓ DO THIS

- Run risk activities at every lifecycle stage.
- Write specific failure statements with consequences.
- Use the Act for obligations and the RMF for risk method.
- Put preventive gates before irreversible agent actions.
- Specify controls with trigger, check, action, owner, log, test.
- Test human review for capacity and effectiveness.
- Report specific failures and control tests to executives.

× AVOID THIS

- Risk assessment once, at approval only.
- Category-level risks (“bias: amber”).
- Assuming minimal legal risk means minimal business risk.
- Relying on after-the-fact log review for agents.
- “Human oversight” as an untested assurance.
- Concluding “no errors” from tiny samples.
- Using seeded-error results to blame individuals.

2.17 Module 2 cheat sheet

MODULE 2 · ONE-PAGE SUMMARY

AI Risk Management

LIFECYCLE

Design · data · build/buy · test · deploy · operate · change/ retire. Material change → re-assess.

CATEGORIES

Model · data · regulatory · third-party · security · operational/ people · agentic.

ACT + RMF

EU AI Act: what must we do, by when (legal classes). **NIST AI RMF**: what could go wrong, how to measure/manage (Map–Measure–Manage; GenAI profile).

FAILURE STATEMENTS

System · does what wrong · condition · to whom · consequence. Rate and control failures, not categories.

VENDOR RISK

Documentation · change notice & regression rights · no training on data · security · fourth parties · exit · AI Act role allocation.

AGENTIC RISK

Compounding errors · permission creep · prompt injection · irreversible actions · runaway costs · accountability gaps · scope creep. Gates **before** irreversible actions.

BUILDABLE CONTROLS

Trigger · check · action/default · owner · log · test. Types: guardrail (preventive) · approval gate · detective · escalation path · kill switch & fallback.

HITL TESTS

Capacity (volume × rate × minutes vs hours) · time-on-task · **seeded-error catch rate** · override health · information · authority. Sample $n \geq \ln 0.05 \div \ln(1-p)$ (≈149 for 2%).

AUDIT & REPORTING

Audit file: inventory, register, control tests, decisions, vendor file, monitoring, training, change log. KRIs with thresholds. Executives: specific failures, control tests, accepted risks, decisions.

2.18 Knowledge check

Answer all questions before turning the page.

K1 **Which is a specific failure statement?**

(A) "Vendor risk: amber."
(B) "Hallucination risk is high."
(C) "The policy assistant cites a superseded rule when retrieval returns an archived document; customers receive wrong advice."
(D) "AI is risky."

K2 **How should the EU AI Act and NIST AI RMF be used together?**

(A) Choose one; they conflict
(B) NIST replaces the Act outside the US
(C) The Act is only for providers
(D) The Act for legal obligations and deadlines; the RMF to identify, measure and manage risks beyond the legal minimum

K3 **A vendor silently updates its model and accuracy drops unnoticed. Which controls address this?**

(A) Change-notification clause, regression testing and accuracy monitoring
(B) Higher licence fees
(C) Removing logs
(D) More user training

K4 **An agent can make payments to suppliers. Which control is essential for bank-detail changes?**

(A) Monthly log review
(B) An approval gate (and no ability to change master data) before payment to changed bank details
(C) A faster model
(D) A disclaimer

K5 **What makes a control "buildable"?**

(A) It is approved by the board
(B) It is described in the policy
(C) It has a defined trigger, check, action or default, owner, log and test
(D) It uses AI

K6 **A seeded-error test inserts 50 known errors and reviewers catch 31. What is the conclusion?**

(A) Oversight is not effective at this load; fix staffing, tools or information and re-test
(B) Reviewers should be disciplined
(C) The test is invalid
(D) Oversight is effective

K7 **A reviewer samples 20 outputs and finds no errors. What can be concluded about a possible 2% error rate?**

- (A) Sampling is useless
- (B) Little: about 149 items are needed for 95% confidence of seeing at least one error at 2%
- (C) The error rate is exactly 0%
- (D) There are no errors

K8 **Which is the best way to report AI risk to executives?**

- (A) A category heat map
- (B) Only regulatory news
- (C) All open actions
- (D) Top specific failures with consequence, control status and owner; KRIs vs thresholds; decisions needed

2.19 Knowledge check: answers and explanations

K1 **Answer: C** Review: 2.4

Which is a specific failure statement?

- ✗ (A) "Vendor risk: amber." — **Wrong.** A heat-map category.
- ✗ (B) "Hallucination risk is high." — **Wrong.** A category, not a failure.
- ✓ (C) "The policy assistant cites a superseded rule when retrieval returns an archived document; customers receive wrong advice." — **Correct.** It names system, error, condition, affected people and consequence.
- ✗ (D) "AI is risky." — **Wrong.** Too vague.

K2 **Answer: D** Review: 2.3

How should the EU AI Act and NIST AI RMF be used together?

- ✗ (A) Choose one; they conflict — **Wrong.** They are complementary.
- ✗ (B) NIST replaces the Act outside the US — **Wrong.** NIST is voluntary; the Act applies in the EU market.
- ✗ (C) The Act is only for providers — **Wrong.** Deployers have obligations too.
- ✓ (D) The Act for legal obligations and deadlines; the RMF to identify, measure and manage risks beyond the legal minimum — **Correct.** Each answers a different question.

K3 **Answer: A** Review: 2.7

A vendor silently updates its model and accuracy drops unnoticed. Which controls address this?

- ✓ (A) Change-notification clause, regression testing and accuracy monitoring — **Correct.** Contract and detective controls.
- ✗ (B) Higher licence fees — **Wrong.** Irrelevant.
- ✗ (C) Removing logs — **Wrong.** Counterproductive.
- ✗ (D) More user training — **Wrong.** Does not detect model changes.

K4 **Answer: B** Review: 2.8

An agent can make payments to suppliers. Which control is essential for bank-detail changes?

- ✗ (A) Monthly log review — **Wrong.** Too late for irreversible payments.
- ✓ (B) An approval gate (and no ability to change master data) before payment to changed bank details — **Correct.** Preventive control before an irreversible action.
- ✗ (C) A faster model — **Wrong.** Irrelevant.
- ✗ (D) A disclaimer — **Wrong.** Not a control.

K5 **Answer: C** Review: 2.9

What makes a control "buildable"?

- ✗ (A) It is approved by the board — **Wrong.** Approval does not make it buildable.
- ✗ (B) It is described in the policy — **Wrong.** Policy wording is not a specification.
- ✓ (C) It has a defined trigger, check, action or default, owner, log and test — **Correct.** A team can implement and verify it.
- ✗ (D) It uses AI — **Wrong.** Irrelevant.

K6 **Answer: A** Review: 2.10

A seeded-error test inserts 50 known errors and reviewers catch 31. What is the conclusion?

✓ **A** Oversight is not effective at this load; fix staffing, tools or information and re-test — **Correct.** The control fails the effectiveness test.

✗ B Reviewers should be disciplined — **Wrong.** Results should fix the system, not blame individuals.

✗ C The test is invalid — **Wrong.** No basis for that.

✗ D Oversight is effective — **Wrong.** A 62% catch rate is well below a typical 90% target.

K7 **Answer: B** Review: 2.10

A reviewer samples 20 outputs and finds no errors. What can be concluded about a possible 2% error rate?

✗ A Sampling is useless — **Wrong.** Sampling works with adequate size.

✓ **B** Little: about 149 items are needed for 95% confidence of seeing at least one error at 2% — **Correct.** Small samples give weak evidence.

✗ C The error rate is exactly 0% — **Wrong.** Unsupported.

✗ D There are no errors — **Wrong.** A sample of 20 is too small to detect a 2% rate reliably.

K8 **Answer: D** Review: 2.12

Which is the best way to report AI risk to executives?

✗ A A category heat map — **Wrong.** Hides specifics.

✗ B Only regulatory news — **Wrong.** Insufficient.

✗ C All open actions — **Wrong.** Too much detail.

✓ **D** Top specific failures with consequence, control status and owner; KRIs vs thresholds; decisions needed — **Correct.** Enables governance by exception.

R1 Course at a glance: the master cheat sheet

BOTH MODULES ON ONE PAGE

AI Governance, Risk, and Executive Communication

M1 · GOVERNANCE AND ETHICS

Governance = decision rights · policy · classification · controls · assurance · evidence · literacy. Instruments by question: **framework** (NIST AI RMF) · **certifiable standard** (ISO/IEC 42001) · **law** (EU AI Act, GDPR). Policy turns ethics into rules. Classification → obligations with owners and evidence. EU AI Act as amended by Reg. (EU) 2026/1744: prohibitions & literacy (now "take measures to support") since Feb 2025 · GPAI Aug 2025 · transparency Aug 2026 · high-risk Annex III **2 Dec 2027**, Annex I **2 Aug 2028**. Oversight at volume. Monitoring and L1–L4 escalation. Data boundaries. Bias evaluation. Decision records with accepted gaps. Executive reporting.

M2 · RISK MANAGEMENT

Lifecycle: design → data → build/buy → test → deploy → operate → change/retire. Categories: model · data · regulatory · vendor · security · operational · agentic. Act for obligations, RMF for method. **Specific failure statements**. Vendor controls incl. change notice and fourth parties. Agentic: gates **before** irreversible actions, least privilege, limits, kill switch. **Buildable controls**: trigger · check · action · owner · log · test. **HITL tests**: capacity, time-on-task, seeded errors, override health; sample size $n \geq \ln 0.05 \div \ln(1-p)$. Audit file, KRIs, governance and risk plan, exception-based executive reports.

R2 Template index

Code	Template	Use when
T1.1	Governance Instrument Map	Which instruments apply
T1.2	Responsible AI Policy Checklist	Policy design
T1.3	Risk Classification & Obligations Map	Classification
T1.4	Data Boundary Matrix	Data rules
T1.5	Bias Evaluation Plan	Fairness
T1.6	Governance Decision Record	Every material decision
T1.7	Monitoring & Escalation Plan	Operations
T2.1	Lifecycle Risk Register	Risk identification
T2.2	Failure Statement Builder	Specific failures
T2.3	AI Vendor Risk Checklist	Third parties
T2.4	Control Specification	Buildable controls
T2.5	Agentic Control Checklist	Agents
T2.6	HITL Volume & Effectiveness Test	Before relying on review
T2.7	AI Governance & Risk Plan Outline	Plan and audit file

R3 Glossary

A

Accountable executive
Senior person answerable for an AI system or programme's outcomes.

Agentic AI
AI that plans and executes actions through tools, sometimes without review of each step.

AI literacy (Article 4)
EU AI Act duty, as amended in 2026, for providers and deployers to take measures to support AI literacy of staff and others operating AI on their behalf.

AI management system
Organization-wide system of policies, roles, processes and controls for AI; specified by ISO/IEC 42001.

AI Omnibus
Regulation (EU) 2026/1744, amending the EU AI Act; in force 27 July 2026; deferred high-risk obligations.

Annex I (EU AI Act)
EU product legislation under which AI safety components can be high-risk; obligations apply from 2 August 2028 (as amended).

Annex III (EU AI Act)
List of high-risk use areas (e.g., employment, creditworthiness, education, essential services); obligations apply from 2 December 2027 (as amended).

Approval gate
Control requiring a human decision before an output or action takes effect.

Audit file
Organized evidence showing a system is governed: inventory, register, control tests, decisions, vendor file, monitoring, training, change log.

Automation bias
Tendency of reviewers to accept automated outputs without adequate checking.

B

Bias evaluation
Structured assessment of whether a system treats groups differently in unjustified ways.

Buildable control
Control specified with trigger, check, action/default, owner, log and test.

C

Capacity test (HITL)
Comparison of review hours needed with qualified hours available, per phase and at peak.

Classification record
Documented legal class and internal tier for a system, with basis, standard and date.

D

Data boundary policy
Rules on which data classes may be used with which AI tools, including outputs and retention.

Deployer
Organization using an AI system under its authority (EU AI Act).

Detective control
Control that finds problems after they occur, e.g., sampling, reconciliation, monitoring.

DPIA
Data Protection Impact Assessment under GDPR.

E

Escalation ladder
Defined levels (investigate, restrict, suspend, report) with triggers, decision-holders and time limits.

Ethics review trigger
Condition that sends a use to ethical review, e.g., novel use or vulnerable people affected.

EU AI Act
Regulation (EU) 2024/1689 laying down rules on AI by risk; amended by Regulation (EU) 2026/1744.

F

Failure statement
Specific description of how a system fails: system, error, condition, affected party, consequence.

Fairness metric
Quantitative measure of group differences (selection rates, error rates, calibration); metrics can conflict.

Fourth-party risk
Risk from a vendor's own suppliers, such as upstream model providers.

FRIA
Fundamental-rights impact assessment required of certain deployers of high-risk systems.

G

General-purpose AI model
Model usable for many tasks; providers have specific obligations under the EU AI Act.

Governance decision record
Record of a material governance decision: reasoning, owners, controls, monitoring, escalation, accepted gaps, review date.

Guardrail
Preventive control that automatically blocks harmful outputs or actions.

H

High-risk AI system
System in an Annex III use area or a safety component under Annex I product law.

Human-in-the-loop (HITL)
Design in which humans review or approve AI outputs or actions.

I

ISO/IEC 23894
Guidance standard on AI risk management (not certifiable).

ISO/IEC 42001
Certifiable AI management-system standard.

K

Key risk indicator (KRI)
Metric with threshold signalling rising risk, e.g., median review time, override rate.

Kill switch
Tested mechanism and named owner able to stop an AI system immediately, with fallback.

L

Least privilege
Granting only the permissions strictly needed.

M

Model risk
Risk from model inaccuracy, instability, drift, opacity or misuse.

N

NIST AI 600-1
Generative AI Profile of the NIST AI RMF, cataloguing generative-AI risks.

NIST AI RMF
Voluntary AI Risk Management Framework (Govern, Map, Measure, Manage).

P

Prohibited practices
AI uses banned under the EU AI Act, applying since 2 February 2025.

Prompt injection
Instructions hidden in content that manipulate an AI system's behaviour.

Provider
Organization that develops an AI system and places it on the market or into service (EU AI Act).

R

Regulatory risk
Risk of breaching applicable law through misclassification or unmet obligations.

Residual risk
Risk remaining after controls; must be accepted by a named owner.

Responsible AI policy
Organizational policy turning principles into scope, roles, classification, uses, boundaries, oversight and incidents.

Risk tier (internal)
Organization's own risk scale, mapped to legal classes as a minimum.

S

Seeded-error test
Inserting known errors into a review queue to measure reviewers' catch rate.

Serious incident
Incident with severe consequences that may trigger reporting duties under the EU AI Act.

T

Third-party AI risk
Risk arising from AI products and services supplied by vendors.

Transparency obligations
Duties to inform people about AI interaction or AI-generated content (EU AI Act Article 50).

V

Voluntary framework
Non-binding guidance for managing risk, e.g., NIST AI RMF.

R4 Mock exam

Format: 36 multiple-choice questions · one best answer each · **Time:** 60 minutes · **Suggested pass mark:** 70% (25 correct)

Questions 1–16 cover Module 1, 17–32 Module 2 and 33–36 are integrated scenarios. Answer every question before checking the answer key, which explains why each option is right or wrong.

Q1 **Which instrument answers "what must we do, and what happens if we don't?"**

- A Binding law such as the EU AI Act
- B OECD AI Principles
- C NIST AI RMF
- D ISO/IEC 23894

Q2 **Which instrument can be certified by an accredited body?**

- A OECD Principles
- B NIST AI RMF
- C ISO/IEC 42001
- D ISO/IEC 23894

Q3 **A firm is ISO/IEC 42001 certified. What does that establish?**

- A Full EU AI Act compliance for every system
- B Nothing
- C That models are unbiased
- D That an AI management system operates and was audited against the standard

Q4 **Which is the best test of a responsible AI policy's operability?**

- A Whether it lists many principles
- B Whether staff can correctly apply it to realistic scenarios
- C Its length
- D Whether legal approved it

Q5 **Which ethical question best reflects proportionality?**

- A "Is AI necessary for this purpose, or are we using it because we can?"
- B "Is it cheap?"
- C "Is the vendor large?"
- D "Is it legal?"

Q6 **Which use is high-risk under Annex III?**

- A Spam filtering
- B Code completion for developers
- C Internal translation
- D AI used to rank job applicants

Q7 **As amended in 2026, when do obligations for AI safety components in products under Annex I apply?**

- A 2 December 2027
- B 2 August 2028
- C 2 August 2026
- D Never

Q8 **Which obligation falls on deployers of high-risk systems?**

- A Conformity assessment of the system
- B Writing technical documentation
- C Assigning human oversight to people with competence, training and authority, and keeping logs under their control
- D Registering the provider's QMS

Q9 **Which statement about the deferral of high-risk obligations is correct?**

- A The Act was repealed.
- B Only GDPR now applies.
- C High-risk obligations were deferred, while prohibitions, AI literacy, GPAI and transparency duties continue
- D All AI Act obligations are paused.

Q10 **What makes a risk classification useful?**

- A Many tiers
- B Colours and labels
- C Vendor agreement
- D Each class triggers named obligations with owners and evidence

Q11 **12,000 reviews per month at 4 minutes, with 320 hours available. What is the effective review time?**

- A About 1.6 minutes
- B 8 minutes
- C 4 minutes
- D 0.4 minutes

Q12 **Which monitoring metric best indicates oversight health?**

- A Number of users
- B Median review time and override/edit rates
- C Server uptime
- D Licence count

Q13 **Which escalation level involves reducing autonomy or scope?**

- A L1 investigate
- B L3 suspend
- C L4 report
- D L2 restrict

Q14 **Under a typical data boundary policy, may restricted health data be pasted into a public consumer AI tool?**

- A No
- B Only in emergencies
- C Yes, with manager approval
- D Yes, if anonymised by the user

Q15 **Why is choosing a fairness metric a governance decision?**

- A Because vendors require it
- B Because metrics can conflict, so the choice depends on the use and legal context and must be recorded with reasoning
- C Because it is purely technical
- D Because metrics are always identical

Q16 **Which field distinguishes a governance decision record from meeting minutes?**

- A Date
- B Location
- C Accepted residual risks with owner and review date
- D Attendees

Q17 **At which lifecycle stage are HITL capacity and effectiveness tests typically gate evidence?**

- A Data
- B Retire
- C Deploy
- D Design

Q18 **Which is a data risk?**

- A Rubber-stamping
- B Prompt injection
- C Vendor insolvency
- D Training data used without a legal basis

Q19 **What does the NIST AI RMF Map function contribute to classification?**

- A Context, purpose, impacts and risk tolerance
- B Certification
- C Legal categories
- D Penalties

Q20 **Which risk does NIST's Generative AI Profile call "confabulation"?**

- A Data residency
- B Confidently stated but false or fabricated content
- C Model theft
- D High cost

Q21 **Which element is missing from “bias risk: amber”?**

(A) Nothing
(B) A logo
(C) The system, error, condition, affected people and consequence
(D) Colour

Q22 **Which practice addresses model risk after a vendor prompt or model change?**

(A) Increase usage
(B) Remove monitoring
(C) Ignore minor changes
(D) Change control with re-validation on a regression set

Q23 **Which control best addresses misclassification risk?**

(A) More users
(B) A classification record with legal review of borderline cases
(C) A faster model
(D) A disclaimer

Q24 **Which contract clause addresses fourth-party risk?**

(A) Disclosure of upstream model providers with flow-down terms
(B) Logo use
(C) Payment terms
(D) Price cap

Q25 **An agent reads supplier e-mails. What is the main prompt-injection control?**

(A) Let the agent follow instructions in e-mails
(B) Use a larger model
(C) Train users
(D) Treat e-mail content as data, restrict actions after reading untrusted content, and gate master-data changes

Q26 **Why is after-the-fact log review insufficient for irreversible agent actions?**

(A) Agents do not create logs
(B) The harm has already happened; preventive gates or limits are required
(C) Logs are expensive
(D) Logs are illegal

Q27 **Which specification is buildable?**

(A) “Payments > €500 or to new suppliers require AP approval within 4 h, otherwise auto-reject; logged; tested monthly.”
(B) “Payments are monitored.”
(C) “Ensure oversight of payments.”
(D) “Be careful with payments.”

Q28 **Which control type stops a harmful action automatically?**

- A Detective
- B Escalation path
- C Guardrail (preventive)
- D Report

Q29 **What does a seeded-error test measure?**

- A Server speed
- B Cost
- C Reviewers' catch rate for known errors under normal working conditions
- D Model accuracy

Q30 **About how many items must be sampled for 95% confidence of seeing at least one error if the true error rate is 2%?**

- A 1,000
- B About 149
- C 20
- D 5

Q31 **Which belongs in an AI audit file?**

- A Control specifications with test results, including HITL tests
- B Employee birthdays
- C Vendor sales decks
- D Marketing brochures

Q32 **Which report helps executives govern AI risk by exception?**

- A A maturity score
- B A full action list
- C Vendor news
- D One page per material system: top failures, control status, KRIs vs thresholds, decisions needed

Q33 **Integrated: An HR screening tool is classified high-risk, but its oversight relies on two recruiters reviewing 6,000 rankings a month. What is the governance-and-risk response?**

- A Remove human review
- B Run capacity and seeded-error tests, redesign oversight until it passes, and record the decision and residual risks
- C Approve — high-risk obligations are deferred to 2027
- D Ask the vendor to confirm compliance

Q34 **Integrated: A minimal-risk internal agent can delete records in the CRM. How should risk be treated?**

- A Use the RMF to identify specific failures, then gate irreversible deletions and apply least privilege
- B Classify as prohibited
- C No controls — legally minimal
- D Ignore until audit

Q35 **Integrated: The board asks “are we compliant with AI regulation?”. What is the best answer format?**

A “Yes — we are ISO certified.”

B “Mostly.”

C A list of frameworks read

D Tier summary of systems, dated obligations, control-test results, accepted gaps with owners, and decisions needed

Q36 **Integrated: Which sequence reflects this course?**

A Audit → design → classify

B Buy tool → write policy later

C Map instruments → set policy and classification → tie to obligations → identify specific failures → specify buildable controls → test oversight at volume → monitor, document and report

D Report → classify → buy

R5 Mock exam: answer key and explanations

Quick answer grid

1 · A	**2** · C	**3** · D	**4** · B	**5** · A	**6** · D	**7** · B	**8** · C	**9** · C	**10** · D
11 · A	**12** · B	**13** · D	**14** · A	**15** · B	**16** · C	**17** · C	**18** · D	**19** · A	**20** · B
21 · C	**22** · D	**23** · B	**24** · A	**25** · D	**26** · B	**27** · A	**28** · C	**29** · C	**30** · B
31 · A	**32** · D	**33** · B	**34** · A	**35** · D	**36** · C				

Score	Interpretation	Next step
31–36	Exam-ready	Review missed items; run a real assessment with the templates
25–30	Pass level	Re-read cheat sheets for modules where you missed two or more
below 25	Not yet	Re-study weak modules using the “Review” references; retake in one week

Full explanations

Q1 **Answer: A** Review: M1 · 1.2

Which instrument answers “what must we do, and what happens if we don’t?”

✓ **A** Binding law such as the EU AI Act — **Correct.** Law sets mandatory obligations and penalties.

✗ **B** OECD AI Principles — **Wrong.** Principles, not binding obligations.

✗ **C** NIST AI RMF — **Wrong.** A voluntary framework.

✗ **D** ISO/IEC 23894 — **Wrong.** A guidance standard.

Q2 **Answer: C** Review: M1 · 1.2

Which instrument can be certified by an accredited body?

✗ **A** OECD Principles — **Wrong.** Not certifiable.

✗ **B** NIST AI RMF — **Wrong.** No conformity regime.

✓ **C** ISO/IEC 42001 — **Correct.** A certifiable management-system standard.

✗ **D** ISO/IEC 23894 — **Wrong.** Guidance only.

Q3 **Answer: D** Review: M1 · 1.2

A firm is ISO/IEC 42001 certified. What does that establish?

✗ **A** Full EU AI Act compliance for every system — **Wrong.** Certification does not establish system-level legal compliance.

✗ **B** Nothing — **Wrong.** It shows a management system.

✗ **C** That models are unbiased — **Wrong.** Not established.

✓ **D** That an AI management system operates and was audited against the standard — **Correct.** That is what certification shows.

Q4 **Answer: B** Review: M1 · 1.3

Which is the best test of a responsible AI policy's operability?

✗ (A) Whether it lists many principles — **Wrong.** Principles need rules.

✓ (B) Whether staff can correctly apply it to realistic scenarios — **Correct.** Operability is shown in use.

✗ (C) Its length — **Wrong.** Irrelevant.

✗ (D) Whether legal approved it — **Wrong.** Necessary but not sufficient.

Q5 **Answer: A** Review: M1 · 1.4

Which ethical question best reflects proportionality?

✓ (A) "Is AI necessary for this purpose, or are we using it because we can?" — **Correct.** Proportionality.

✗ (B) "Is it cheap?" — **Wrong.** Cost, not ethics.

✗ (C) "Is the vendor large?" — **Wrong.** Irrelevant.

✗ (D) "Is it legal?" — **Wrong.** That is compliance.

Q6 **Answer: D** Review: M1 · 1.5

Which use is high-risk under Annex III?

✗ (A) Spam filtering — **Wrong.** Minimal.

✗ (B) Code completion for developers — **Wrong.** Minimal.

✗ (C) Internal translation — **Wrong.** Minimal.

✓ (D) AI used to rank job applicants — **Correct.** Employment is an Annex III area.

Q7 **Answer: B** Review: M1 · 1.6

As amended in 2026, when do obligations for AI safety components in products under Annex I apply?

✗ (A) 2 December 2027 — **Wrong.** That is the date for stand-alone Annex III systems.

✓ (B) 2 August 2028 — **Correct.** Deferred by the AI Omnibus.

✗ (C) 2 August 2026 — **Wrong.** Not this date.

✗ (D) Never — **Wrong.** They apply from 2028.

Q8 **Answer: C** Review: M1 · 1.6

Which obligation falls on deployers of high-risk systems?

✗ (A) Conformity assessment of the system — **Wrong.** A provider duty.

✗ (B) Writing technical documentation — **Wrong.** A provider duty.

✓ (C) Assigning human oversight to people with competence, training and authority, and keeping logs under their control — **Correct.** Deployer duties.

✗ (D) Registering the provider's QMS — **Wrong.** A provider duty.

Q9 **Answer: C** Review: M1 · 1.6

Which statement about the deferral of high-risk obligations is correct?

✗ (A) The Act was repealed. — **Wrong.** It was amended.

✗ (B) Only GDPR now applies. — **Wrong.** No.

✓ (C) High-risk obligations were deferred, while prohibitions, AI literacy, GPAI and transparency duties continue — **Correct.** Deferred is not paused.

✗ (D) All AI Act obligations are paused. — **Wrong.** Prohibitions, literacy, transparency and GPAI rules continue.

Q10 **Answer: D** Review: M1 · 1.7

What makes a risk classification useful?

✗ A Many tiers — **Wrong.** Not the point.

✗ B Colours and labels — **Wrong.** Cosmetic.

✗ C Vendor agreement — **Wrong.** Irrelevant.

✓ D Each class triggers named obligations with owners and evidence — **Correct.** Classification must change what happens.

Q11 **Answer: A** Review: M1 · 1.8

12,000 reviews per month at 4 minutes, with 320 hours available. What is the effective review time?

✓ A About 1.6 minutes — **Correct.** 320 h × 60 ÷ 12,000 = 1.6.

✗ B 8 minutes — **Wrong.** Incorrect.

✗ C 4 minutes — **Wrong.** That is the requirement.

✗ D 0.4 minutes — **Wrong.** Incorrect.

Q12 **Answer: B** Review: M1 · 1.9

Which monitoring metric best indicates oversight health?

✗ A Number of users — **Wrong.** Adoption.

✓ B Median review time and override/edit rates — **Correct.** These reveal rubber-stamping.

✗ C Server uptime — **Wrong.** Technical, not oversight.

✗ D Licence count — **Wrong.** Irrelevant.

Q13 **Answer: D** Review: M1 · 1.9

Which escalation level involves reducing autonomy or scope?

✗ A L1 investigate — **Wrong.** Investigation only.

✗ B L3 suspend — **Wrong.** Stop the system.

✗ C L4 report — **Wrong.** External reporting.

✓ D L2 restrict — **Correct.** Restrict autonomy or scope.

Q14 **Answer: A** Review: M1 · 1.10

Under a typical data boundary policy, may restricted health data be pasted into a public consumer AI tool?

✓ A No — **Correct.** Restricted data is never allowed in public tools.

✗ B Only in emergencies — **Wrong.** No.

✗ C Yes, with manager approval — **Wrong.** Approval does not change the tool's risk.

✗ D Yes, if anonymised by the user — **Wrong.** User anonymisation is unreliable; public tools are excluded.

Q15 **Answer: B** Review: M1 · 1.11

Why is choosing a fairness metric a governance decision?

✗ A Because vendors require it — **Wrong.** No.

✓ B Because metrics can conflict, so the choice depends on the use and legal context and must be recorded with reasoning — **Correct.** A judgment requiring accountability.

✗ C Because it is purely technical — **Wrong.** It has ethical and legal dimensions.

✗ D Because metrics are always identical — **Wrong.** They can conflict.

Q16 **Answer: C** Review: M1 · 1.12

Which field distinguishes a governance decision record from meeting minutes?

✗ (A) Date — **Wrong.** Both have dates.

✗ (B) Location — **Wrong.** Irrelevant.

✓ (C) Accepted residual risks with owner and review date — **Correct.** Key to defensibility.

✗ (D) Attendees — **Wrong.** Minutes include these.

Q17 **Answer: C** Review: M2 · 2.1

At which lifecycle stage are HITL capacity and effectiveness tests typically gate evidence?

✗ (A) Data — **Wrong.** Not this stage.

✗ (B) Retire — **Wrong.** Too late.

✓ (C) Deploy — **Correct.** The go-live gate needs oversight evidence.

✗ (D) Design — **Wrong.** Too early to test at volume.

Q18 **Answer: D** Review: M2 · 2.2

Which is a data risk?

✗ (A) Rubber-stamping — **Wrong.** Operational.

✗ (B) Prompt injection — **Wrong.** Security.

✗ (C) Vendor insolvency — **Wrong.** Third-party.

✓ (D) Training data used without a legal basis — **Correct.** Rights and provenance are data risks.

Q19 **Answer: A** Review: M2 · 2.3

What does the NIST AI RMF Map function contribute to classification?

✓ (A) Context, purpose, impacts and risk tolerance — **Correct.** Context-based identification.

✗ (B) Certification — **Wrong.** Not part of NIST.

✗ (C) Legal categories — **Wrong.** That is the AI Act.

✗ (D) Penalties — **Wrong.** No.

Q20 **Answer: B** Review: M2 · 2.3

Which risk does NIST's Generative AI Profile call "confabulation"?

✗ (A) Data residency — **Wrong.** No.

✓ (B) Confidently stated but false or fabricated content — **Correct.** Commonly called hallucination.

✗ (C) Model theft — **Wrong.** Different risk.

✗ (D) High cost — **Wrong.** No.

Q21 **Answer: C** Review: M2 · 2.4

Which element is missing from "bias risk: amber"?

✗ (A) Nothing — **Wrong.** It is not controllable as written.

✗ (B) A logo — **Wrong.** Irrelevant.

✓ (C) The system, error, condition, affected people and consequence — **Correct.** A failure statement needs these.

✗ (D) Colour — **Wrong.** It has one.

Q22 **Answer: D** Review: M2 · 2.5

Which practice addresses model risk after a vendor prompt or model change?

✗ A Increase usage — **Wrong.** Irrelevant.

✗ B Remove monitoring — **Wrong.** Counterproductive.

✗ C Ignore minor changes — **Wrong.** Changes can alter behaviour.

✓ D Change control with re-validation on a regression set — **Correct.** Material changes trigger re-testing.

Q23 **Answer: B** Review: M2 · 2.6

Which control best addresses misclassification risk?

✗ A More users — **Wrong.** Irrelevant.

✓ B A classification record with legal review of borderline cases — **Correct.** Regulatory risk control.

✗ C A faster model — **Wrong.** Irrelevant.

✗ D A disclaimer — **Wrong.** Insufficient.

Q24 **Answer: A** Review: M2 · 2.7

Which contract clause addresses fourth-party risk?

✓ A Disclosure of upstream model providers with flow-down terms — **Correct.** Addresses the vendor's own suppliers.

✗ B Logo use — **Wrong.** Irrelevant.

✗ C Payment terms — **Wrong.** Commercial.

✗ D Price cap — **Wrong.** Commercial.

Q25 **Answer: D** Review: M2 · 2.8

An agent reads supplier e-mails. What is the main prompt-injection control?

✗ A Let the agent follow instructions in e-mails — **Wrong.** This is the vulnerability.

✗ B Use a larger model — **Wrong.** Does not solve injection.

✗ C Train users — **Wrong.** Insufficient alone.

✓ D Treat e-mail content as data, restrict actions after reading untrusted content, and gate master-data changes — **Correct.** Prevents manipulated actions.

Q26 **Answer: B** Review: M2 · 2.8

Why is after-the-fact log review insufficient for irreversible agent actions?

✗ A Agents do not create logs — **Wrong.** They should.

✓ B The harm has already happened; preventive gates or limits are required — **Correct.** Controls must act before irreversible actions.

✗ C Logs are expensive — **Wrong.** Not the reason.

✗ D Logs are illegal — **Wrong.** No.

Q27 **Answer: A** Review: M2 · 2.9

Which specification is buildable?

✓ A "Payments > €500 or to new suppliers require AP approval within 4 h, otherwise auto-reject; logged; tested monthly." — **Correct.** Trigger, action, owner, log, test.

✗ B "Payments are monitored." — **Wrong.** Vague.

✗ C "Ensure oversight of payments." — **Wrong.** Vague.

✗ D "Be careful with payments." — **Wrong.** Vague.

Q28 **Answer: C** Review: M2 · 2.9

Which control type stops a harmful action automatically?

✗ A Detective — **Wrong.** Finds problems afterwards.

✗ B Escalation path — **Wrong.** Routes problems.

✓ C Guardrail (preventive) — **Correct.** Blocks automatically.

✗ D Report — **Wrong.** Communicates.

Q29 **Answer: C** Review: M2 · 2.10

What does a seeded-error test measure?

✗ A Server speed — **Wrong.** No.

✗ B Cost — **Wrong.** No.

✓ C Reviewers' catch rate for known errors under normal working conditions — **Correct.** Effectiveness of human review.

✗ D Model accuracy — **Wrong.** Not directly.

Q30 **Answer: B** Review: M2 · 2.10

About how many items must be sampled for 95% confidence of seeing at least one error if the true error rate is 2%?

✗ A 1,000 — **Wrong.** More than needed for this purpose.

✓ B About 149 — **Correct.** $n \geq \ln 0.05 \div \ln 0.98$.

✗ C 20 — **Wrong.** Far too few.

✗ D 5 — **Wrong.** Far too few.

Q31 **Answer: A** Review: M2 · 2.11

Which belongs in an AI audit file?

✓ A Control specifications with test results, including HITL tests — **Correct.** Core audit evidence.

✗ B Employee birthdays — **Wrong.** Irrelevant.

✗ C Vendor sales decks — **Wrong.** Not evidence of control.

✗ D Marketing brochures — **Wrong.** Not evidence.

Q32 **Answer: D** Review: M2 · 2.12

Which report helps executives govern AI risk by exception?

✗ A A maturity score — **Wrong.** Not actionable.

✗ B A full action list — **Wrong.** Too detailed.

✗ C Vendor news — **Wrong.** Insufficient.

✓ D One page per material system: top failures, control status, KRIs vs thresholds, decisions needed — **Correct.** Governance by exception.

Q33 **Answer: B** Review: M1 · 1.6–1.8, M2 · 2.10

Integrated: An HR screening tool is classified high-risk, but its oversight relies on two recruiters reviewing 6,000 rankings a month. What is the governance-and-risk response?

✗ (A) Remove human review — **Wrong.** Increases risk.

✓ (B) Run capacity and seeded-error tests, redesign oversight until it passes, and record the decision and residual risks — **Correct.** Combines classification, oversight testing and decision records.

✗ (C) Approve — high-risk obligations are deferred to 2027 — **Wrong.** Deferral does not make an inadequate control acceptable, and the use affects people now.

✗ (D) Ask the vendor to confirm compliance — **Wrong.** The deployer's oversight is the deployer's responsibility.

Q34 **Answer: A** Review: M2 · 2.3, 2.8

Integrated: A minimal-risk internal agent can delete records in the CRM. How should risk be treated?

✓ (A) Use the RMF to identify specific failures, then gate irreversible deletions and apply least privilege — **Correct.** Business risk beyond the legal minimum.

✗ (B) Classify as prohibited — **Wrong.** Incorrect.

✗ (C) No controls — legally minimal — **Wrong.** Legal class sets a floor, not the business risk.

✗ (D) Ignore until audit — **Wrong.** Risky.

Q35 **Answer: D** Review: M1 · 1.13, M2 · 2.12

Integrated: The board asks "are we compliant with AI regulation?". What is the best answer format?

✗ (A) "Yes — we are ISO certified." — **Wrong.** Certification ≠ legal compliance.

✗ (B) "Mostly." — **Wrong.** Vague.

✗ (C) A list of frameworks read — **Wrong.** Not an answer.

✓ (D) Tier summary of systems, dated obligations, control-test results, accepted gaps with owners, and decisions needed — **Correct.** Specific, traceable and decision-oriented.

Q36 **Answer: C** Review: M1–M2

Integrated: Which sequence reflects this course?

✗ (A) Audit → design → classify — **Wrong.** Out of order.

✗ (B) Buy tool → write policy later — **Wrong.** Governance first.

✓ (C) Map instruments → set policy and classification → tie to obligations → identify specific failures → specify buildable controls → test oversight at volume → monitor, document and report — **Correct.** The course arc.

✗ (D) Report → classify → buy — **Wrong.** Out of order.

R6 Further reading and next steps

Useful public companions to this course. Check for the latest versions.

Resource	Why it is useful
EU AI Act — Regulation (EU) 2024/1689, consolidated with Regulation (EU) 2026/1744 (AI Omnibus)	Binding obligations and current dates; check Commission guidance and national authorities
NIST AI RMF 1.0, Playbook, and Generative AI Profile (NIST AI 600-1)	Risk identification and measurement method
ISO/IEC 42001:2023 and ISO/IEC 23894:2023	AI management system and risk-management guidance
GDPR and data-protection authority guidance on AI	Legal basis, automated decisions, DPIAs
Sector guidance (e.g., banking model-risk expectations, medical-device rules)	Integrating AI into existing risk regimes

Practice assignments

1. For one AI system you know, complete T1.1 and T1.3: which instruments apply, what class, what obligations, who owns them.
2. Draft a data boundary matrix (T1.4) and a governance decision record (T1.6) for its approval.
3. Write five specific failure statements (T2.2) and a buildable control for each (T2.4).
4. Run a paper capacity test and design a seeded-error test (T2.6); then assemble the governance and risk plan (T2.7).

ENTERPRISE AI TRANSFORMATION AND CONSULTING PROFESSIONAL

COURSE 7 · STUDY BOOK

AI Transformation Advisory Package

The capstone: a complete advisory engagement for a realistic manufacturer — opportunity assessment, portfolio, readiness, roadmap, business case and governance plan, integrated into one recommendation, presented to executives and documented as a portfolio piece.

4 MODULES · CASE STUDIES · TEMPLATES · CHEAT SHEETS · MOCK EXAM

Study edition · 2026

Includes full answer explanations

How to use this book

This study book accompanies **Course 7, AI Transformation Advisory Package**, of the Enterprise AI Transformation and Consulting Professional certificate. It is written for participants completing the certificate who want to practise the whole advisory engagement end to end, and to leave with a portfolio piece that demonstrates professional judgment.

Each module follows the same rhythm so that you always know where you are:

Element	What you get	How to use it
Module opener	Learning objectives, section list, study time	Read first; return to it to self-test at the end
Overview & key takeaways	The module on one screen	Skim before, re-read after
Core sections	Concepts, tables, diagrams, worked examples	Study actively: annotate, restate in your own words
Callout boxes	Key ideas, examples, pitfalls, exam traps, consultant tips	Exam traps are the most frequently tested ideas
Case study	A realistic enterprise scenario with model answers	Answer the questions before reading the model answers
Templates	Fill-in consulting tools (T-codes), one per page or less	Copy into your own engagements
Do / Avoid	Best practices and pitfalls side by side	Use as a pre-meeting checklist
Cheat sheet	One-page module summary	Revise from these in the final week
Exercises & model answers	Case-based tasks, each followed by a model solution and a rubric	Do the work first; compare; score yourself honestly
Knowledge check	Scenario questions with explanations for **every** option	Understanding why the wrong answers are wrong matters as much as getting the right one

Callout key

KEY IDEA
The idea to remember if you remember nothing else.

EXAM TRAP
A frequently tested distinction or a tempting wrong answer.

EXAMPLE
A worked illustration of the concept.

CONSULTANT'S TIP
Practical wording or technique from the field.

PITFALL
A common mistake made in real engagements.

DEFINITION
A precise definition worth learning word for word.

> **NOTE · A capstone, not a new course**
> This book introduces almost no new theory. It asks you to **apply** the frameworks of Courses 1–6 to one realistic case, in the order a real engagement would, and to produce the documents a client would receive. Every exercise names the course and template it draws on, gives you the inputs, asks you to work first, and then shows a model answer with a scoring rubric. Model answers are *one* defensible solution — yours may differ and still score well if your reasoning is traceable.

Suggested study plan (four weeks)

Week	Study	Practise
1	Read the case dossier; Module 1 exercises 1–5	Complete each exercise before reading the model answer; score yourself with the rubric
2	Module 2 exercises 6–11	Build the roadmap, ROI model, measurement and governance plans; write the null recommendations
3	Module 3	Run the consistency check; build the 12-slide storyline; rehearse Q&A aloud
4	Module 4, final self-assessment and mock exam	Write your reasoning log and hardest-finding narrative; then try the variant case on your own

> **NOTE · About the examples**
> All organizations in case studies and examples are fictional, and their figures are illustrative. Regulatory and framework references are for orientation only; always check current law and guidance, and refer legal interpretation to qualified specialists.

Contents

FRONT **How to use this book** 2

MODULE 1 **Opportunity Assessment, Portfolio and Readiness** 6
The capstone and your engagement plan 7
Case dossier: Brenner Antriebstechnik 8
Case dossier: interview excerpts 9
Case dossier: data extracts and AI landscape 9
Exercise 1 · Frame the ask and the decision rights 11
Exercise 2 · Opportunity inventory with value hypotheses and ledgers 12
Exercise 3 · Screening: gates, label tests and risk by use 14
Exercise 4 · Prioritize into a balanced portfolio 16
Exercise 5 · Readiness summary 18
Common mistakes in this phase 19
Module 1 cheat sheet 20
Knowledge check 21
Knowledge check: answers and explanations 23

MODULE 2 **Roadmap, Business Case and Governance** 25
Exercise 6 · Roadmap with preconditions and tranches 27
Exercise 7 · Framing the scheduling agent — and deciding against it 29
Exercise 8 · ROI model and business case 31
Exercise 9 · Measurement plan 33
Exercise 10 · Governance and risk plan 34
Exercise 11 · Pressure-test the plan 36
Recommending against: four null recommendations 37
Common mistakes in this phase 37
Module 2 cheat sheet 38
Knowledge check 39
Knowledge check: answers and explanations 41

MODULE 3 **Integrating the Package and the Executive Presentation** 43
What integration means 44
Exercise 12 · Consistency check 45
The single recommendation 46
Package architecture 46
Exercise 13 · The executive storyline 47
Exercise 14 · Investment committee memo 49
Delivering the hardest message 49
Q&A preparation 50
Red-teaming the package 51
Consulting templates 52
Module 3 cheat sheet 54
Knowledge check 55
Knowledge check: answers and explanations 57

MODULE 4 **Documenting Reasoning: The Portfolio Piece** 59
What a portfolio piece must show 60
Structure of the portfolio piece 61

Exercise 15 · Reasoning log 61
Evidence and assumption register 62
Exercise 16 · The hardest finding 63
Limitations and what you would do differently 64
Anonymisation and confidentiality 64
Exercise 17 · Package self-assessment 65
Talking about the work 66
Your development plan 66
Consulting templates 67
Module 4 cheat sheet 69
Knowledge check 70
Knowledge check: answers and explanations 72

BACK **The Advisor's Playbook: the master cheat sheet 74**
Framework and template index 75
Glossary of capstone terms 76
Integrated mock exam 77
Mock exam: answer key and explanations 83
Practice case and further reading 91

MODULE 1

Opportunity Assessment, Portfolio and Readiness

Meet the client — Brenner Antriebstechnik, a mid-size drive-systems manufacturer — and apply the frameworks of Courses 1–3 to frame the ask, build an opportunity inventory, screen and prioritize it into a portfolio, and summarise readiness.

YOU WILL BE ABLE TO

- Plan an AI advisory engagement end to end
- Work from a realistic case dossier: interviews, data extracts and an AI landscape
- Reframe the sponsor's ask and map decision rights
- Build an opportunity inventory from three channels with value hypotheses and ledgers
- Screen candidates with gates and separable dimensions
- Prioritize into a balanced portfolio with stated dispositions
- Write a readiness summary with proceed, phase or not-yet findings

IN THIS MODULE

1. The capstone and your engagement plan
2. Case dossier: the company
3. Case dossier: interviews
4. Case dossier: data and AI landscape
5. Exercise 1 · Frame the ask
6. Exercise 2 · Opportunity inventory
7. Exercise 3 · Screening
8. Exercise 4 · Portfolio
9. Exercise 5 · Readiness summary
10. Common mistakes · cheat sheet · knowledge check

Suggested study time · 8–10 hours (incl. exercises)

In this capstone you are the lead AI transformation advisor on a 12-week engagement. Module 1 gives you the full case dossier and walks you through the first half of the work: framing the ask, discovering opportunities, screening and prioritizing them into a portfolio, and assessing readiness. Each exercise asks you to work first; a model answer and a rubric follow.

KEY TAKEAWAYS

1. An engagement is a **sequence of decisions**, each needing its own artifact — plan them from day one.
2. Start from the **underlying problem** behind the CEO's ask, and name who decides what.
3. Use **all three inventory channels**: the best and the riskiest items rarely come from interviews alone.
4. Screen with **gates** (data, prohibited/high-risk use) before scoring; score on **separable dimensions**.
5. A portfolio needs **dispositions for everything** — including fund-now, prepare, test, defer, redirect and reject.
6. Readiness is a **chain**: one weak dimension can turn an attractive use case into "not yet".

1.1 The capstone and your engagement plan

You will produce the package a real client would receive. The table lists every artifact, the course it draws on and where in this book you build it.

#	Artifact	Draws on	Built in
1	Engagement charter and ask-to-problem reframing	C1 M1, C4 M4	Exercise 1
2	Opportunity inventory with value hypotheses and ledgers	C2 M1	Exercise 2
3	Screening results and dispositions	C2 M2	Exercise 3
4	Prioritized portfolio	C2 M3–M4	Exercise 4
5	Readiness summary	C3	Exercise 5
6	Roadmap with preconditions and tranches	C4 M1–M2	Exercise 6
7	Agent opportunity frame (and null recommendation)	C4 M3, C5 M4	Exercise 7
8	ROI model and business case	C5 M1–M2	Exercise 8
9	Measurement plan	C5 M3	Exercise 9
10	Governance and risk plan	C6	Exercise 10
11	Pressure-test log	C4 M4	Exercise 11
12	Integrated package, IC memo and executive deck	C5 M4, C7 M3	Exercises 12–14
13	Portfolio piece with reasoning log and hardest finding	C7 M4	Exercises 15–17

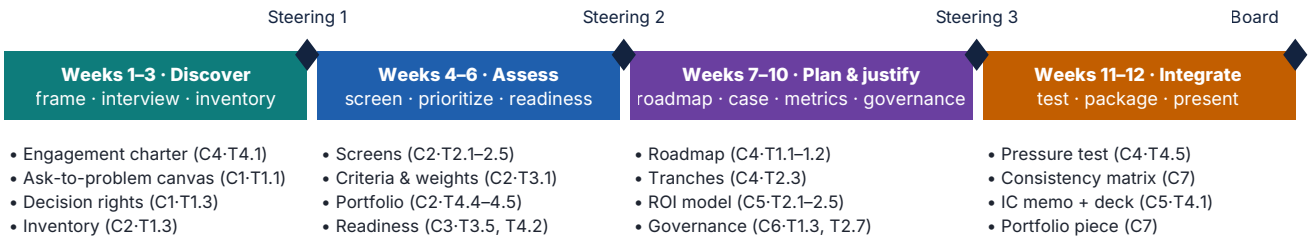


Figure 1.1 — A 12-week engagement plan with steering points

CONSULTANT'S TIP · How to use the exercises

Set a timer for the suggested time. Produce your answer in the referenced template (or a page of notes). Then read the model answer and score yourself on the rubric. Where your answer differs, ask: is my reasoning traceable to evidence in the dossier? If yes, a different answer can be equally good.

1.2 Case dossier: Brenner Antriebstechnik

CASE DOSSIER · COMPANY

Brenner Antriebstechnik GmbH (fictional) designs and manufactures industrial gearboxes and drive systems for conveyors, mixers and wind-power auxiliaries. Family-owned, third generation; the CEO, Dr Katrin Brenner, took over two years ago.

Fact	Value
Revenue / EBIT margin	€620m / 6.1% (down from 8.4% three years ago)
Employees	2,400 (1,300 production, 1,100 office and engineering incl. 120 service technicians)
Plants	Heilbronn (DE, lines 1–5), Plauen (DE, heat treatment and assembly), Brno (CZ, machining and assembly)
Products	~60% standard catalogue, ~40% engineered-to-order (1,400 custom quotes a year)
After-sales	Spare parts and service ≈ 18% of revenue, higher margin
Systems	ERP (migration to a common instance across plants due Q3 2027); MES at Heilbronn and Brno; separate CMMS for maintenance; document management system for manuals
Finance parameters	Discount rate 9%; evaluation horizon 5 years; preferred payback < 3 years; IRR hurdle 15%
Governance context	Works council (Betriebsrat) at both German plants; group DPO; no AI policy yet; IT security policy from 2021

Strategic context	Detail
Margin pressure	Energy and labour costs up; Asian competitors on standard gearboxes
Board pressure	Advisory board asked in March: "What is our AI plan?"
Competitor signal	A competitor announced "AI-optimised production" at Hannover Messe

Strategic context	Detail
CEO's trigger	At the same fair, Dr Brenner saw an autonomous production-scheduling agent demo and wants "something like that by next year"
Budget	CFO has indicated up to €2m over 18 months could be available "if the case is solid"

1.3 Case dossier: interview excerpts

Dr Katrin Brenner, CEO
"I want Brenner to be the AI leader in our niche. That scheduling agent could run our plants better than we do — it showed 8% more output in the demo. Also, IT wants to give everyone Copilot. Let's move fast."

Thomas Weiler, CFO
"I've seen three 'digital' cases in five years. None of the savings appeared in my numbers. If you bring me an AI case, show me where the money lands and who signs for it. I'm fine with saying no to things."

Andrea Kovács, COO
"Our real pain is quality on lines 4 and 5 at Heilbronn — scrap and rework cost us about €2.6m a year, plus customer claims. And unplanned downtime on the grinding machines. Scheduling? The planners juggle a lot of informal rules; I'm not sure the data shows how things really run."

Dr Jonas Feld, Quality Director
"The vendor PoC on line 4 last year was promising — 18,000 labelled surface images, and it caught most defects the inspectors catch. Line 5 has different camera positions. One camera on line 4 also sees the operators' hands and faces."

Miriam Hofer, Head of Service
"My 120 technicians spend ages searching manuals and old tickets. First-time fix is 78%. We pay about €450k a year for subcontracted service hours in peak season. The Czech manuals are badly out of date. The service share drive also has all customer contracts in it."

Stefan Brandt, CIO
"Copilot for all 850 office users — €30 per seat per month — would save an hour a week each. That's €2m of value. Also, the ERP vendor switched on 'AI insights', and HR's suite has some CV-ranking feature that went live in an update."

Lukas Maier, Head of Production Planning
"Nobody asked us about the scheduling agent. A quarter of operations have no actual times in MES because operators book at shift end. We have rules about customers, tooling and people that aren't in any system."

Petra Schulz, Works Council Chair
"We'll support AI that makes work better. We need an agreement before any system can monitor performance — cameras and anything that logs what people do. And nobody gets ranked by a machine for an apprenticeship."

Erik Lund, Head of Sales Engineering
"Quotes take nine days. Half the work is finding similar past designs. But our drawings are the crown jewels — I know people paste them into public chatbots to get descriptions written."

1.4 Case dossier: data extracts and AI landscape

Area	Data point (source)
Quality	Scrap + rework: line 4 €1.9m, line 5 €0.7m per year; customer claims €0.4m/yr (controlling)
Quality inspection	210,000 parts/yr on lines 4–5; PoC: 18,000 labelled images (line 4 only); PoC recall 92% on vendor-selected test set (E2–E3)

Area	Data point (source)
Maintenance	Grinding machines: 38 CNC; sensors on 40%; 37% of 2025 work orders lack failure codes (CMMS profile)
Planning	MES: 23% of operations lack actual times; 14 planners; on-time delivery 84%
Energy	Heat treatment (Plauen) energy €3.2m/yr; furnace sensor data complete since 2023; vendor references 4–8% savings
Service	9,000 visits/yr; first-time fix 78%; subcontracted hours €450k/yr; manuals: German current, Czech 40% outdated
Spare parts	€27m inventory; obsolete write-offs €1.1m/yr; part master differs between plants until ERP migration
Sales engineering	1,400 custom quotes/yr; 9-day cycle; 22% win rate; 18 engineers
Finance AP	95,000 supplier invoices/yr; 35% already structured e-invoices, rising as German B2B e-invoicing rolls out; vendor claims €140k/yr AI-extraction saving
HR	≈ 400 apprentice applications/yr; HR suite CV-ranking feature switched on by update in May

AI landscape finding	Source channel
310 of 640 survey respondents use public chatbots weekly; ~70 admit pasting drawings or customer specs	Observed use (anonymous survey)
Quality inspectors keep an Excel defect catalogue with photos to train new staff	Observed use (shop-floor visit)
ERP “AI insights” (forecast suggestions) enabled; nobody uses them	Standing audit (SaaS admin review)
HR suite CV-ranking feature live since May; used on 60 applications	Standing audit
Vendor PoC on line 4 completed; no owner for next step	Solicited (Quality interview)
Heat-treatment energy flagged in energy audit as top cost item	Standing audit (energy review)

1.5 Exercise 1 · Frame the ask and the decision rights

EXERCISE 1 **From "an AI plan" to the real problems** ⏱ 45 min

Draws on: C1 M1 (ask → problem, decision rights), C4 M4 (scope)

Inputs: Interviews and strategic context (1.2–1.3)

YOUR TASK

1. Write the CEO's stated ask verbatim and the trigger behind it.
2. Reframe it into 2–3 underlying business problems with metrics and baselines.
3. State the engagement's decision: what the board must decide, by when.
4. Draft a decision-rights register for five key decisions.
5. Write three scope exclusions.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 1

Stated ask: "Make Brenner the AI leader in our niche — something like the scheduling agent by next year; give everyone Copilot." **Triggers:** board question in March; competitor announcement; trade-fair demo.

Underlying problems: (1) margin erosion — EBIT 6.1% vs 8.4%; (2) quality cost on lines 4–5 (€2.6m scrap and rework + €0.4m claims); (3) service productivity and first-time fix (78%; €450k subcontracting); (4) uncontrolled AI use putting IP at risk (≈ 70 staff pasting drawings into public tools).

Engagement decision: "Which AI initiatives should Brenner fund in the next 18 months, in what order, under what rules — and which should it decline?" Decision-holder: CEO with CFO, at the advisory-board meeting in week 12.

Decision	Accountable	Consult
Portfolio and funding envelope	CEO	CFO, COO, advisory board
Tranche release at gates	CFO	COO, use-case owners
Rules of use and data boundaries (AI policy)	CEO (policy owner: CIO + DPO)	Works council, Legal, Sales Eng
Monitoring-capable systems (cameras, logs)	COO, subject to works agreement	Works council, DPO
Status of HR CV-ranking feature	Head of HR	Works council, Legal, DPO

Exclusions: vendor selection and contract negotiation; technical build; ERP migration itself.

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Reframing	Restates the ask	Names business problems	Problems with metrics, baselines and owners
Decision	Vague ("AI strategy")	Names a decision	Decision, holder, date, and includes declining
Decision rights	Committee as owner	Named roles	Named roles incl. works council and HR feature

1.6 Exercise 2 · Opportunity inventory with value hypotheses and ledgers

EXERCISE 2 **Ten candidates, three channels, honest value** ⏱ 60 min

Draws on: C2 M1 (work-pull, value hypothesis, ledger), C1 M2 (classes, shadow AI)

Inputs: Dossier 1.2–1.4

YOUR TASK

1. List at least ten candidate opportunities and tag each with its source channel.
2. Write a value hypothesis for each (mechanism → metric → € → landing line).
3. For three candidates, name the ledger: who verifies, corrects, escalates or waits.
4. Flag which candidates are claims you cannot yet trace to a landing line.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 2

ID	Candidate	Channel	Value hypothesis (planning range)	Traceable?
O1	Visual surface inspection, lines 4–5	Solicited + observed	Fewer escapes and less rework → scrap/ rework CC + claims; €400–600k/yr	Yes (Quality)
O2	Service-technician assistant (manuals + tickets)	Solicited	Less search time; FTF +3 pts → subcontract hours and repeat visits; €150–260k/yr	Partly (subcontract budget)
O3	Heat-treatment furnace energy optimisation	Standing audit	4–6% energy → Plauen energy CC; €130–190k/yr	Yes (Plant)
O4	Predictive maintenance, grinding	Solicited	Less unplanned downtime → output, overtime; €300–700k/yr (weak data)	Partly
O5	Spare-parts demand forecasting	Standing audit	Lower obsolescence write-offs (€1.1m/yr) by 15–25%	Yes (after ERP)
O6	Quote drafting from past designs	Solicited	Shorter cycle → win rate; new value, uncertain	No (new value)
O7	Autonomous production-scheduling agent	Solicited (CEO)	Vendor demo claims +8% output; evidence E2	No
O8	AI CV screening for apprentices	Standing audit (live feature)	≈ €25k/yr HR time	Yes, tiny
O9	Supplier-invoice extraction	Solicited (vendor)	Vendor claims €140k/yr; shrinking as e-invoicing rises	Declining
O10	Copilot for all 850 office staff	Solicited (CIO)	Claimed €2.15m “time value”; hard landing line ≈ none; real driver = shadow-AI risk	No (soft)

Ledgers (examples): O1 — quality inspectors *verify* flagged parts (≈ 6% of 210,000 = 12,600/yr); false rejects *wait* for re-inspection. O2 — technicians *verify* answers; content owners *correct* manuals; complex cases *escalate* to engineering. O7 — planners *verify and override* schedules; supervisors *escalate* conflicts; customers *wait* if delivery promises slip.

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Coverage	< 8 candidates; interviews only	≥ 10; two channels	≥ 10; three channels incl. shadow and embedded AI
Value hypotheses	“Productivity”	Mechanism and metric	Mechanism, metric, range, landing line
Honesty	Takes vendor/CIO claims at face value	Flags some	Flags every untraceable claim explicitly

1.7 Exercise 3 · Screening: gates, label tests and risk by use

EXERCISE 3 **Separate what is worth study from what will stall** ⏱ 60 min

Draws on: C2 M2 (normalization, data gate, label test, risk by use, dispositions)

Inputs: Your inventory; dossier data extracts

YOUR TASK

1. Normalize O1, O4 and O7 (user, trigger, input, AI task, output, decision, autonomy, volume, metric).
2. Apply the data gate and, for predictive items, the label test.
3. Size risk by use for O1, O7 and O8.
4. Give every candidate a screening disposition with a reason.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 3

ID	Data gate / label test	Risk by use	Screening disposition
O1	Pass for line 4 (18k labelled images); conditional for line 5 (camera geometry)	Tier 2: product quality; human re-inspection; camera sees operators → works agreement	Advance (line 4) with condition for line 5
O2	Conditional: Czech manuals outdated; permissions overshare customer contracts	Tier 2: internal; logs → works agreement	Advance with conditions
O3	Pass: furnace data complete since 2023	Tier 1	Advance
O4	Fail now: 37% missing failure codes; sensors on 40% of machines	Tier 2	Park — trigger: 6 months of coded failures on instrumented machines
O5	Conditional: part master inconsistent until ERP migration (Q3 2027)	Tier 1	Park with trigger (ERP migration)
O6	Conditional: past designs exist; drawings are restricted IP	Tier 2; IP exposure	Advance with conditions (data boundary, owner)
O7	Fail: 23% missing actual times; tacit rules not recorded (label test for “good schedule” fails)	Tier 3: autonomous actions affecting delivery promises	Do not advance as specified; reframe as decision-support experiment
O8	Pass technically	High-risk use (employment, EU AI Act Annex III); works council objects; value ≈ €25k	Reject; switch off live feature pending assessment
O9	Pass	Tier 1	Redirect: structured e-invoice processing (ordinary automation)
O10	n/a	Tier 2 (data boundary)	Redirect: governed assistant for targeted groups

EXAM TRAP · Where most learners slip

Treating the live HR CV-ranking feature as a future "opportunity" to score. It is a **live, unassessed high-risk use** — a governance action is needed now, regardless of portfolio scoring.

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Gates	Scores everything	Applies data gate	Applies data gate, label test and prohibited/high-risk screen before scoring
Risk sizing	By technique	By use	By use, including works-council and live-feature implications
Dispositions	Missing for some	All, with reasons	All, with reasons and revisit triggers

1.8 Exercise 4 · Prioritize into a balanced portfolio

EXERCISE 4 **Value, effort, balance and dispositions** ⏱ 60 min

Draws on: C2 M3–M4 (criteria first, effort by component, error pricing, matrix, balance, consolidation)

Inputs: Screened candidates; CFO's parameters

YOUR TASK

1. Fix criteria and weights before scoring (write them down).
2. Score value and effort by component for the advanced candidates.
3. Place them on a value–effort matrix and assign portfolio dispositions.
4. Run two balance checks (capacity and risk) and one consolidation check.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 4

Criteria (fixed first): value = net financial (40), risk reduction (25), strategic fit (20), time to value (15). Effort components: data, integration, compliance, adoption, operational capacity (1–5 each, equal weights).

ID	Value (1–5)	Effort D/I/C/A/O	Effort	Quadrant	Disposition
O1 Inspection (L4)	5	2/3/2/2/2	2.2	Quick win	Fund now (line 4); line 5 after camera fix
O2 Service assistant	3	3/2/2/2/2	2.2	Quick win / learn	Fund pilot (30 technicians, DE)
O3 Furnace energy	3	1/2/1/1/1	1.2	Quick win — but marginal economics	Fund a test (one furnace) before full spend
O4 Predictive maintenance	4	5/3/1/2/3	2.8	Strategic bet	Prepare: fund failure-code capture and sensors
O5 Spare-parts forecast	3	3/3/1/2/2	2.2	Prepare	H2 after ERP migration
O6 Quote drafting	3 (uncertain)	3/3/3/3/2	2.8	Strategic bet	Defer to H2; prerequisites: data boundary, owner
O7 Scheduling agent	2 (after evidence haircut)	5/4/2/5/4	4.0	Money pit as specified	Not now; small decision-support experiment in H2
O8 CV screening	1	2/2/5/4/2	3.0	Money pit	Reject
O9 Invoice extraction	1–2 (declining)	2/2/1/1/1	1.4	Fill-in	Redirect to e-invoicing automation
O10 Copilot all staff	2 (risk)	1/1/2/3/2	1.8	—	Redirect: governed assistant, 120 targeted seats

Balance: capacity — Quality and IT integration both needed by O1 and O2; stagger O2 integration by six weeks. Risk — no high-risk uses in wave 1 (O8 rejected). **Consolidation:** O2, O6 and O10 all need a governed retrieval assistant with data boundaries → one shared enabler ("Brenner Assistant"), configured per group.

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Discipline	Weights after scoring	Weights before scoring	Weights before scoring, with sensitivity check
Effort	Single number	Components	Components naming who supplies the effort
Portfolio	Top-N list	Dispositions	Dispositions, balance checks and consolidation

1.9 Exercise 5 · Readiness summary

EXERCISE 5 **Proceed, phase or not yet — with evidence grades** ⏱ 60 min

Draws on: C3 (seven dimensions, model vs organization risk, conditions not assurances, evidence grading)

Inputs: Dossier; your portfolio

YOUR TASK

1. Record the standard and date your readiness is measured against.
2. For O1, O2, O3 and O7, rate the seven dimensions and name the weakest link.
3. Grade the key evidence (V1–U).
4. Write a readiness finding for each with remediation and sequence.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 5

Baseline: measured against Brenner's draft AI Policy v0.9 (to be approved in week 10), NIST AI RMF 1.0, ISO/IEC 42001 (selected controls) and the EU AI Act as amended by Regulation (EU) 2026/1744, as of week 6 of the engagement.

	O1 Inspection	O2 Service assistant	O3 Furnace	O7 Scheduling agent
Data	Strong L4 (V1 sample of 300 labels) / weak L5	Partial: CZ manuals outdated (V2)	Strong (V1 profile)	Weak: 23% missing actuals (V1)
Technology	Partial: MES interface exists	Partial: DMS API; no permission-aware retrieval yet	Strong	Weak: MES write-back untested
Ownership	Strong: Quality Director (V2)	Strong: Head of Service	Strong: Plauen plant manager	Weak: no owner; planners not consulted
Adoption	Partial: inspectors keen (A1)	Partial: technicians keen (A1); CZ language	n/a (operator advisory)	Weak: planners sceptical (A1)
Governance	Weak: no works agreement (cameras)	Weak: oversharing of contracts (V1)	Partial	Weak
Measurement	Strong: claims ppm, scrap CC baseline	Partial: FTF baseline; no time study	Strong: metered energy	Weak: OEE baseline disputed
Trust	Unknown at L5	Unknown	Partial	Weak (E2 demo only)
Finding	**Phase**: L4 proceed after works agreement; L5 not yet	**Phase**: DE pilot after permission fix and works agreement; CZ after manual update	**Proceed** with one-furnace test	**Not yet**

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Baseline	None	Standard named	Standard, version and date
Dimensions	Averaged	Each rated	Each rated, weakest link named, evidence graded
Findings	Yes/no	Proceed/phase/not yet	With remediation, owners and sequence

1.10 Common mistakes in this phase

✓ DO THIS

- Spend the first week on problems, not solutions.
- Visit the shop floor — observed use found two candidates.
- Treat live embedded features as governance actions now.
- Flag every claim you cannot trace to a landing line.
- Give every candidate a disposition, including “no”.
- Grade evidence before writing findings.

× AVOID THIS

- Starting with the CEO’s favourite and working backwards.
- Accepting the CIO’s €2m Copilot “value”.
- Scoring the CV-ranking feature as an opportunity.
- Using the vendor PoC recall as production evidence.
- Treating “prepare” items as rejected.
- Averaging readiness into one score.

1.11 Module 1 cheat sheet

MODULE 1 · ONE-PAGE SUMMARY

Opportunity Assessment, Portfolio and Readiness

ENGAGEMENT ARC

Discover (1–3) · assess (4–6) · plan & justify (7–10) · integrate (11–12). Steering points at each stage; board in week 12.

FRAME

Ask verbatim → trigger → business problems with baselines → engagement decision → decision rights → exclusions.

INVENTORY

Three channels. Value hypothesis with landing line. Ledger. Flag untraceable claims.

SCREEN

Normalize · data gate · label test · risk by use · prohibited/ high-risk screen · dispositions with triggers.

PORTFOLIO

Criteria first · effort by component · matrix · dispositions for all · balance · consolidation.

READINESS

Baseline (standard, date) · seven dimensions · weakest link · graded evidence · proceed / phase / not yet.

BRENNER AT A GLANCE (MODEL)

Fund now: inspection L4, service assistant pilot, furnace test. Prepare: maintenance data, spare-parts forecast (after ERP), quote drafting (after data boundary). Not now: scheduling agent. Reject: CV screening (and switch off live feature). Redirect: invoice extraction → e-invoicing; Copilot-for-all → governed assistant for 120 targeted users.

1.12 Knowledge check

Answer all questions before turning the page.

K1 **Which is the best reframing of Dr Brenner's ask?**

(A) "Roll out Copilot to all staff."
(B) "Deploy the scheduling agent by next year."
(C) "Become the AI leader."
(D) "Decide which AI initiatives to fund over 18 months, in what order and under what rules — and which to decline — to address margin, quality, service and IP-risk problems."

K2 **Which channel surfaced the IP risk of drawings pasted into public chatbots?**

(A) Observed use (anonymous survey), corroborated by Sales Engineering
(B) ERP admin review
(C) Solicited input from the CEO
(D) Vendor briefing

K3 **How should the CIO's "€2m value" for Copilot-for-all be treated?**

(A) Ignore Copilot entirely
(B) As an untraceable soft claim; the real driver is shadow-AI risk, which a governed, targeted assistant can address
(C) As hard ROI
(D) As option value

K4 **Predictive maintenance fails the data gate (37% missing failure codes). What is the right disposition?**

(A) Fund full deployment now
(B) Reject permanently
(C) Park or prepare, with an enabler to capture coded failures and a revisit trigger
(D) Ask the vendor to fill the gaps

K5 **The HR suite's CV-ranking feature is already live. What should happen?**

(A) Expand it to engineering hires
(B) Treat it as an immediate governance action: switch off pending assessment, involve HR, works council, legal and DPO
(C) Leave it because the vendor enabled it
(D) Score it in the portfolio

K6 **Why is the furnace optimisation disposed as "fund a test" rather than "fund now"?**

(A) The plant manager is opposed
(B) It is high-risk
(C) Its data fails the gate
(D) Its economics are marginal around the 4% saving level, so a cheap test should confirm the saving before full spend

K7 **Which readiness finding fits the inspection use case?**

(A) Proceed everywhere
(B) Reject
(C) Phase: line 4 after the works agreement; line 5 not yet
(D) Not yet everywhere

K8 **Why consolidate O2, O6 and O10 into one shared assistant enabler?**

(A) They need the same governed retrieval capability with data boundaries; one platform avoids duplicate automation and fragmented controls
(B) To reduce the number of slides
(C) Because they are all high-risk
(D) Because they have the same owner

1.13 Knowledge check: answers and explanations

K1 **Answer: D** — Review: Exercise 1

Which is the best reframing of Dr Brenner's ask?

- ✗ A "Roll out Copilot to all staff." — **Wrong.** A solution, not a problem.
- ✗ B "Deploy the scheduling agent by next year." — **Wrong.** That repeats the solution-shaped ask.
- ✗ C "Become the AI leader." — **Wrong.** An aspiration without a decision.
- ✓ D "Decide which AI initiatives to fund over 18 months, in what order and under what rules — and which to decline — to address margin, quality, service and IP-risk problems." — **Correct.** A decision tied to underlying problems.

K2 **Answer: A** — Review: Exercise 2

Which channel surfaced the IP risk of drawings pasted into public chatbots?

- ✓ A Observed use (anonymous survey), corroborated by Sales Engineering — **Correct.** Observed use reveals shadow AI.
- ✗ B ERP admin review — **Wrong.** That found ERP AI insights.
- ✗ C Solicited input from the CEO — **Wrong.** The CEO did not raise it.
- ✗ D Vendor briefing — **Wrong.** No.

K3 **Answer: B** — Review: Exercise 2, 4

How should the CIO's "€2m value" for Copilot-for-all be treated?

- ✗ A Ignore Copilot entirely — **Wrong.** The shadow-AI risk is real and needs a response.
- ✓ B As an untraceable soft claim; the real driver is shadow-AI risk, which a governed, targeted assistant can address — **Correct.** Honest classification and reframing.
- ✗ C As hard ROI — **Wrong.** There is no landing line or owner.
- ✗ D As option value — **Wrong.** Not the right category.

K4 **Answer: C** — Review: Exercise 3–4

Predictive maintenance fails the data gate (37% missing failure codes). What is the right disposition?

- ✗ A Fund full deployment now — **Wrong.** It would stall.
- ✗ B Reject permanently — **Wrong.** The value is real; the gap is fixable.
- ✓ C Park or prepare, with an enabler to capture coded failures and a revisit trigger — **Correct.** Readiness remediation becomes an enabler.
- ✗ D Ask the vendor to fill the gaps — **Wrong.** Labels cannot be invented.

K5 **Answer: B** — Review: Exercise 3

The HR suite's CV-ranking feature is already live. What should happen?

- ✗ A Expand it to engineering hires — **Wrong.** Increases risk.
- ✓ B Treat it as an immediate governance action: switch off pending assessment, involve HR, works council, legal and DPO — **Correct.** Live high-risk uses need action now.
- ✗ C Leave it because the vendor enabled it — **Wrong.** Deployers are responsible for their use.
- ✗ D Score it in the portfolio — **Wrong.** It is a live, unassessed high-risk use.

K6 **Answer: D** Review: Exercise 4; Module 2

Why is the furnace optimisation disposed as "fund a test" rather than "fund now"?

✗ A The plant manager is opposed — **Wrong.** The manager is keen.

✗ B It is high-risk — **Wrong.** It is Tier 1.

✗ C Its data fails the gate — **Wrong.** Its data passes.

✓ D Its economics are marginal around the 4% saving level, so a cheap test should confirm the saving before full spend — **Correct.** Evidence should precede commitment.

K7 **Answer: C** Review: Exercise 5

Which readiness finding fits the inspection use case?

✗ A Proceed everywhere — **Wrong.** Line 5 and the works agreement are not ready.

✗ B Reject — **Wrong.** The value is strong.

✓ C Phase: line 4 after the works agreement; line 5 not yet — **Correct.** A ready subset exists.

✗ D Not yet everywhere — **Wrong.** Line 4 is ready subject to one condition.

K8 **Answer: A** Review: Exercise 4

Why consolidate O2, O6 and O10 into one shared assistant enabler?

✓ A They need the same governed retrieval capability with data boundaries; one platform avoids duplicate automation and fragmented controls — **Correct.** Consolidation.

✗ B To reduce the number of slides — **Wrong.** Not the reason.

✗ C Because they are all high-risk — **Wrong.** They are not.

✗ D Because they have the same owner — **Wrong.** They do not.

MODULE 2

Roadmap, Business Case and Governance

Turn the portfolio into a precondition-based roadmap, an ROI business case finance can verify and a governance plan that works at real volume — and write evidence-based recommendations against the ideas the evidence does not support, including the CEO's favourite.

YOU WILL BE ABLE TO

- Build a roadmap sequenced by checkable preconditions with tranche-based funding
- Frame an agent opportunity and recognise when to recommend against it
- Build an ROI model with hidden costs, scenarios, sensitivity and breakeven
- Write a measurement plan with baselines, paired metrics and attribution
- Write a governance and risk plan tied to classification, with a human-oversight volume test
- Pressure-test the plan and choose dispositions
- Write credible null recommendations

IN THIS MODULE

1. Exercise 6 · Roadmap and tranches
2. Exercise 7 · The scheduling agent
3. Exercise 8 · ROI model and business case
4. Exercise 9 · Measurement plan
5. Exercise 10 · Governance and risk plan
6. Exercise 11 · Pressure test
7. Recommending against: four null recommendations
8. Common mistakes · cheat sheet · knowledge check

Suggested study time · 10–12 hours (incl. exercises)

With a portfolio and readiness findings in hand, the engagement turns to **planning and justification**. This is where advisory credibility is won or lost: the roadmap must start initiatives only when they are ready, the business case must survive the CFO, the governance plan must work at real volume — and the package must say **no** clearly where the evidence does not support an idea, even when the idea belongs to the CEO.

KEY TAKEAWAYS

1. Sequence by **preconditions**; fund by **tranches** released at gates with observable criteria.
2. Frame agents by **agentic surface, reach, reversibility, ownership and residual work** — then decide honestly.
3. Build the ROI model with **hidden costs**, **production-adjusted benefits**, scenarios and **breakeven**.
4. Measure against **baselines** with **paired quality metrics** and a stated **attribution level**.
5. Tie governance to **classification**, works agreements and a **human-oversight volume test**.
6. Pressure-test, then choose **dispositions**.
7. Recommend **against** unsupported ideas with numbers, alternatives and revisit triggers.

2.1 Exercise 6 · Roadmap with preconditions and tranches

EXERCISE 6 **From portfolio to a fundable plan** ⏱ 75 min

Draws on: C4 M1–M2 (preconditions, horizons, capacity, operating model, tranches)

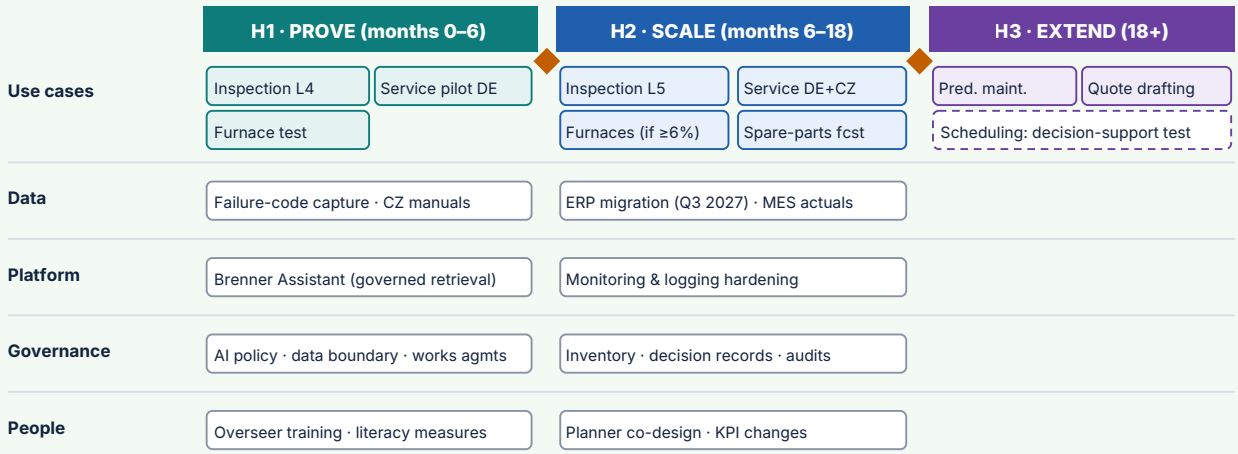
Inputs: Your portfolio and readiness findings

YOUR TASK

1. Lay out H1–H3 with use cases and enabler workstreams.
2. Write checkable start preconditions for four initiatives.
3. Define Gate 1 criteria and a Tranche 1 within the CFO's envelope.
4. Choose an operating model and assign ownership.
5. Identify the binding capacity constraint and a WIP limit.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 6



Model roadmap — Brenner AI programme

Initiative	Starts when (checkable preconditions)
Inspection line 4	Works agreement on camera use signed (operator area masked); Quality Director confirmed as owner; baseline (claims ppm, scrap CC) signed
Service assistant pilot (30 DE technicians)	Customer contracts removed from indexed scope and permissions fixed; works agreement on logs; answer-citation requirement configured
Full furnace rollout	One-furnace test shows ≥ 6% metered saving over 8 weeks at comparable load
Predictive maintenance pilot	≥ 6 months of coded failures (≥ 85% coded) on instrumented machines
Scheduling decision-support test	MES actual-time capture ≥ 95% for 3 months; planners co-design; owner named (Head of Production Planning)

Tranche	Amount	Funds	Gate to next
T1 (now)	€0.52m	Inspection L4 (180), service pilot (90), governance basics incl. governed assistant (120), furnace test (60), failure-code capture (70)	Gate 1 (month 6): escapes −25% vs baseline on L4 with false-reject ≤ 3%; service answer acceptance ≥ 70% and FTF +2 pts vs comparison group; furnace ≥ 6%; policy approved; works agreements signed
T2	≈ €0.6m	Inspection L5, service rollout DE+CZ, furnace rollout (if ≥ 6%), monitoring hardening	Gate 2 (month 12): value tracked in controlling; HITL tests passed at scaled volume
T3	≈ €0.35m	Spare-parts forecasting, PdM pilot, quote-drafting pilot (on preconditions)	Per-initiative gates

Envelope: ≈ €1.5m over 18 months incl. run costs and 10% contingency — inside the CFO's €2m. **Operating model:** hub-and-spoke — CIO hosts the platform and policy (hub); plants and service own use cases (spokes). **Binding constraint:** one MES/DMS integration engineer → WIP limit 2 integrations at a time; service integration starts six weeks after inspection.

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Preconditions	Dates only	Some conditions	Observable, verifiable, owned conditions for each start
Funding	Full envelope up front	Tranches	Tranches with observable gate criteria and stop rules
Capacity & model	Ignored	Mentioned	Binding constraint named; WIP limit; hub-and-spoke ownership

2.2 Exercise 7 · Framing the scheduling agent — and deciding against it

EXERCISE 7 **The CEO's favourite idea** ⏱ 60 min

Draws on: C4 M3 (agentic surface, reach, reversibility, ownership, residual work), C1 M5 (evidence), C5 M4 (null recommendation)

Inputs: CEO, COO and Planning interviews; MES data; vendor demo claims

YOUR TASK

1. Read the vendor claim back to its conditions.
2. Map the scheduling workflow and isolate any agentic surface.
3. Assess reach, reversibility, ownership and residual work.
4. Decide: fund, test, defer or reject — and write the recommendation for the CEO.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 7

Claim read-back: "+8% output" came from a vendor demo on a reference plant with clean, complete MES data and no tacit rules (evidence E2). Brenner has 23% of operations without actual times and undocumented customer, tooling and staffing rules. The claim does not transfer.

Framing factor	Assessment
Agentic surface	Sequencing on two bottleneck grinding cells, where trade-offs are complex — plausible for suggestions; the rest of scheduling is rule-based
Reach	All orders and customer delivery promises — wide
Reversibility	Reversible until released to the shop floor; costly afterwards (set-ups, missed deliveries)
Ownership	No owner; planners not consulted; cross-plant
Residual work	Planners verify and override; tacit rules must be captured first
Readiness	Not yet (data, ownership, adoption)

> **Recommendation to the CEO — not now, with a route back.** Do not procure an autonomous scheduling agent in this programme. The demo's 8% depends on data Brenner does not yet have, and on planning rules nobody has written down. Instead: (1) fix MES actual-time capture (in the data workstream); (2) in H3, run a 10-week, ~€60k decision-support test at Heilbronn where the tool *suggests* bottleneck sequences that planners accept or reject (autonomy level 1). **What would change this:** MES actuals ≥ 95% for three months and the test showing ≥ 3% throughput gain on bottlenecks with planner acceptance ≥ 60%.

KEY IDEA

This is the engagement's **hardest finding**: it contradicts the CEO's visible enthusiasm. Module 4 shows how to document it; Module 3 shows how to present it without losing the room.

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Evidence	Repeats vendor claim	Questions it	Reads it back to conditions; grades it
Framing	“Agent good/bad”	Some factors	Surface, reach, reversibility, ownership, residual work
Recommendation	Avoids a position	Says no	Says not now, with alternative, criteria and revisit trigger

2.3 Exercise 8 · ROI model and business case

EXERCISE 8 **A case the CFO can verify** ⏱ 90 min

Draws on: C5 M1–M2 (traceability, pilot adjustment, TCO, hidden costs, NPV, scenarios, breakeven)

Inputs: Data extracts; CFO parameters (9%, 5 years); your roadmap

YOUR TASK

1. Trace each wave-1 benefit to a landing line and owner.
2. Adjust the inspection PoC for production.
3. Build TCO including hidden costs.
4. Compute NPV, payback and IRR per use case and for the wave; run conservative and upside scenarios and breakeven.
5. State what the numbers say about each use case.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 8 — benefits and adjustments

Benefit	Mechanism	Landing line / owner	€k / yr (steady)
Inspection: scrap and rework	€2.6m × 18% planning reduction	Scrap & rework CC, Heilbronn / Quality Director	468
Inspection: customer claims	€0.4m × 15%	Warranty & claims / Quality Director	60
Service: subcontracted hours	25% of €450k (owner commits budget cut after Gate 2)	Service subcontracting / Head of Service	≈ 112
Service: repeat visits	FTF +3 pts × 9,000 visits ≈ 270 fewer × ≈ €360	Service cost / Head of Service	≈ 97
Furnace (if test ≥ 6%)	€3.2m × 6% (excluded from wave-1 NPV until test)	Energy CC, Plauen	(192)

Inspection pilot-to-production adjustment: PoC suggested ~35% reduction potential → × 0.85 (vendor-selected test set) × 0.85 (line 5 geometry and mix) × 0.80 (adoption) ≈ 20% − 2 pts ledger (false rejects re-inspected) ≈ **18%**.

MODEL ANSWER · Exercise 8 — costs and results (€k, 9%, 5 years, year-1 benefits at 50%)

	Inspection	Service assistant	Furnace (at 4%)	Wave 1 package*
One-off	380	220	260	780
Recurring / yr (licences, inference, review, upkeep, governance)	95	110	45	265
Steady benefit / yr	528	210	128	738
NPV	1,062	73	≈ 4	**721**

	Inspection	Service assistant	Furnace (at 4%)	Wave 1 package*
Payback	1.5 yrs	3.3 yrs	3.9 yrs	**2.4 yrs**
IRR	77%	19%	10%	**35%**

*Wave 1 = inspection + service assistant + governance enabler (one-off 120, recurring 60) + furnace test (60, no benefit counted).

Scenario (wave 1)	Assumptions	NPV	Payback
Conservative	Benefits −25%, costs +15%	−183	4.7 yrs
Base	As above	721	2.4 yrs
Upside	Benefits +20%	1,228	2.0 yrs

Breakeven: NPV = 0 at about **72% of planned benefits**. **What the numbers say:** inspection carries the case; the service assistant is marginal financially (conservative NPV ≈ −€205k) and is justified as a pilot that also builds the shared assistant; the furnace breaks even at a ~4% saving, so a €60k test must show ≥ 6% before the €260k rollout. The conservative scenario is negative → **staged funding**, not full commitment.

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Traceability	Hours × salary	Landing lines	Landing lines, owners and budget commitments
Costs	Licences only	TCO	TCO with review, upkeep, inference, governance
Analysis	Base NPV only	Scenarios	Scenarios, breakeven and use-case-level conclusions

2.4 Exercise 9 · Measurement plan

EXERCISE 9 **Making value provable** ⏱ 45 min

Draws on: C5 M3 (metric hierarchy, baselines, paired metrics, comparison designs, attribution, targets)

Inputs: Your business case and roadmap

YOUR TASK

1. Define a metric hierarchy for inspection and service.
2. Pair every productivity metric with a quality metric.
3. Choose comparison designs and state attribution levels.
4. Set year-1 and year-2 targets aligned with the ramp.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 9

Use case	Value metric (owner's report)	Paired quality metric	Design / attribution	Y1 target	Y2 target
Inspection	Scrap + rework € (monthly quality report); claims ppm	False-reject rate ≤ 3%; inspector re-check time	Line 4 before/after with line 5 as comparison until L5 goes live (A2–A3)	−10% scrap/ rework L4	−18% L4+L5
Service	Subcontract € ; FTF rate (service KPI pack)	Repeat-visit rate; customer complaints	Pilot group vs comparison group of technicians; staggered DE→CZ (A3)	FTF +2 pts pilot	FTF +3 pts; subcontract −25%
Furnace test	kWh per tonne treated (metered)	Hardness spec compliance	Alternating weeks on/ off on one furnace (A3–A4)	≥ 6% in test	Rollout target per furnace
Programme	Adoption & usage quality (leading)	Override / edit rates	—	≥ 75% eligible use	Stable edit rates

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Hierarchy	Usage only	Outcomes	Usage → quality → outcome → € in owners' reports
Rigour	Before/after only	Comparison groups	Design matched to roadmap; attribution stated
Targets	Year 2 only	Y1 and Y2	Y1/Y2 aligned with ramp and gates

2.5 Exercise 10 · Governance and risk plan

EXERCISE 10 **Rules that work at real volume** ⏱ 75 min

Draws on: C6 M1–M2 (instruments, classification → obligations, data boundary, HITL tests, failure statements, buildable controls)

Inputs: Landscape findings; works council and DPO positions; roadmap

YOUR TASK

1. Map instruments and classify wave-1 systems plus the HR feature.
2. Draft the data boundary for drawings and customer data.
3. Write three specific failure statements and buildable controls.
4. Run the HITL capacity calculation for inspection and design a seeded-error test.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 10

System	Classification (EU AI Act, as amended)	Internal tier	Key obligations / actions
Inspection (L4)	Not high-risk as a quality-control use (confirm with counsel that it is not a safety component under product law)	Tier 2	Works agreement (cameras capture operators; masking); DPIA screening; oversight spec; monitoring
Service assistant	Minimal risk; internal use	Tier 2	Data boundary; permission-aware retrieval; works agreement on logs; literacy measures
Furnace optimisation	Minimal risk	Tier 1	Change control on setpoints; process-engineer approval
HR CV ranking (live)	High-risk (employment, Annex III; obligations from 2 Dec 2027) — and objected to by works council	Tier 4 (stop)	Switch off now; decision record; legal/ DPO review
Governed assistant (120 seats)	Minimal risk; transparency where content is AI-generated externally	Tier 2	Approved tool replaces public chatbots; DLP for drawings

Data boundary: technical drawings and customer specifications = *Restricted (IP)* → never in public tools; only in the governed assistant's Sales-Engineering space with access by role; outputs inherit the classification.

Specific failure	Buildable control
Inspection model passes cracked gears on line 5 because camera geometry differs from line 4 training data → customer claims	Guardrail: model active only on validated line(s); L5 activation requires validation ≥ 95% recall on 2,000 L5 images; owner: Quality Director; log: activation record; test: monthly seeded defects
Service assistant cites superseded Czech manual → wrong repair, repeat visit	Answers must cite a current, owner-approved document; outdated documents excluded from index; weekly regression of 50 questions

Specific failure	Buildable control
Staff paste drawings into public chatbots → IP leak	DLP block on public AI domains for drawing file types; governed assistant available; monthly shadow-use report

HITL capacity (inspection): ~6% of 210,000 parts flagged = 12,600/yr ≈ 1,050/month × 1.5 min ≈ **26 hours/month** — well within inspectors' capacity (pass). **Seeded-error test:** 40 known-defect images seeded per quarter into the review queue; target ≥ 90% caught; agreed with the works council; results used to improve tools, not to assess individuals.

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Classification	Labels only	Classes with obligations	Classes → obligations → owners, incl. live HR feature
Controls	Vague	Specific	Buildable: trigger, check, action, owner, log, test
Oversight	Assumed	Capacity calculated	Capacity and seeded-error effectiveness test designed

2.6 Exercise 11 · Pressure-test the plan

EXERCISE 11 **Attack your own roadmap** ⏱ 45 min

Draws on: C4 M4 (assumptions vs dependencies, premortem, control load, dispositions)

Inputs: Your roadmap, case and governance plan

YOUR TASK

1. List the top five assumptions and three fragile dependencies.
2. Run a 20-minute premortem ("it is month 18 and the programme failed because…").
3. State concrete consequences and choose dispositions.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 11

Weakness	Type	Concrete consequence	Disposition
Head of Service will cut subcontract budget only if FTF improves	Assumption	≈ €112k/yr benefit not captured; service NPV negative	Correct: make the budget cut a Gate 2 condition signed by owner
Inspection model generalises to line 5	Assumption	L5 value (≈ 27% of inspection benefit) delayed or lost	Resequence: L5 only after validation; case shows L4-only NPV
Single integration engineer for MES and DMS	Fragile dependency	Both wave-1 integrations stall for months if he leaves	Fallback: contractor framework; document interfaces
Works agreements take longer than 8 weeks	Fragile dependency	Inspection and service starts slip; Gate 1 moves	Resequence: start negotiations in week 2; furnace test (no worker data) proceeds
Inspectors over-trust model at peak	Control load	Defects pass during peak months	Correct: seeded tests at peak; minimum review time KPI

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Classification of weaknesses	Generic risks	Assumptions vs dependencies	Separated, each with a different fix
Consequences	"Delay risk"	Some specifics	What, when, who, cost
Dispositions	None	Some	Lightest effective disposition for each, with owner

2.7 Recommending against: four null recommendations

A credible package says “no” or “not now” as clearly as “yes”. Each null recommendation below follows the structure from Course 5: plain statement, driving evidence, what is valuable anyway, what would change the answer, alternative, revisit trigger.

Idea	Recommendation	Driving evidence	Alternative	Revisit trigger
Autonomous scheduling agent (CEO)	Not now	E2 demo; 23% missing MES actuals; tacit rules; no owner	Decision-support test in H3 (~€60k)	MES actuals ≥ 95% for 3 months
AI CV screening (HR, live feature)	Reject; switch off	High-risk use; ≈ €25k value; works-council objection	Structured interview guides; faster scheduling of interviews	Only with a fresh, full assessment and works agreement
Supplier-invoice AI extraction	Redirect	Structured e-invoices rising from 35%; value declining	E-invoice processing via ordinary automation	If unstructured volume stays > 50% after 2028
Copilot for all 850 office staff	Redirect	No landing line for €2.15m claim; €306k/yr licences; the real issue is IP leakage	Governed assistant for 120 targeted users (≈ €43k/yr) + DLP	After 6 months of usage and value evidence

CONSULTANT'S TIP · Say no to the idea, yes to the problem

Every null recommendation above solves the underlying problem in a better way: delivery reliability (MES data, decision support), hiring speed (process), AP effort (e-invoicing), staff productivity and IP protection (governed assistant). That is what keeps sponsors on side.

2.8 Common mistakes in this phase

✓ DO THIS

- Tie every start to checkable preconditions.
- Let the conservative scenario shape the funding structure.
- Show use-case-level economics, not only the total.
- Make benefit capture a gate condition signed by owners.
- Classify live features, not just planned ones.
- Pair every “no” with a better route to the problem.

× AVOID THIS

- A Gantt chart that starts everything in Q1.
- Letting inspection’s strength hide the marginal service case.
- Counting the furnace saving before the test.
- Treating the works council as a late sign-off.
- “Human review” without a capacity number.
- A null recommendation without a revisit trigger.

2.9 Module 2 cheat sheet

MODULE 2 · ONE-PAGE SUMMARY

Roadmap, Business Case and Governance

ROADMAP

H1 prove · H2 scale · H3 extend; enabler lanes; preconditions; WIP limit on binding constraint; hub-and-spoke.

FUNDING

Envelope ≈ €1.5m / 18 months; T1 €0.52m now; Gate 1 at month 6 with observable criteria.

AGENT FRAMING

Claim read-back · surface · reach · reversibility · ownership · residual work · readiness → decide.

BUSINESS CASE

Landing lines and owners · pilot adjusted (35% → ~18%) · TCO · wave-1 NPV €721k, payback 2.4 yrs, IRR 35% · conservative −€183k · breakeven ~72%.

MEASUREMENT

Owners' reports · paired quality · comparison groups / staggered · attribution stated · Y1/Y2 targets.

GOVERNANCE

Classify (incl. live HR feature → stop) · data boundary (drawings restricted) · works agreements · buildable controls · HITL 26 h/month · seeded tests.

PRESSURE TEST

Assumptions vs dependencies · premortem · consequences · dispositions (correct, resequence, fallback…).

NULL RECOMMENDATIONS

Scheduling agent (not now) · CV screening (reject) · invoice AI (redirect) · Copilot-for-all (redirect). Each with alternative and trigger.

2.10 Knowledge check

Answer all questions before turning the page.

K1 **Which is a checkable precondition for the inspection start?**

- (A) "Q1 next year"
- (B) "Works agreement on camera use signed, with operator area masked"
- (C) "Model good enough"
- (D) "Stakeholders aligned"

K2 **Why is the wave-1 funding staged rather than committed in full?**

- (A) The base NPV is negative
- (B) Vendors require it
- (C) The conservative scenario is negative (≈ −€183k) and breakeven sits at ~72% of planned benefits, so evidence should precede full spend
- (D) The CFO dislikes AI

K3 **The furnace optimisation breaks even at about a 4% saving. What follows?**

- (A) Reject it
- (B) Add strategic value to justify it
- (C) Fund the full rollout now
- (D) Run a cheap one-furnace test; roll out only if it shows ≥ 6%

K4 **What is the strongest reason the scheduling agent is "not now"?**

- (A) The vendor's 8% depends on clean MES data and documented rules Brenner lacks (23% missing actuals; tacit rules), with no owner
- (B) It is high-risk under the EU AI Act
- (C) The CEO likes it
- (D) Agents are always too risky

K5 **Which item is a fragile dependency rather than an assumption?**

- (A) Line 5 will behave like line 4
- (B) FTF will rise by 3 points
- (C) Adoption will reach 75%
- (D) A single integration engineer handles both MES and DMS integrations

K6 **What should happen to the live HR CV-ranking feature?**

- (A) Keep it; the vendor enabled it
- (B) Switch it off, record the decision, and review with legal, DPO and works council
- (C) Expand it
- (D) Score it in H2

K7 **Inspection review needs ~26 hours a month. What else must be tested before relying on it?**

- (A) Effectiveness: a seeded-error test showing reviewers catch ≥ 90% of known defects
- (B) Vendor references
- (C) Nothing
- (D) Server speed

K8 **What makes the Copilot redirect credible to the CIO?**

(A) Blaming the vendor

(B) Saying AI assistants are useless

(C) It solves the real problem (IP leakage, productivity for targeted groups) at ≈ €43k/yr instead of €306k/yr, with a trigger to expand on evidence

(D) Delaying the decision

2.11 Knowledge check: answers and explanations

K1 **Answer: B** — Review: Exercise 6

Which is a checkable precondition for the inspection start?

- ✗ (A) "Q1 next year" — **Wrong.** A date.
- ✓ (B) "Works agreement on camera use signed, with operator area masked" — **Correct.** Observable and verifiable.
- ✗ (C) "Model good enough" — **Wrong.** Not thresholded.
- ✗ (D) "Stakeholders aligned" — **Wrong.** Vague.

K2 **Answer: C** — Review: Exercise 8

Why is the wave-1 funding staged rather than committed in full?

- ✗ (A) The base NPV is negative — **Wrong.** The base NPV is positive.
- ✗ (B) Vendors require it — **Wrong.** No.
- ✓ (C) The conservative scenario is negative (≈ −€183k) and breakeven sits at ~72% of planned benefits, so evidence should precede full spend — **Correct.** Staging buys information.
- ✗ (D) The CFO dislikes AI — **Wrong.** Not the reason.

K3 **Answer: D** — Review: Exercise 8

The furnace optimisation breaks even at about a 4% saving. What follows?

- ✗ (A) Reject it — **Wrong.** It may be valuable.
- ✗ (B) Add strategic value to justify it — **Wrong.** Inflation.
- ✗ (C) Fund the full rollout now — **Wrong.** Margin for error is thin.
- ✓ (D) Run a cheap one-furnace test; roll out only if it shows ≥ 6% — **Correct.** Evidence before commitment.

K4 **Answer: A** — Review: Exercise 7

What is the strongest reason the scheduling agent is "not now"?

- ✓ (A) The vendor's 8% depends on clean MES data and documented rules Brenner lacks (23% missing actuals; tacit rules), with no owner — **Correct.** Evidence and readiness.
- ✗ (B) It is high-risk under the EU AI Act — **Wrong.** It is not an Annex III use.
- ✗ (C) The CEO likes it — **Wrong.** Irrelevant to the evidence.
- ✗ (D) Agents are always too risky — **Wrong.** Overgeneralisation.

K5 **Answer: D** — Review: Exercise 11

Which item is a fragile dependency rather than an assumption?

- ✗ (A) Line 5 will behave like line 4 — **Wrong.** Assumption.
- ✗ (B) FTF will rise by 3 points — **Wrong.** Assumption.
- ✗ (C) Adoption will reach 75% — **Wrong.** Assumption.
- ✓ (D) A single integration engineer handles both MES and DMS integrations — **Correct.** True today but easily broken.

K6 **Answer: B** Review: Exercise 10

What should happen to the live HR CV-ranking feature?

✗ (A) Keep it; the vendor enabled it — **Wrong.** Deployer responsibility.

✓ (B) Switch it off, record the decision, and review with legal, DPO and works council — **Correct.** A Tier-4 stop.

✗ (C) Expand it — **Wrong.** No.

✗ (D) Score it in H2 — **Wrong.** It is live now.

K7 **Answer: A** Review: Exercise 10

Inspection review needs ~26 hours a month. What else must be tested before relying on it?

✓ (A) Effectiveness: a seeded-error test showing reviewers catch ≥ 90% of known defects — **Correct.** Capacity and effectiveness.

✗ (B) Vendor references — **Wrong.** Not oversight evidence.

✗ (C) Nothing — **Wrong.** Capacity is necessary but not sufficient.

✗ (D) Server speed — **Wrong.** Irrelevant.

K8 **Answer: C** Review: 2.7

What makes the Copilot redirect credible to the CIO?

✗ (A) Blaming the vendor — **Wrong.** Unhelpful.

✗ (B) Saying AI assistants are useless — **Wrong.** Untrue and alienating.

✓ (C) It solves the real problem (IP leakage, productivity for targeted groups) at ≈ €43k/yr instead of €306k/yr, with a trigger to expand on evidence — **Correct.** Yes to the problem; no to the unsupported scope.

✗ (D) Delaying the decision — **Wrong.** Avoidance.

MODULE 3

Integrating the Package and the Executive Presentation

Turning separate advisory components into one consistent package — one recommendation, one set of numbers, one story — and an executive presentation that leads with the decision, survives challenge and delivers the hardest message without losing the room.

YOU WILL BE ABLE TO

- Explain what integration means and why stapled packages fail
- Run a consistency check across all artifacts and fix contradictions
- Write a single governing recommendation
- Structure the package so each document has one job
- Build a 12-slide executive storyline with action titles
- Write an investment committee memo for the package
- Pre-wire and deliver a difficult finding; prepare for executive questions
- Red-team the package before it leaves your hands

IN THIS MODULE

1. What integration means
2. Exercise 12 · Consistency check
3. The single recommendation
4. Package architecture
5. Exercise 13 · Executive storyline
6. Exercise 14 · IC memo
7. Delivering the hardest message
8. Q&A preparation
9. Red-teaming the package
10. Templates · cheat sheet · knowledge check

Suggested study time · 6–8 hours (incl. exercises)

Eleven good artifacts do not make a good advisory package. Clients receive a package as **one argument**: if the roadmap contradicts the readiness finding, or the memo quotes a different number from the model, credibility collapses. This module teaches integration — consistency, a single recommendation, a clear document architecture — and the executive presentation that carries it, including how to deliver a finding the CEO will not want to hear.

KEY TAKEAWAYS

1. Integration = **one recommendation, one set of numbers, one story** across every document.
2. Check every **hand-off**: portfolio → readiness → roadmap → case → metrics → governance → memo.
3. Write a **single governing recommendation** that names the decision, the amount and what is declined.
4. Give each document **one job**; put detail in appendices.
5. Build slides with **action titles** that read as the argument on their own.
6. **Pre-wire** difficult findings one-to-one before the group meeting.
7. Prepare for the **predictable questions**; red-team the package before sending it.

3.1 What integration means

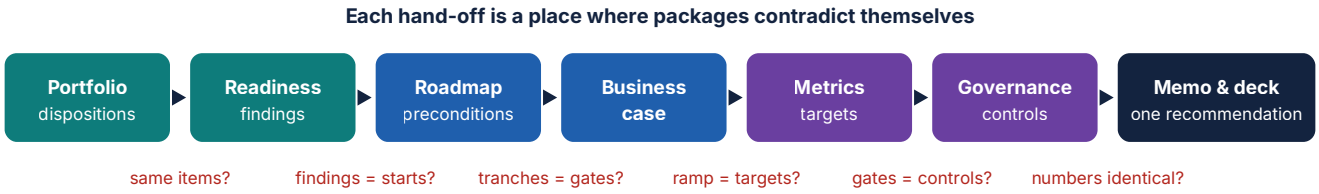


Figure 3.1 — The consistency chain

Stapled package	Integrated package
Each document written by a different person, in a different voice	One storyline; documents written to the same governing recommendation
Numbers differ slightly between memo, model and deck	One model is the single source; every number traces to it
Roadmap ignores readiness findings	Readiness findings become roadmap preconditions
Governance plan as an annex nobody reads	Governance controls appear as gate criteria and costs
Null recommendations hidden or missing	Declined ideas visible, with reasons and alternatives

3.2 Exercise 12 · Consistency check

EXERCISE 12 **Find the six contradictions** ⏱ 30 min

Draws on: All earlier exercises; C1 M3 (artifact test)

Inputs: The draft package excerpts below (prepared by a junior colleague)

YOUR TASK

- **ROADMAP:**
  "H1: inspection lines 4 and 5; service assistant DE and CZ; scheduling agent pilot in H2."
- **BUSINESS CASE:**
  "Wave-1 steady benefits €930k/yr (incl. furnace €192k)."
- **MEASUREMENT PLAN:**
  "Year-1 target: scrap and rework −18% on lines 4–5."
- **GOVERNANCE PLAN:**
  "Service assistant — Tier 1, no special controls."
- **EXECUTIVE SUMMARY:**
  "We request a €2.0m envelope."
- **NULL RECOMMENDATIONS:**
  "Scheduling agent: not now; decision-support test in H3."

For each excerpt, name the contradiction, the source of truth and the fix.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 12

Excerpt	Contradiction	Source of truth	Fix
Roadmap: L5 and CZ in H1	Readiness says L5 and CZ are not yet ready	Readiness findings (Ex. 5)	Move L5 and CZ to H2 with preconditions
Roadmap: scheduling pilot in H2	Null recommendation says H3 decision-support test	Null recommendation (Ex. 7)	H3, decision support (level 1), on precondition
Case: €930k incl. furnace	Furnace benefit counted before the test	ROI model (Ex. 8)	€738k; furnace shown as conditional upside
Metrics: Y1 −18%	Model ramps to steady state; L5 not live in year 1	ROI model ramp	Y1 −10% on L4; Y2 −18% on L4+L5
Governance: service Tier 1	Data boundary issue (customer contracts) and logs need works agreement	Governance plan (Ex. 10)	Tier 2 with permission fix, citation rule and works agreement
Summary: €2.0m	Plan totals ≈ €1.5m; €2m is the CFO's ceiling	Tranche plan (Ex. 6)	Request ≈ €1.5m envelope; T1 €0.52m now

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Detection	≤ 3 found	5 found	All 6, plus any you introduced yourself
Source of truth	Not named	Named	Named, with one model as single source for numbers
Fix	Edits one document	Edits both sides	Fixes and re-runs the chain check

3.3 The single recommendation

A package needs one **governing recommendation**: a single sentence that states the decision, the amount, the scope and what is declined. Everything else in the package supports, qualifies or implements it.

Weak governing statement	Why it fails
"AI offers significant opportunities for Brenner."	No decision
"We recommend a holistic AI transformation journey."	No amount, scope or test
"Approve ten AI initiatives."	Ignores readiness and evidence; nothing declined

KEY IDEA · Model governing recommendation

"Approve a staged, ≈ €1.5m, 18-month AI programme: release €0.52m now for quality inspection on line 4, a service-assistant pilot, a furnace-energy test and AI governance — and decline or redirect four ideas, including the autonomous scheduling agent, until the evidence supports them."

CONSULTANT'S TIP · Three tests for the sentence

- **Decision test:** could the board vote on it?
- **Money test:** does it say how much now and in total?
- **Courage test:** does it say what is declined?

3.4 Package architecture

Document	Job (one each)	Audience	Length
Executive deck	Win the decision	Advisory board, CEO, CFO	12 slides + appendix
Investment committee memo	Record the decision request and conditions	CFO, board	2 pages
Business case and ROI model	Let finance verify and challenge	CFO, controlling	8–12 pages + model
Roadmap and tranche plan	Sequence and fund the work	COO, PMO	3–4 pages
Governance and risk plan	Make AI use safe and defensible	CIO, DPO, works council, audit	6–8 pages
Measurement plan	Prove value	Owners, controlling	2 pages
Appendices	Evidence: inventory, screens, readiness, pressure test, null recommendations	Reviewers	As needed

3.5 Exercise 13 · The executive storyline

EXERCISE 13 **Twelve slides, one argument** ⏱ 60 min

Draws on: C5 M4 (60-second test, conservative first, null recommendations), C1 M3 (artifact test)

Inputs: Your integrated package

YOUR TASK

1. Write 12 slide titles as full-sentence assertions (action titles) that together tell the argument.
2. For each slide, note the one exhibit that proves the title.
3. Check: can a board member read only the titles and understand the recommendation, the money, the risks and what is declined?

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 13 — model storyline

1	**We recommend a staged €1.5m AI programme: €0.52m now, four ideas declined or redirected**	Governing recommendation; decision requested today; what the board gets in month 6.
2	**AI should address Brenner's margin, quality, service and IP-risk problems — not be an end in itself**	Problems with baselines: EBIT 6.1%, €3.0m quality cost, FTF 78%, ~70 staff pasting drawings.
3	**We assessed ten ideas from interviews, the shop floor and a systems audit**	Inventory by channel; how screening worked.
4	**Three are ready to start, three need preparation, four should not proceed as proposed**	Portfolio matrix with dispositions.
5	**Quality inspection carries the case: ~€530k a year, payback under two years**	Traceability to scrap CC and claims; pilot-adjusted 18%; owner.
6	**The programme returns €0.7m NPV in the base case — but turns negative if benefits fall 28%, so we fund in stages**	Scenario table, conservative first; breakeven 72%; tranche plan.
7	**Each step starts only when it is ready**	Roadmap with preconditions; Gate 1 criteria.
8	**The autonomous scheduling agent is not ready for Brenner yet — fix the data, then test decision support**	Claim read-back; MES 23% gap; route back and trigger.
9	**Three other ideas are better solved differently**	CV screening (reject; switched off), invoice AI → e-invoicing, Copilot-for-all → governed assistant.

10	**AI governance protects our drawings and our people from day one** Policy, data boundary, works agreements, HITL test results.
11	**We will know in six months whether it works — and stop if it doesn't** Measurement plan; Gate 1 criteria; stop rules.
12	**Decision today: approve Tranche 1 and the AI policy; next decision in month 6** Decision request box; owners; dates.

EXAM TRAP

Topic titles ("Business case", "Governance") fail the read-the-titles test. Action titles state the conclusion of the slide. Notice also that slide 6 puts the **downside first** and slide 8 gives the hardest finding its own slide — hiding it in an appendix would destroy trust when the CEO asks about it.

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Titles	Topic labels	Some assertions	All 12 assertions; titles alone tell the story
Order	Analysis first	Recommendation early	Recommendation first; downside before upside; hardest finding visible
Evidence	Text-heavy	One exhibit per slide	One exhibit per slide that proves the title

3.6 Exercise 14 · Investment committee memo

EXERCISE 14 **Two pages the board can approve** ⏱ 45 min

Draws on: C5 M4 (IC memo, evidence maturity, decision request)

Inputs: Your storyline and model

YOUR TASK

1. Write a two-page IC memo following the C5 template.
2. Tag the top five claims with evidence maturity (E1–E5).
3. Apply the 60-second test.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 14 — page one

Decision requested: Approve Tranche 1 (€0.52m) of an ≈ €1.5m, 18-month AI programme, and the Brenner AI Policy v1.0. Next decision at Gate 1 (month 6).

Recommendation: Start three ready initiatives — quality inspection on line 4, a service-assistant pilot for 30 technicians, a one-furnace energy test — plus AI governance and maintenance-data capture. Decline or redirect four ideas: the autonomous scheduling agent (not now), AI CV screening (reject; feature switched off), invoice AI (redirect to e-invoicing) and Copilot for all staff (redirect to a governed assistant for 120 users).

Return: wave-1 NPV €0.7m at 9% over 5 years (base); conservative −€0.18m; upside €1.2m; payback 2.4 years; breakeven at ~72% of planned benefits.

Top claims	Evidence maturity
Inspection reduces scrap/rework by 18% (planning)	E3 (own PoC, adjusted)
Service subcontract −25% after Gate 2	E3 (owner commitment; budget mechanism)
Repeat visits −270/yr	E2–E3 (benchmarks + pilot to confirm)
Furnace saving ≥ 6%	E2 (vendor references) → test
HITL inspection capacity 26 h/month	E3 (calculation from volumes)

Conditions: works agreements signed; contracts removed from assistant scope; baselines signed. **Stop rule:** if Gate 1 criteria are missed, no Tranche 2 for that initiative.

3.7 Delivering the hardest message

The scheduling-agent finding contradicts the CEO's publicly visible enthusiasm. Delivered for the first time in front of the advisory board, it would embarrass her and probably be rejected. The professional sequence:

Step	What you do
1 · Pre-wire one-to-one	Meet the CEO alone (week 10) before the steering meeting; show the evidence, not the conclusion first

Step	What you do
2 · Start from her goal	"You want Brenner to plan better than competitors. Here's what stands between us and that today."
3 · Show the route back	Data fix in H1–H2, decision-support test in H3, criteria for scaling autonomy
4 · Offer ownership of the alternative	The CEO announces the "AI-ready production data" initiative as her programme
5 · Agree the board wording	No surprises in the room; she can present the decision as hers

CONSULTANT'S TIP · Script

> *"I share your ambition for smarter scheduling. The demo's 8% came from a plant with complete data and written rules. A quarter of our operations have no actual times, and our planners carry rules no system knows. If we buy the agent now, I expect it to stall within months, and the board will remember it. If we fix the data this year and test decision support next year, you'll have evidence to scale — and it will be your programme that made it possible."*

PITFALL · Softening the finding

Under pressure, advisors drift to "the scheduling agent is promising, with some challenges". That sentence will be heard as "yes". Keep the disposition unambiguous — **not now** — and put the warmth into the route back.

3.8 Q&A preparation

Likely question	Asked by	Prepared answer (and slide / appendix)
"Why not just buy the scheduling agent and fix data as we go?"	CEO	Training and validating it requires the data; stalls cost more than waiting; route back (slide 8)
"Where exactly will the €0.5m a year show up?"	CFO	Scrap CC, claims, subcontract budget — owners named; budget cuts as gate conditions (slide 5, case §4)
"What if the inspection model fails on line 5?"	COO	L5 only after validation; L4 alone has positive NPV (case appendix)
"Will cameras monitor my colleagues?"	Works council	Operator area masked; works agreement before go-live; no performance use (governance §3)
"Why not give everyone Copilot? Competitors do."	CIO	No landing line for €2m claim; governed assistant now, expansion on evidence (slide 9)
"Is any of this high-risk under the AI Act?"	Board member	Wave 1 no; HR ranking would be — switched off (slide 10)
"What do we get in six months?"	Board	Gate 1 results against named criteria; decision on Tranche 2 (slide 11)

3.9 Red-teaming the package

Red-team question	Check
Does every number trace to the model?	Spot-check five numbers across deck, memo and case
Do readiness findings match roadmap starts?	Compare item by item
Is the downside visible before the upside?	Slide order and memo
Is every declined idea explained with an alternative?	Null recommendations present in deck and memo
Would the CFO find a hidden cost?	Hidden-cost checklist (C5 T2.3)
Would the works council be surprised by anything?	Governance and roadmap reviewed with them before the board
Can a reader find the hardest finding in 60 seconds?	Slide 8 and memo page one

3.10 Consulting templates

T7.1 Package Consistency Matrix

Purpose: Check every hand-off for contradictions.

Check	Document A	Document B	Consistent?	Source of truth	Fix
Portfolio items = roadmap items	*Portfolio*	*Roadmap*			
Readiness findings = start preconditions	*Readiness*	*Roadmap*			
Tranches = gates = case funding	*Roadmap*	*Business case*			
Ramp = Y1/Y2 targets	*ROI model*	*Measurement plan*			
Gate criteria = controls in place	*Roadmap*	*Governance plan*			
Numbers identical	*ROI model*	*Memo & deck*			
Declined ideas consistent	*Null recs*	*Roadmap & deck*			

Governing recommendation (one sentence): .....

T7.2 Executive Storyline Board

Purpose: Write the argument as action titles, one exhibit each.

#	Action title (full-sentence assertion)	Exhibit that proves it	Source document

- ☐ Titles alone tell the recommendation, money, risks and what is declined
- ☐ Downside shown before upside
- ☐ Hardest finding has its own slide

T7.3 Executive Q&A Preparation Sheet

Purpose: Prepare answers to predictable challenges.

Likely question	Who asks	Short answer (≤ 2 sentences)	Evidence / slide / appendix	Pre-wired?

3.11 Module 3 cheat sheet

MODULE 3 · ONE-PAGE SUMMARY

Integrating the Package and the Executive Presentation

INTEGRATION

One recommendation · one set of numbers (one model) · one story. Check every hand-off.

CONSISTENCY CHAIN

Portfolio → readiness → roadmap → case → metrics → governance → memo & deck.

GOVERNING RECOMMENDATION

Decision test · money test · courage test (what is declined).

ARCHITECTURE

Deck (win the decision) · memo (record it) · case (verify) · roadmap (sequence) · governance (protect) · metrics (prove) · appendices (evidence).

STORYLINE

12 action titles; recommendation first; downside before upside; hardest finding on its own slide; one exhibit per slide.

HARDEST MESSAGE

Pre-wire 1:1 · start from their goal · show route back · give them ownership of the alternative · agree board wording · never soften the disposition.

Q&A AND RED TEAM

Predictable questions by role with short answers and evidence. Red-team: numbers trace · findings = starts · downside visible · declined ideas explained · hidden costs · no surprises for works council · hardest finding findable in 60 seconds.

3.12 Knowledge check

Answer all questions before turning the page.

K1 **The memo says €2.0m; the tranche plan totals ≈ €1.5m. What should happen?**

- (A) Remove numbers from the memo
- (B) Keep €2.0m — it gives headroom
- (C) Use one model as the single source; request ≈ €1.5m and note the €2m as the CFO's ceiling
- (D) Average the two

K2 **Which governing recommendation passes the decision, money and courage tests?**

- (A) "Begin an AI journey."
- (B) "AI offers significant opportunities."
- (C) "Approve ten initiatives."
- (D) "Approve a staged ≈ €1.5m programme, €0.52m now for three ready initiatives and governance, and decline or redirect four ideas including the scheduling agent."

K3 **What is an action title?**

- (A) The company logo
- (B) A full-sentence assertion stating the slide's conclusion
- (C) A question to the audience
- (D) A one-word topic such as "Governance"

K4 **Why does the hardest finding get its own slide rather than an appendix?**

- (A) Hiding it would destroy trust when it surfaces; visible, it can be discussed with its route back
- (B) Board rules require it
- (C) Appendices are not read
- (D) To embarrass the CEO

K5 **What is the first step in delivering the scheduling-agent finding?**

- (A) Omit it
- (B) Pre-wire it with the CEO one-to-one, starting from her goal and showing the evidence and route back
- (C) Announce it in the board meeting
- (D) E-mail the board in advance

K6 **Under pressure, an advisor writes "the scheduling agent is promising, with some challenges". What is wrong?**

- (A) It softens an unambiguous "not now" into implied approval
- (B) Nothing — it is diplomatic
- (C) It should say "reject forever"
- (D) It is too long

K7 **Which red-team check would have caught the furnace benefit counted before the test?**

- (A) Vendor references
- (B) Spelling
- (C) Every number traces to the model, and the model's assumptions match the roadmap's preconditions
- (D) Slide colours

K8 **What is the single job of the IC memo in the package?**

(A) Win hearts
(B) Explain the technology
(C) Present all the evidence
(D) Record the decision request, return, conditions and stop rules for approval

3.13 Knowledge check: answers and explanations

K1 **Answer: C** — Review: Exercise 12

The memo says €2.0m; the tranche plan totals ≈ €1.5m. What should happen?

- ✗ (A) Remove numbers from the memo — **Wrong.** Executives need them.
- ✗ (B) Keep €2.0m — it gives headroom — **Wrong.** Inconsistent numbers destroy credibility.
- ✓ (C) Use one model as the single source; request ≈ €1.5m and note the €2m as the CFO's ceiling — **Correct.** One set of numbers.
- ✗ (D) Average the two — **Wrong.** Arbitrary.

K2 **Answer: D** — Review: 3.3

Which governing recommendation passes the decision, money and courage tests?

- ✗ (A) "Begin an AI journey." — **Wrong.** Vague.
- ✗ (B) "AI offers significant opportunities." — **Wrong.** No decision.
- ✗ (C) "Approve ten initiatives." — **Wrong.** Ignores evidence; nothing declined.
- ✓ (D) "Approve a staged ≈ €1.5m programme, €0.52m now for three ready initiatives and governance, and decline or redirect four ideas including the scheduling agent." — **Correct.** Decision, money and what is declined.

K3 **Answer: B** — Review: Exercise 13

What is an action title?

- ✗ (A) The company logo — **Wrong.** No.
- ✓ (B) A full-sentence assertion stating the slide's conclusion — **Correct.** Titles alone tell the argument.
- ✗ (C) A question to the audience — **Wrong.** Not an action title.
- ✗ (D) A one-word topic such as "Governance" — **Wrong.** Topic label.

K4 **Answer: A** — Review: Exercise 13

Why does the hardest finding get its own slide rather than an appendix?

- ✓ (A) Hiding it would destroy trust when it surfaces; visible, it can be discussed with its route back — **Correct.** Transparency builds credibility.
- ✗ (B) Board rules require it — **Wrong.** No.
- ✗ (C) Appendices are not read — **Wrong.** Partly true, but not the reason.
- ✗ (D) To embarrass the CEO — **Wrong.** The opposite of the intent.

K5 **Answer: B** — Review: 3.7

What is the first step in delivering the scheduling-agent finding?

- ✗ (A) Omit it — **Wrong.** Dishonest.
- ✓ (B) Pre-wire it with the CEO one-to-one, starting from her goal and showing the evidence and route back — **Correct.** No surprises; shared ownership.
- ✗ (C) Announce it in the board meeting — **Wrong.** That causes surprise and resistance.
- ✗ (D) E-mail the board in advance — **Wrong.** Bypasses the sponsor.

K6 **Answer: A** Review: 3.7

Under pressure, an advisor writes "the scheduling agent is promising, with some challenges". What is wrong?

✓ **A** It softens an unambiguous "not now" into implied approval — **Correct.** Keep the disposition clear; put the warmth in the route back.

✗ B Nothing — it is diplomatic — **Wrong.** It will be heard as "yes".

✗ C It should say "reject forever" — **Wrong.** That would also be wrong.

✗ D It is too long — **Wrong.** Irrelevant.

K7 **Answer: C** Review: 3.9

Which red-team check would have caught the furnace benefit counted before the test?

✗ A Vendor references — **Wrong.** Not the issue.

✗ B Spelling — **Wrong.** Irrelevant.

✓ **C** Every number traces to the model, and the model's assumptions match the roadmap's preconditions — **Correct.** Consistency chain.

✗ D Slide colours — **Wrong.** Irrelevant.

K8 **Answer: D** Review: 3.4

What is the single job of the IC memo in the package?

✗ A Win hearts — **Wrong.** That is closer to the deck.

✗ B Explain the technology — **Wrong.** Not its job.

✗ C Present all the evidence — **Wrong.** That is the appendices' job.

✓ **D** Record the decision request, return, conditions and stop rules for approval — **Correct.** One document, one job.

MODULE 4

Documenting Reasoning: The Portfolio Piece

Making the package stand alone: a reasoning log that shows how conclusions were reached, an evidence register, an honest account of the hardest finding, limitations, and a portfolio piece you can show clients and employers with confidence.

YOU WILL BE ABLE TO

- Explain what a portfolio piece must demonstrate to a reader who was not in the room
- Structure the portfolio piece so the package stands alone
- Keep a reasoning log and decision journal
- Maintain a graded evidence and assumption register
- Write the hardest finding as a professional narrative
- State limitations and what you would do differently
- Anonymise client work responsibly
- Assess the full package against an advisor-ready rubric and plan your development

IN THIS MODULE

1. What a portfolio piece must show
2. Structure of the portfolio piece
3. Exercise 15 · Reasoning log
4. Evidence and assumption register
5. Exercise 16 · The hardest finding
6. Limitations and what you would do differently
7. Anonymisation and confidentiality
8. Exercise 17 · Package self-assessment
9. Talking about the work
10. Your development plan
11. Templates · cheat sheet · knowledge check

Suggested study time · 5–6 hours (incl. exercises)

A strong package can still be a weak portfolio piece if the reader cannot see **how** you reached its conclusions. Clients and employers do not hire your recommendations; they hire your **judgment**. This module turns the Brenner package into a stand-alone portfolio piece: the reasoning behind each major decision, the evidence and its strength, the hardest finding told honestly, the limits of your work, and what you would do differently.

KEY TAKEAWAYS

1. A portfolio piece shows **judgment**: what you concluded, why, on what evidence, and what you rejected.
2. Keep a **reasoning log** during the engagement — reconstructing it afterwards produces hindsight, not reasoning.
3. Grade evidence and list assumptions so a reader can judge your confidence.
4. Tell the **hardest finding** fully: the finding, the evidence, why it was hard, how you handled it, what would change it.
5. State **limitations** and **what you would do differently** — maturity reads as strength.
6. **Anonymise** real client work and get permission; never expose confidential data.
7. Score the package against the **advisor-ready rubric** and plan your development.

4.1 What a portfolio piece must show

Figure 4.1 — Package + reasoning + learning = portfolio piece

A reader should be able to see…	Evidence in your piece
That you framed the right problem	Ask-to-problem reframing; engagement decision
That you found opportunities systematically	Inventory by channel, incl. shadow and embedded AI
That you judged evidence critically	Graded evidence; claim read-backs; pilot adjustments
That you could say no	Null recommendations with alternatives
That your numbers are honest	Model with scenarios, breakeven, hidden costs
That you made AI safe to use	Governance plan with tested controls
That you can lead executives to a decision	Governing recommendation; storyline; memo
That you know your limits	Limitations; scope of authority; specialist referrals

4.2 Structure of the portfolio piece

Section	Content	Length
1 · Overview	Client context (anonymised), your role, engagement question, governing recommendation	1 page
2 · Approach	Engagement plan, methods, frameworks used and why	1 page
3 · Key findings and recommendation	Portfolio, readiness, roadmap, case, governance — summary	2–3 pages
4 · Reasoning log	Major decisions, alternatives, evidence, reasoning	2 pages
5 · The hardest finding	Narrative (Exercise 16)	1 page
6 · Evidence and assumptions	Graded register; key sensitivities	1–2 pages
7 · Limitations and lessons	What the work could not establish; what you would do differently	1 page
Appendices	Executive deck, memo, selected templates	As needed

4.3 Exercise 15 · Reasoning log

EXERCISE 15 **Make your judgment visible** ⏱ 60 min

Draws on: All modules; C3 M4 (method log), C6 M1 (decision records)

Inputs: Your Brenner package

YOUR TASK

1. Identify the eight most consequential decisions you made in the engagement.
2. For each, record: options considered, evidence (graded), reasoning, what would have changed your mind, date.
3. Mark which decisions were contested and by whom.

✋ **Stop here and do the exercise before reading the model answer.**

MODEL ANSWER · Exercise 15 — excerpt

#	Decision	Options considered	Evidence (grade)	Reasoning	Would change if…
D1	Reframe from “AI plan” to four business problems	Accept ask; reframe	Interviews (A1); controlling data (V2)	Board needs a funding decision, not a technology list	CEO mandate were explicitly technology-scouting
D3	Treat HR CV-ranking feature as governance action, not opportunity	Score it; switch off; ignore	Admin audit (V2); works council (A1)	Live high-risk use; small value; objection	Never — unless full assessment and works agreement
D5	Plan inspection at 18%, not PoC’s 35%	35%, 25%, 18%	PoC on vendor set (E2–E3)	Visible adjustments for selection, L5, adoption, ledger	L4 production data in month 3 shows ≥ 25%

#	Decision	Options considered	Evidence (grade)	Reasoning	Would change if…
D6	Stage funding; T1 €0.52m	Full €1.5m; staged	Model: conservative −€183k; breakeven 72%	Downside too large for full commitment	Conservative positive after Gate 1
D7	Scheduling agent: not now (contested by CEO)	Fund; pilot; not now; reject	Demo (E2); MES gap (V1); planners (A1)	Data, ownership and adoption all weak	MES actuals ≥ 95% for 3 months
D8	Copilot-for-all → governed assistant (contested by CIO)	850 seats; 120 seats; none	Survey (V2); no landing line	Risk problem real; value claim untraceable	Targeted pilot shows measurable value

CONSULTANT'S TIP · Log as you go

Spend ten minutes at the end of each engagement week writing the decisions you made and why. A log written afterwards drifts towards a tidy story; a log written during the work shows real judgment — including changes of mind, which are a strength.

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Coverage	Few decisions	Major decisions	Major decisions incl. contested ones
Reasoning	Conclusions only	Reasons given	Options, graded evidence, reasoning and change conditions
Honesty	No changes of mind	Some	Records where and why your view changed

4.4 Evidence and assumption register

Item	Type	Grade / evidence	Sensitivity	Status
Scrap + rework €2.6m/yr	Fact	V2 controlling	High	Verified
Inspection planning reduction 18%	Assumption	E3 (adjusted PoC)	High (drives ~60% of benefit)	Test at Gate 1
Subcontract −25% after Gate 2	Commitment	A1 owner + budget mechanism	Medium	Gate condition
23% MES operations lack actuals	Fact	V1 profile	Decisive for D7	Verified
Furnace saving ≥ 6%	Assumption	E2 vendor references	High for furnace only	Test
Works agreements within 8 weeks	Dependency	A2	Medium (timing)	Started week 2

4.5 Exercise 16 · The hardest finding

EXERCISE 16 **Tell it fully and fairly** ⏱ 45 min

Draws on: C3 M4 (professional "not yet"), C5 M4 (null recommendation), C7 M3 (delivering the message)

Inputs: Exercise 7 and your reasoning log

YOUR TASK

1. Write a one-page narrative of the hardest finding using the six-part structure below.
2. Have someone who was not involved read it: can they explain the finding and why it was right?

✋ **Stop here and do the exercise before reading the model answer.**

Part	Question it answers
1 · The finding	What did you conclude, in one sentence?
2 · The evidence	What showed it, graded?
3 · Why it was hard	Whose expectations did it contradict, and what was at stake?
4 · How you handled it	Sequence, conversations, alternatives offered
5 · What would change it	Criteria and revisit trigger
6 · What you learned	What you would do differently next time

MODEL ANSWER · Exercise 16 — hardest finding narrative

1 · The finding. Brenner should not procure the autonomous production-scheduling agent the CEO saw at Hannover Messe; it should fix production data first and test decision support later.

2 · The evidence. The vendor's 8% output gain came from a demonstration on a reference plant with complete data (E2). At Brenner, 23% of MES operations lack actual times (V1, data profile of 410,000 operations); planners apply customer, tooling and staffing rules that exist in no system (A1, interviews with 6 of 14 planners, corroborated by the COO); no one had agreed to own the system; and released schedules are costly to reverse on the shop floor.

3 · Why it was hard. The CEO had spoken publicly about the agent; the advisory board had asked for an AI plan; the competitor's announcement created urgency. Recommending against it risked the sponsor's support for the whole programme.

4 · How I handled it. I pre-wired the finding with the CEO one-to-one in week 10, starting from her goal (better planning than competitors), showing the data before the conclusion, and offering a route back she could own: an "AI-ready production data" initiative now and a decision-support test in H3. We agreed the board wording together. The disposition stayed unambiguous: not now.

5 · What would change it. MES actual-time capture ≥ 95% for three months; planners co-design the test; a 10-week decision-support test shows ≥ 3% throughput gain on bottleneck cells with ≥ 60% planner acceptance.

6 · What I learned. I raised the data issue too late — in week 8. Next time I would profile the data behind any sponsor's favourite idea in week 2, so the conversation starts from shared facts rather than from a conclusion.

SELF-ASSESSMENT RUBRIC

Criterion	Developing (1)	Proficient (2)	Advisor-ready (3)
Clarity	Finding implied	Finding stated	Stated in one sentence with disposition
Fairness	One-sided	Acknowledges sponsor's view	Explains stakes and the sponsor's reasons respectfully
Learning	None	Generic	Specific change in your own practice

4.6 Limitations and what you would do differently

Limitation (Brenner model)	Why it matters
No production data yet for inspection; planning value relies on an adjusted PoC	The largest benefit is E3 — hence staging
Service time savings not measured by time study	Repeat-visit benefit depends on benchmarks until the pilot reports
Legal classification of inspection as non-safety component not yet confirmed by counsel	Referred; scope of authority respected
Brno plant interviews limited to two people	Czech adoption risk under-explored

KEY IDEA

Stating limitations does not weaken a portfolio piece — it shows you know exactly how strong your conclusions are. Readers trust advisors who mark the edges of their evidence.

4.7 Anonymisation and confidentiality

Rule	Practice
Get permission	Written client consent before using real work; respect NDAs
Anonymise identity	Rename the company; generalise industry and geography; remove people's names
Disguise numbers	Scale or index figures; keep ratios and logic intact
Remove confidential content	No drawings, contracts, personal data, internal system names
Check re-identification	Could an insider or competitor recognise the client? If yes, generalise further
Mark it	"Anonymised and modified for portfolio use"

NOTE

The Brenner case is fictional and can be used in your portfolio as a **capstone exercise**. Label it as such — never present a training case as client work.

4.8 Exercise 17 · Package self-assessment

EXERCISE 17 **Score the whole package** ⏱ 45 min

Draws on: All courses

Inputs: Your complete Brenner package and portfolio piece

YOUR TASK

1. Score each dimension of the rubric below (1–3).
2. For any score below 3, write the single change that would raise it.
3. Ask a peer to score it independently; discuss differences.

✋ **Stop here and do the exercise before reading the model answer.**

Dimension	Developing (1)	Proficient (2)	Advisor-ready (3)
Problem framing	Solution-led	Problems named	Problems with baselines, decision and decision rights
Opportunity discovery	Interviews only	Two channels	Three channels incl. shadow and embedded AI
Evidence discipline	Claims repeated	Some grading	Every material claim graded; read-backs; adjustments visible
Portfolio judgment	Top-N list	Dispositions	Dispositions for all, balance, consolidation
Readiness	Generic maturity	Dimensions rated	Weakest link, graded evidence, sequenced remediation
Roadmap	Dates	Phases	Preconditions, tranches, capacity, operating model
Business case	Headline ROI	NPV and scenarios	Traceable benefits, TCO, conservative first, breakeven
Measurement	Usage metrics	Outcomes	Baselines, paired metrics, attribution, Y1/Y2
Governance	Policy text	Classification	Obligations, data boundary, buildable controls, HITL tests
Courage	No declines	Some declines	Clear null recommendations incl. sponsor’s favourite, with routes back
Integration	Stapled documents	Mostly consistent	One recommendation, one model, consistency matrix passed
Communication	Topic slides	Clear deck	Action-title storyline; memo passes 60-second test; Q&A prepared
Reflection	None	Some	Reasoning log, hardest finding, limitations, lessons

CONSULTANT'S TIP · Reading your score

Advisor-ready on at least 10 of 13 dimensions, with no dimension at Developing, is a strong capstone. Courage and evidence discipline matter most: clients forgive a missing chart, not an unsupported recommendation.

4.9 Talking about the work

When presenting your portfolio piece to a client or employer, use a simple structure: **situation → question → approach → recommendation → hardest call → result or lesson**. Two minutes, then let them ask.

EXAMPLE · A two-minute version

> *"A family-owned drive-systems manufacturer asked for an AI plan after a trade fair. I reframed it into four business problems and assessed ten ideas from interviews, the shop floor and a systems audit. I recommended a staged €1.5m programme — inspection, a service assistant and a furnace test first — and declined four ideas. The hardest call was the CEO's favourite, an autonomous scheduling agent: the data wasn't there. I pre-wired it with her and gave her a route back she could own. The lesson: profile the data behind the sponsor's idea in week two, not week eight."*

4.10 Your development plan

Competency	Course	Self-rating 1–3	Next practice
Framing and decision rights	C1		
Capability mapping and evidence due diligence	C1		
Opportunity discovery and screening	C2		
Prioritization and portfolio design	C2		
Readiness assessment and reporting	C3		
Roadmaps, horizons and funding	C4		
Workflow and agent framing	C4		
Business cases, ROI and value metrics	C5		
Governance and risk management	C6		
Integration and executive communication	C7		

4.11 Consulting templates

T7.4 Reasoning Log

Purpose: Record consequential decisions as you make them. **Use when:** weekly during any engagement

#	Date	Decision	Options considered	Evidence (grade)	Reasoning	Contested by	Would change if...

T7.5 Hardest Finding Write-up

Purpose: Tell the most difficult finding fully and fairly.

1 · The finding (one sentence, with disposition):

2 · The evidence (graded):

3 · Why it was hard (whose expectations, what was at stake):

4 · How I handled it (sequence, conversations, alternatives):

5 · What would change it (criteria, revisit trigger):

6 · What I learned / would do differently:

T7.6 Portfolio Piece Outline & Anonymisation Check

Purpose: Assemble a stand-alone portfolio piece safely.

OUTLINE

- [ ] Overview: context, role, question, governing recommendation
- [ ] Approach: plan, methods, frameworks
- [ ] Key findings and recommendation
- [ ] Reasoning log (major and contested decisions)
- [ ] Hardest finding narrative
- [ ] Evidence and assumption register
- [ ] Limitations and lessons
- [ ] Appendices: deck, memo, selected templates

ANONYMISATION

- [ ] Written client permission / NDA checked
- [ ] Company, people and system names removed
- [ ] Numbers scaled or indexed; logic intact
- [ ] No confidential documents, drawings or personal data
- [ ] Re-identification test passed
- [ ] Labelled “anonymised and modified” or “capstone exercise”

4.12 Module 4 cheat sheet

MODULE 4 · ONE-PAGE SUMMARY

Documenting Reasoning: The Portfolio Piece

PURPOSE

Show judgment: what, why, on what evidence, what was rejected, how strong it is.

STRUCTURE

Overview · approach · findings & recommendation · reasoning log · hardest finding · evidence & assumptions · limitations & lessons · appendices.

REASONING LOG

Decision · options · graded evidence · reasoning · contested by · would change if. Write weekly, not afterwards.

HARDEST FINDING

Finding · evidence · why hard · how handled · what would change it · what I learned.

LIMITATIONS

Mark the edges of your evidence; refer what is outside your authority.

ANONYMISE

Permission · rename · scale numbers · remove confidential content · re-identification test · label.

SELF-ASSESSMENT & TALKING ABOUT IT

13-dimension rubric; advisor-ready ≥ 10 with none at Developing; courage and evidence discipline weigh most. Two-minute story: situation → question → approach → recommendation → hardest call → lesson.

4.13 Knowledge check

Answer all questions before turning the page.

K1 **What distinguishes a portfolio piece from the client package?**

- (A) It makes the reasoning visible: decisions, alternatives, graded evidence, the hardest finding and limitations
- (B) It is longer
- (C) It omits the recommendation
- (D) It has more charts

K2 **Why should the reasoning log be written during the engagement?**

- (A) It is faster
- (B) Logs reconstructed afterwards drift into a tidy story and lose real reasoning and changes of mind
- (C) To bill more hours
- (D) Clients require it

K3 **Which element of the hardest-finding narrative most shows professional maturity?**

- (A) Calling the CEO wrong
- (B) Omitting the evidence
- (C) Blaming the sponsor
- (D) A specific lesson about your own practice, such as profiling the data behind sponsor ideas earlier

K4 **Stating limitations in a portfolio piece…**

- (A) Is unnecessary
- (B) Should be in a hidden appendix
- (C) Shows you know the strength of your conclusions and where specialists are needed
- (D) Weakens your credibility

K5 **You want to use real client work in your portfolio. What comes first?**

- (A) Publish quickly
- (B) Rename the company
- (C) Obtain written client permission and check NDAs
- (D) Remove all numbers

K6 **How should the fictional Brenner case appear in your portfolio?**

- (A) Labelled as a capstone exercise
- (B) Only as slides
- (C) Without mention of its origin
- (D) As client work

K7 **Which rubric dimensions weigh most for clients?**

- (A) Number of frameworks named
- (B) Courage and evidence discipline
- (C) Slide design and colours
- (D) Length

K8 **Which two-minute structure fits presenting the work?**

- (A) Numbers only
- (B) Biography → company history
- (C) Frameworks → tools → slides
- (D) Situation → question → approach → recommendation → hardest call → lesson

4.14 Knowledge check: answers and explanations

K1 **Answer: A** Review: 4.1

What distinguishes a portfolio piece from the client package?

✓ **A** It makes the reasoning visible: decisions, alternatives, graded evidence, the hardest finding and limitations — **Correct.** Readers judge your judgment.

✗ **B** It is longer — **Wrong.** Length is not the point.

✗ **C** It omits the recommendation — **Wrong.** The recommendation stays.

✗ **D** It has more charts — **Wrong.** Irrelevant.

K2 **Answer: B** Review: Exercise 15

Why should the reasoning log be written during the engagement?

✗ **A** It is faster — **Wrong.** Not the reason.

✓ **B** Logs reconstructed afterwards drift into a tidy story and lose real reasoning and changes of mind — **Correct.** Contemporaneous records show genuine judgment.

✗ **C** To bill more hours — **Wrong.** No.

✗ **D** Clients require it — **Wrong.** Not necessarily.

K3 **Answer: D** Review: Exercise 16

Which element of the hardest-finding narrative most shows professional maturity?

✗ **A** Calling the CEO wrong — **Wrong.** Not the point.

✗ **B** Omitting the evidence — **Wrong.** Weakens it.

✗ **C** Blaming the sponsor — **Wrong.** Unprofessional.

✓ **D** A specific lesson about your own practice, such as profiling the data behind sponsor ideas earlier — **Correct.** Self-critical, specific learning.

K4 **Answer: C** Review: 4.6

Stating limitations in a portfolio piece…

✗ **A** Is unnecessary — **Wrong.** It is expected.

✗ **B** Should be in a hidden appendix — **Wrong.** Make it visible.

✓ **C** Shows you know the strength of your conclusions and where specialists are needed — **Correct.** Marks the edges of evidence.

✗ **D** Weakens your credibility — **Wrong.** The opposite.

K5 **Answer: C** Review: 4.7

You want to use real client work in your portfolio. What comes first?

✗ **A** Publish quickly — **Wrong.** No.

✗ **B** Rename the company — **Wrong.** Necessary but not first.

✓ **C** Obtain written client permission and check NDAs — **Correct.** Permission precedes anonymisation.

✗ **D** Remove all numbers — **Wrong.** Scale them; keep the logic.

K6 **Answer: A** Review: 4.7

How should the fictional Brenner case appear in your portfolio?

✓ A Labelled as a capstone exercise — **Correct.** Honest labelling.

✗ B Only as slides — **Wrong.** Format is not the issue.

✗ C Without mention of its origin — **Wrong.** Misleading.

✗ D As client work — **Wrong.** Misrepresentation.

K7 **Answer: B** Review: Exercise 17

Which rubric dimensions weigh most for clients?

✗ A Number of frameworks named — **Wrong.** No.

✓ B Courage and evidence discipline — **Correct.** Clients do not forgive unsupported recommendations.

✗ C Slide design and colours — **Wrong.** Clients forgive missing polish.

✗ D Length — **Wrong.** No.

K8 **Answer: D** Review: 4.9

Which two-minute structure fits presenting the work?

✗ A Numbers only — **Wrong.** Lacks story.

✗ B Biography → company history — **Wrong.** Irrelevant.

✗ C Frameworks → tools → slides — **Wrong.** Method-centric.

✓ D Situation → question → approach → recommendation → hardest call → lesson — **Correct.** Judgment-centric story.

R1 The Advisor's Playbook: the master cheat sheet

THE WHOLE PROGRAM ON ONE PAGE

The AI Transformation Advisor's Playbook

1 · FRAME (C1)

Ask → business problem → decision → decision-holder. Baselines before ideas. Engagement plan with gates.

2 · DISCOVER (C2)

Three channels: interviews, shop floor/process, systems audit. Include shadow and embedded AI. One inventory with IDs.

3 · SCREEN & PRIORITIZE (C2)

Fit, value, feasibility, risk. Grade evidence. Dispositions for every idea: do now, prepare, test, not now, reject, redirect.

4 · READINESS (C3)

Dimensions with graded evidence; find the weakest link per use case; sequence remediation; say "not yet" professionally.

5 · ROADMAP (C4)

Preconditions → tranches → gates. Capacity and WIP limits. Operating model. Horizons H1–H3.

6 · BUSINESS CASE (C5)

Traceable benefits, full TCO, pilot adjustment, conservative case first, breakeven, staged funding, Y1/Y2 measurement.

7 · GOVERN (C6)

Classify; obligations; data boundary; failure statements → buildable controls; HITL capacity and seeded tests.

8 · INTEGRATE (C7 M3)

One governing recommendation, one model, one set of numbers. Consistency matrix. Action-title storyline; one-page memo.

9 · DOCUMENT (C7 M4) — AND THE RULES THAT NEVER CHANGE

Reasoning log, evidence register, hardest finding, limitations, lessons. **Always:** evidence before enthusiasm · every idea gets a disposition · conservative case first · recommend against what the evidence does not support, with a route back · pre-wire hard messages · stay inside your authority and refer legal questions · never split the story across inconsistent documents.

R2 Framework and template index

Where each framework from the program comes from, and where you applied it in this capstone.

Framework / tool	Course	Applied in
Ask-to-problem reframing; engagement plan	C1	Ex 1
Capability map; vendor claim due diligence	C1	Ex 2, 7, 11
Three-channel discovery; opportunity inventory	C2	Ex 2
Screening and scoring; evidence grades	C2	Ex 3
Portfolio dispositions and balance	C2	Ex 4, §2.7
Readiness dimensions; weakest link; "not yet"	C3	Ex 5, 7
Roadmap: preconditions, tranches, gates, capacity	C4	Ex 6
Operating model (hub-and-spoke)	C4	Ex 6
Agent / workflow framing	C4	Ex 7
NPV, payback, IRR, scenarios, breakeven	C5	Ex 8
Value measurement plan; attribution	C5	Ex 9
Classification, data boundary, buildable controls, HITL tests	C6	Ex 10
Executive pressure test; Q&A	C6	Ex 11, §3
Consistency matrix (T7.1)	C7	Ex 12
Storyline board (T7.2) and Q&A sheet (T7.3)	C7	Ex 13, 14
Reasoning log (T7.4); hardest finding (T7.5); portfolio outline (T7.6)	C7	Ex 15–17

Course 7 templates

Code	Template	Use when
T7.1	Package Consistency Matrix	Before any package leaves your hands
T7.2	Executive Storyline Board	Building the executive deck
T7.3	Executive Q&A Preparation Sheet	Preparing for board or steering meetings
T7.4	Reasoning Log	Weekly during every engagement
T7.5	Hardest Finding Write-up	After the engagement; for the portfolio
T7.6	Portfolio Piece Outline & Anonymisation Check	Assembling a portfolio piece

R3 Glossary of capstone terms

A

Advisory package
The integrated set of deliverables (portfolio, readiness, roadmap, case, governance, deck, memo) built around one recommendation.

Anonymisation
Removing or disguising identifying information so client work can be shown without breaching confidentiality.

B

Breakeven
Value of a key assumption at which NPV equals zero.

C

Conservative case
Scenario using cautious assumptions; shown first to build trust.

Consistency matrix
Cross-check that every figure, name, date and disposition is identical across the package's documents.

D

Disposition
The decision for each idea: do now, prepare, test, not now, reject or redirect.

E

Evidence grade
Label for how strong a piece of evidence is (e.g., verified data, attested, external claim, estimate).

G

Gate
A decision point where named criteria decide whether the next tranche of funding is released.

Governing recommendation
The single sentence the whole package supports; what you want the decision-holder to approve.

H

Hardest finding
The conclusion that most contradicted a stakeholder's expectations; documented fully in the portfolio piece.

N

Null recommendation
A recommendation not to do something, with reasons, alternatives and a revisit trigger.

P

Pilot adjustment
Reduction of pilot or PoC results for selection effects, scope, adoption and other differences from production.

Portfolio piece
A stand-alone document showing your package and your reasoning, for clients or employers.

Pre-wiring
Discussing a difficult finding with a key stakeholder one-to-one before the formal meeting.

R

Reasoning log
Contemporaneous record of major decisions, options, evidence, reasoning and change conditions.

Red team
Deliberate attempt to find weaknesses in the package before the client does.

Route back
The conditions under which a declined idea would be reconsidered, offered to its sponsor.

S

Storyline
Sequence of action titles that, read alone, makes the case for the governing recommendation.

T

Tranche
A portion of the funding envelope released at a gate.

W

WIP limit
Maximum number of use cases in active build, set by scarce capacity.

R4 Integrated mock exam

Format: 36 scenario-based multiple-choice questions · one best answer each · **Time:** 60 minutes · **Suggested pass mark:** 70% (25 correct)

All questions use the Brenner case unless stated. Questions 1–10 cover Module 1, 11–20 Module 2, 21–28 Module 3, 29–33 Module 4 and 34–36 are integrated scenarios across the whole program. Answer every question before checking the key, which explains why each option is right or wrong.

Q1 **A CEO asks for "an AI plan" after a trade fair. What is the advisor's first move?**

- (A) Start a vendor selection
- (B) Reframe the ask into business problems, the decision needed and who makes it
- (C) Produce a long list of AI technologies
- (D) Run a company-wide AI survey only

Q2 **Why use three discovery channels (interviews, shop floor, systems audit)?**

- (A) To bill more days
- (B) Interviews are unreliable and should be avoided
- (C) Clients require it by law
- (D) Each channel finds ideas and evidence the others miss, including shadow and embedded AI

Q3 **The systems audit finds a live CV-ranking feature in the HR suite. How should it be treated?**

- (A) As a governance action — assess and switch off pending review
- (B) As a quick-win opportunity
- (C) Expand it to all hiring
- (D) Ignore it — it came with the software

Q4 **310 of 640 office staff use public chatbots, about 70 paste drawings. What does this most directly show?**

- (A) AI demand is fake
- (B) Ban all AI immediately and stop
- (C) A real need and an urgent data-protection risk to address with a governed alternative and a data boundary
- (D) Buy Copilot for all 850 users

Q5 **A vendor PoC claims 35% scrap reduction on a vendor-selected image set. What grade is it, and what do you do?**

- (A) Verified data; use as is
- (B) Double it for the upside case
- (C) External claim / test evidence; plan with a visible adjustment and test on production data
- (D) Irrelevant; ignore

Q6 **Which best describes a good portfolio of dispositions?**

- (A) A top-5 list
- (B) Ideas ranked by vendor enthusiasm
- (C) Only ideas you will fund
- (D) Every idea has a disposition — do now, prepare, test, not now, reject or redirect — with balance and consolidation

Q7 **The readiness summary rates data "amber" overall, but MES actuals are missing for 23% of operations. Why does that matter more for scheduling than for inspection?**

- A Averages are always enough
- B Scheduling depends on actual times; inspection depends on images — the weakest link differs per use case
- C It doesn't
- D Inspection uses MES actuals

Q8 **What makes a readiness finding "professional" when the answer is not yet?**

- A Clear disposition, graded evidence, the remediation sequence and a trigger to revisit
- B Recommending a different vendor
- C Omitting the finding
- D Vague language to avoid offence

Q9 **Why should every material claim be read back to its owner?**

- A To confirm the number, its source and the owner's commitment before it enters the case
- B To shorten the engagement
- C Politeness
- D Legal requirement

Q10 **The engagement plan has 12 weeks. When should the data behind the sponsor's favourite idea be profiled?**

- A Only if the vendor agrees
- B Early — around week 2 — so the conversation starts from shared facts
- C Week 11, before the board
- D Never — it is the sponsor's idea

Q11 **Why does the roadmap release only €0.52m of a ≈ €1.5m envelope now?**

- A To make the project look small
- B Vendors require it
- C Staging limits downside: the conservative case is negative, so later tranches depend on gate evidence
- D Budget is short

Q12 **Brenner has one integration engineer. What follows for the roadmap?**

- A Ignore it — hire later
- B Outsource everything without plan
- C Run all builds in parallel
- D Set a WIP limit and sequence builds accordingly

Q13 **The disposition for the autonomous scheduling agent is "not now". Which pairing is right?**

- A Not now, with no conditions
- B Pilot immediately to please the CEO
- C Not now, plus an AI-ready data initiative, a later decision-support test and a trigger (MES actuals ≥ 95%)
- D Reject forever

Q14 **Wave 1 NPV is +€721k but the conservative case is −€183k. What is the right recommendation?**

- A Fund a first tranche with gate criteria; show the conservative case first
- B Fund everything now
- C Hide the conservative case
- D Cancel the programme

Q15 **Breakeven for Wave 1 is about 72% of planned benefit. What does that tell the board?**

- A Payback is 72 months
- B Benefits can fall about 28% short before NPV turns negative — a useful margin for the decision
- C IRR is 72%
- D The case fails

Q16 **The furnace-energy idea has NPV ≈ €4k at 4% saving and breakeven at 3.96%. Best disposition?**

- A Ask the vendor to guarantee 10%
- B Reject because NPV is small
- C Fund full roll-out
- D A low-cost test to establish whether savings reach ≈ 6%, before any roll-out

Q17 **Why is the invoice-AI idea redirected to e-invoicing?**

- A AI is banned in finance
- B The CFO dislikes AI
- C E-invoicing uses more AI
- D A simpler, non-AI fix solves most of the problem at lower risk and cost

Q18 **Copilot for all 850 users (€306k/yr) vs a governed assistant for 120 seats (≈ €43k/yr). Why redirect?**

- A Copilot is unsafe
- B Cheaper is always better
- C The risk problem is real but the value claim is untraceable; a targeted, governed pilot addresses both
- D The CIO asked for it

Q19 **A measurement plan pairs an outcome metric with a guardrail. Which pair fits inspection?**

- A Model accuracy and number of logins
- B Scrap + rework cost on line 4 (outcome) and escaped defects / customer claims (guardrail)
- C Vendor satisfaction
- D Number of images labelled

Q20 **The inspection HITL needs about 26 review hours per month. Why test it with 40 seeded images?**

- A To prove reviewers actually catch errors — controls must be tested, not assumed
- B Regulation demands exactly 40
- C To slow production
- D To train the model

Q21 **The deck says T1 = €0.52m, the memo says €0.55m. What is the right action?**

- (A) Trace both to the single model, correct the wrong one and re-run the consistency matrix
- (B) Average them
- (C) Remove the number
- (D) Leave it — close enough

Q22 **What makes a governing recommendation good?**

- (A) It lists every finding
- (B) One sentence stating the action, scale, scope, timing and what is declined
- (C) It is a question
- (D) It avoids numbers

Q23 **An executive storyline is good when…**

- (A) Each slide has a topic title
- (B) Slides follow the order of the work
- (C) Reading the action titles alone makes the case for the recommendation
- (D) It has 40 slides

Q24 **Where should the hardest message appear in the board deck?**

- (A) Only in the memo
- (B) Hidden in the appendix
- (C) Not at all
- (D) Clearly in the main storyline, after it has been pre-wired with the sponsor

Q25 **The IC memo's first page should pass the "60-second test". What does that mean?**

- (A) It has 60 words
- (B) A busy reader grasps the recommendation, the money, the risks and the ask within a minute
- (C) It was written in 60 seconds
- (D) It can be read aloud in 60 seconds verbatim

Q26 **Why prepare a Q&A sheet before the board meeting?**

- (A) To shorten the meeting
- (B) Because boards forbid improvisation
- (C) To anticipate hard questions and ground answers in the package's evidence and numbers
- (D) To memorise scripted answers

Q27 **What is the purpose of a red-team review of the package?**

- (A) To add more slides
- (B) To rewrite it from scratch
- (C) To approve it automatically
- (D) To find contradictions, weak evidence and missing alternatives before the client does

Q28 **A board member asks: "Why not just buy the scheduling agent — our competitor did?" Best answer?**

- (A) Start from the shared goal, show the data gap, then the route back and trigger
- (B) "The CEO is wrong."
- (C) "We will buy it later anyway."
- (D) "AI is hype."

Q29 **What turns a client package into a portfolio piece?**

- (A) Removing the recommendation
- (B) Visible reasoning: decisions, alternatives, graded evidence, hardest finding and limitations
- (C) Adding the vendor's marketing
- (D) Nicer formatting

Q30 **Which entry belongs in a reasoning log?**

- (A) Decision, options considered, graded evidence, reasoning, who contested it, what would change it
- (B) "Recommend inspection."
- (C) Vendor price lists
- (D) Meeting minutes verbatim

Q31 **Which is the best "what I learned" line in a hardest-finding write-up?**

- (A) "AI is hard."
- (B) "Everything went well."
- (C) "The CEO should have listened."
- (D) "I would profile the data behind any sponsor's favourite idea in week 2."

Q32 **Before using real client work in your portfolio you must first…**

- (A) Publish and wait for objections
- (B) Rename the client
- (C) Get written permission and check NDAs
- (D) Delete all numbers

Q33 **Your self-assessment shows Advisor-ready on 11 of 13 dimensions but "Developing" on Courage. Is the package strong?**

- (A) Not yet — courage weighs most; add clear null recommendations with routes back
- (B) Yes — 11 of 13 is enough
- (C) Only design matters
- (D) Yes, courage is optional

Q34 **Integrated: The CFO wants the full €1.5m approved now "to save time". Your model's conservative case is negative. What do you recommend?**

- (A) Agree — speed matters most
- (B) Cut the programme to inspection only
- (C) Keep staging, but pre-agree gate criteria and dates so later tranches release fast when evidence arrives
- (D) Refuse to discuss

Q35 **Integrated: Counsel has not yet confirmed that inspection is not a safety component. What do you write?**

A "Inspection is not high-risk."

B Omit the topic

C Stop the programme

D State the working classification, mark it as a limitation, refer to counsel and make confirmation a Gate 1 condition

Q36 **Integrated: After Gate 1, line 4 production data shows 12% scrap reduction (plan 18%). What next?**

A Cancel immediately

B Re-run the model, compare to breakeven, review causes, and decide at the gate with the owner

C Blame the vendor publicly

D Hide it and proceed

R5 Mock exam: answer key and explanations

Quick answer grid

1 · B	**2** · D	**3** · A	**4** · C	**5** · C	**6** · D	**7** · B	**8** · A	**9** · A	**10** · B
11 · C	**12** · D	**13** · C	**14** · A	**15** · B	**16** · D	**17** · D	**18** · C	**19** · B	**20** · A
21 · A	**22** · B	**23** · C	**24** · D	**25** · B	**26** · C	**27** · D	**28** · A	**29** · B	**30** · A
31 · D	**32** · C	**33** · A	**34** · C	**35** · D	**36** · B				

Score	Interpretation	Next step
31–36	Advisor-ready	Do the Kessler practice case; build your portfolio piece
25–30	Pass level	Redo the exercises for modules where you missed two or more
below 25	Not yet	Revisit the source course cheat sheets via the index (R2); retake in one week

Full explanations

Q1 **Answer: B** — Review: C7 Ex 1

A CEO asks for "an AI plan" after a trade fair. What is the advisor's first move?

- ✗ (A) Start a vendor selection — **Wrong.** Premature.
- ✓ (B) Reframe the ask into business problems, the decision needed and who makes it — **Correct.** Framing comes first.
- ✗ (C) Produce a long list of AI technologies — **Wrong.** Solution-led; skips the problem.
- ✗ (D) Run a company-wide AI survey only — **Wrong.** One channel; no framing.

Q2 **Answer: D** — Review: C7 Ex 2

Why use three discovery channels (interviews, shop floor, systems audit)?

- ✗ (A) To bill more days — **Wrong.** Not the reason.
- ✗ (B) Interviews are unreliable and should be avoided — **Wrong.** Interviews are one channel, not useless.
- ✗ (C) Clients require it by law — **Wrong.** No such rule.
- ✓ (D) Each channel finds ideas and evidence the others miss, including shadow and embedded AI — **Correct.** Triangulation.

Q3 **Answer: A** — Review: C7 Ex 2, D3

The systems audit finds a live CV-ranking feature in the HR suite. How should it be treated?

- ✓ (A) As a governance action — assess and switch off pending review — **Correct.** Live high-risk use needs governance, not scoring.
- ✗ (B) As a quick-win opportunity — **Wrong.** It is already live and high-risk.
- ✗ (C) Expand it to all hiring — **Wrong.** Raises risk further.
- ✗ (D) Ignore it — it came with the software — **Wrong.** Embedded AI is still the deployer's responsibility.

Q4 **Answer: C** Review: C7 M1

310 of 640 office staff use public chatbots, about 70 paste drawings. What does this most directly show?

✗ (A) AI demand is fake — **Wrong.** Demand is real.

✗ (B) Ban all AI immediately and stop — **Wrong.** Bans push use underground.

✓ (C) A real need and an urgent data-protection risk to address with a governed alternative and a data boundary — **Correct.** Both value and risk.

✗ (D) Buy Copilot for all 850 users — **Wrong.** Untraceable value claim.

Q5 **Answer: C** Review: C7 Ex 3, D5

A vendor PoC claims 35% scrap reduction on a vendor-selected image set. What grade is it, and what do you do?

✗ (A) Verified data; use as is — **Wrong.** Vendor-selected sets are not verified production data.

✗ (B) Double it for the upside case — **Wrong.** No basis.

✓ (C) External claim / test evidence; plan with a visible adjustment and test on production data — **Correct.** Adjust and verify.

✗ (D) Irrelevant; ignore — **Wrong.** It is informative.

Q6 **Answer: D** Review: C7 Ex 4

Which best describes a good portfolio of dispositions?

✗ (A) A top-5 list — **Wrong.** Leaves other ideas undecided.

✗ (B) Ideas ranked by vendor enthusiasm — **Wrong.** No.

✗ (C) Only ideas you will fund — **Wrong.** Declines are part of the portfolio.

✓ (D) Every idea has a disposition — do now, prepare, test, not now, reject or redirect — with balance and consolidation — **Correct.** Complete and decisive.

Q7 **Answer: B** Review: C7 Ex 5

The readiness summary rates data "amber" overall, but MES actuals are missing for 23% of operations. Why does that matter more for scheduling than for inspection?

✗ (A) Averages are always enough — **Wrong.** They hide the weakest link.

✓ (B) Scheduling depends on actual times; inspection depends on images — the weakest link differs per use case — **Correct.** Weakest-link logic.

✗ (C) It doesn't — **Wrong.** Readiness is use-case-specific.

✗ (D) Inspection uses MES actuals — **Wrong.** It uses labelled images.

Q8 **Answer: A** Review: C7 Ex 5, C3 M4

What makes a readiness finding "professional" when the answer is not yet?

✓ (A) Clear disposition, graded evidence, the remediation sequence and a trigger to revisit — **Correct.** Specific and constructive.

✗ (B) Recommending a different vendor — **Wrong.** Doesn't address readiness.

✗ (C) Omitting the finding — **Wrong.** Dishonest.

✗ (D) Vague language to avoid offence — **Wrong.** Unclear.

Q9 **Answer: A** Review: C7 M1

Why should every material claim be read back to its owner?

✓ **A** To confirm the number, its source and the owner's commitment before it enters the case — **Correct.** Read-backs prevent untraceable figures.

✗ B To shorten the engagement — **Wrong.** Not the purpose.

✗ C Politeness — **Wrong.** More than that.

✗ D Legal requirement — **Wrong.** Not generally.

Q10 **Answer: B** Review: C7 Ex 16

The engagement plan has 12 weeks. When should the data behind the sponsor's favourite idea be profiled?

✗ A Only if the vendor agrees — **Wrong.** Irrelevant.

✓ **B** Early — around week 2 — so the conversation starts from shared facts — **Correct.** Lesson from the hardest finding.

✗ C Week 11, before the board — **Wrong.** Too late.

✗ D Never — it is the sponsor's idea — **Wrong.** Avoidance.

Q11 **Answer: C** Review: C7 Ex 6, D6

Why does the roadmap release only €0.52m of a ≈ €1.5m envelope now?

✗ A To make the project look small — **Wrong.** Not the reason.

✗ B Vendors require it — **Wrong.** No.

✓ **C** Staging limits downside: the conservative case is negative, so later tranches depend on gate evidence — **Correct.** Staged funding logic.

✗ D Budget is short — **Wrong.** The client has up to €2m.

Q12 **Answer: D** Review: C7 Ex 6

Brenner has one integration engineer. What follows for the roadmap?

✗ A Ignore it — hire later — **Wrong.** Capacity drives sequencing now.

✗ B Outsource everything without plan — **Wrong.** Uncontrolled.

✗ C Run all builds in parallel — **Wrong.** Overloads the bottleneck.

✓ **D** Set a WIP limit and sequence builds accordingly — **Correct.** Capacity constraint.

Q13 **Answer: C** Review: C7 Ex 7

The disposition for the autonomous scheduling agent is "not now". Which pairing is right?

✗ A Not now, with no conditions — **Wrong.** Leaves no route back.

✗ B Pilot immediately to please the CEO — **Wrong.** Ignores evidence.

✓ **C** Not now, plus an AI-ready data initiative, a later decision-support test and a trigger (MES actuals ≥ 95%) — **Correct.** Decline with a route back.

✗ D Reject forever — **Wrong.** Too absolute; evidence could change.

Q14 **Answer: A** Review: C7 Ex 8

Wave 1 NPV is +€721k but the conservative case is −€183k. What is the right recommendation?

✓ A Fund a first tranche with gate criteria; show the conservative case first — **Correct.** Staging plus honest presentation.

✗ B Fund everything now — **Wrong.** Ignores downside.

✗ C Hide the conservative case — **Wrong.** Destroys trust.

✗ D Cancel the programme — **Wrong.** Base and upside are attractive; staging manages risk.

Q15 **Answer: B** Review: C7 Ex 8

Breakeven for Wave 1 is about 72% of planned benefit. What does that tell the board?

✗ A Payback is 72 months — **Wrong.** No.

✓ B Benefits can fall about 28% short before NPV turns negative — a useful margin for the decision — **Correct.** Breakeven explains headroom.

✗ C IRR is 72% — **Wrong.** Different measure.

✗ D The case fails — **Wrong.** It doesn't.

Q16 **Answer: D** Review: C7 Ex 8

The furnace-energy idea has NPV ≈ €4k at 4% saving and breakeven at 3.96%. Best disposition?

✗ A Ask the vendor to guarantee 10% — **Wrong.** Not evidence.

✗ B Reject because NPV is small — **Wrong.** Upside at 6% is meaningful.

✗ C Fund full roll-out — **Wrong.** Value too thin on current evidence.

✓ D A low-cost test to establish whether savings reach ≈ 6%, before any roll-out — **Correct.** Test cheaply; value is assumption-driven.

Q17 **Answer: D** Review: C7 §2.7

Why is the invoice-AI idea redirected to e-invoicing?

✗ A AI is banned in finance — **Wrong.** No.

✗ B The CFO dislikes AI — **Wrong.** Not the reason.

✗ C E-invoicing uses more AI — **Wrong.** It is a process/standard fix.

✓ D A simpler, non-AI fix solves most of the problem at lower risk and cost — **Correct.** Redirect when a better non-AI route exists.

Q18 **Answer: C** Review: C7 §2.7, D8

Copilot for all 850 users (€306k/yr) vs a governed assistant for 120 seats (≈ €43k/yr). Why redirect?

✗ A Copilot is unsafe — **Wrong.** Not the reasoning.

✗ B Cheaper is always better — **Wrong.** Not the principle.

✓ C The risk problem is real but the value claim is untraceable; a targeted, governed pilot addresses both — **Correct.** Redirect to evidence.

✗ D The CIO asked for it — **Wrong.** Not a reason.

Q19 **Answer: B** Review: C7 Ex 9

A measurement plan pairs an outcome metric with a guardrail. Which pair fits inspection?

✗ (A) Model accuracy and number of logins — **Wrong.** Usage and model metrics only.

✓ (B) Scrap + rework cost on line 4 (outcome) and escaped defects / customer claims (guardrail) — **Correct.** Outcome plus guardrail.

✗ (C) Vendor satisfaction — **Wrong.** Irrelevant.

✗ (D) Number of images labelled — **Wrong.** Activity metric.

Q20 **Answer: A** Review: C7 Ex 10

The inspection HITL needs about 26 review hours per month. Why test it with 40 seeded images?

✓ (A) To prove reviewers actually catch errors — controls must be tested, not assumed — **Correct.** Seeded tests check automation bias.

✗ (B) Regulation demands exactly 40 — **Wrong.** No such number.

✗ (C) To slow production — **Wrong.** No.

✗ (D) To train the model — **Wrong.** Not the purpose.

Q21 **Answer: A** Review: C7 Ex 12

The deck says T1 = €0.52m, the memo says €0.55m. What is the right action?

✓ (A) Trace both to the single model, correct the wrong one and re-run the consistency matrix — **Correct.** One source of truth.

✗ (B) Average them — **Wrong.** Creates a third number.

✗ (C) Remove the number — **Wrong.** Hides a key fact.

✗ (D) Leave it — close enough — **Wrong.** Undermines trust.

Q22 **Answer: B** Review: C7 M3

What makes a governing recommendation good?

✗ (A) It lists every finding — **Wrong.** Too broad.

✓ (B) One sentence stating the action, scale, scope, timing and what is declined — **Correct.** Decision-ready.

✗ (C) It is a question — **Wrong.** Recommendation must recommend.

✗ (D) It avoids numbers — **Wrong.** Numbers anchor the decision.

Q23 **Answer: C** Review: C7 Ex 13

An executive storyline is good when…

✗ (A) Each slide has a topic title — **Wrong.** Topics don't argue.

✗ (B) Slides follow the order of the work — **Wrong.** Order should follow the argument.

✓ (C) Reading the action titles alone makes the case for the recommendation — **Correct.** Horizontal logic test.

✗ (D) It has 40 slides — **Wrong.** Length is not quality.

Q24 **Answer: D** Review: C7 M3

Where should the hardest message appear in the board deck?

✗ (A) Only in the memo — **Wrong.** Inconsistent.

✗ (B) Hidden in the appendix — **Wrong.** Avoidance.

✗ (C) Not at all — **Wrong.** Dishonest.

✓ (D) Clearly in the main storyline, after it has been pre-wired with the sponsor — **Correct.** Visible and prepared.

Q25 **Answer: B** Review: C7 Ex 14

The IC memo's first page should pass the "60-second test". What does that mean?

✗ (A) It has 60 words — **Wrong.** No.

✓ (B) A busy reader grasps the recommendation, the money, the risks and the ask within a minute — **Correct.** Decision clarity.

✗ (C) It was written in 60 seconds — **Wrong.** No.

✗ (D) It can be read aloud in 60 seconds verbatim — **Wrong.** Not the meaning.

Q26 **Answer: C** Review: C7 T7.3

Why prepare a Q&A sheet before the board meeting?

✗ (A) To shorten the meeting — **Wrong.** Side effect at most.

✗ (B) Because boards forbid improvisation — **Wrong.** No.

✓ (C) To anticipate hard questions and ground answers in the package's evidence and numbers — **Correct.** Preparation for challenge.

✗ (D) To memorise scripted answers — **Wrong.** Not the point.

Q27 **Answer: D** Review: C7 M3

What is the purpose of a red-team review of the package?

✗ (A) To add more slides — **Wrong.** No.

✗ (B) To rewrite it from scratch — **Wrong.** No.

✗ (C) To approve it automatically — **Wrong.** No.

✓ (D) To find contradictions, weak evidence and missing alternatives before the client does — **Correct.** Pre-emptive critique.

Q28 **Answer: A** Review: C7 M3

A board member asks: "Why not just buy the scheduling agent — our competitor did?" Best answer?

✓ (A) Start from the shared goal, show the data gap, then the route back and trigger — **Correct.** Goal → evidence → route back.

✗ (B) "The CEO is wrong." — **Wrong.** Undermines the sponsor.

✗ (C) "We will buy it later anyway." — **Wrong.** Pre-commits without evidence.

✗ (D) "AI is hype." — **Wrong.** Unsupported.

Q29 **Answer: B** Review: C7 4.1

What turns a client package into a portfolio piece?

✗ (A) Removing the recommendation — **Wrong.** No.

✓ (B) Visible reasoning: decisions, alternatives, graded evidence, hardest finding and limitations — **Correct.** Shows judgment.

✗ (C) Adding the vendor's marketing — **Wrong.** No.

✗ (D) Nicer formatting — **Wrong.** Not enough.

Q30 **Answer: A** Review: C7 Ex 15

Which entry belongs in a reasoning log?

✓ (A) Decision, options considered, graded evidence, reasoning, who contested it, what would change it — **Correct.** Complete entry.

✗ (B) "Recommend inspection." — **Wrong.** Conclusion only.

✗ (C) Vendor price lists — **Wrong.** Not reasoning.

✗ (D) Meeting minutes verbatim — **Wrong.** Too much, wrong focus.

Q31 **Answer: D** Review: C7 Ex 16

Which is the best "what I learned" line in a hardest-finding write-up?

✗ A "AI is hard." — **Wrong.** Generic.

✗ B "Everything went well." — **Wrong.** No learning.

✗ C "The CEO should have listened." — **Wrong.** Blames others.

✓ D "I would profile the data behind any sponsor's favourite idea in week 2." — **Correct.** Specific change in own practice.

Q32 **Answer: C** Review: C7 4.7

Before using real client work in your portfolio you must first…

✗ A Publish and wait for objections — **Wrong.** Breach risk.

✗ B Rename the client — **Wrong.** Necessary, but not first.

✓ C Get written permission and check NDAs — **Correct.** Permission first.

✗ D Delete all numbers — **Wrong.** Scale; keep logic.

Q33 **Answer: A** Review: C7 Ex 17

Your self-assessment shows Advisor-ready on 11 of 13 dimensions but "Developing" on Courage. Is the package strong?

✓ A Not yet — courage weighs most; add clear null recommendations with routes back — **Correct.** Weighted judgment.

✗ B Yes — 11 of 13 is enough — **Wrong.** No dimension should be Developing.

✗ C Only design matters — **Wrong.** No.

✗ D Yes, courage is optional — **Wrong.** Central to advisory.

Q34 **Answer: C** Review: C4–C5–C7

Integrated: The CFO wants the full €1.5m approved now "to save time". Your model's conservative case is negative. What do you recommend?

✗ A Agree — speed matters most — **Wrong.** Ignores downside.

✗ B Cut the programme to inspection only — **Wrong.** Discards evidenced value.

✓ C Keep staging, but pre-agree gate criteria and dates so later tranches release fast when evidence arrives — **Correct.** Speed without abandoning discipline.

✗ D Refuse to discuss — **Wrong.** Unhelpful.

Q35 **Answer: D** Review: C6–C7

Integrated: Counsel has not yet confirmed that inspection is not a safety component. What do you write?

✗ A "Inspection is not high-risk." — **Wrong.** Outside your authority.

✗ B Omit the topic — **Wrong.** Hides a risk.

✗ C Stop the programme — **Wrong.** Disproportionate.

✓ D State the working classification, mark it as a limitation, refer to counsel and make confirmation a Gate 1 condition — **Correct.** Respect scope; plan for it.

Q36 **Answer: B**

Review: C5–C7

Integrated: After Gate 1, line 4 production data shows 12% scrap reduction (plan 18%). What next?

✗ (A) Cancel immediately — **Wrong.** May still be above breakeven.

✓ (B) Re-run the model, compare to breakeven, review causes, and decide at the gate with the owner — **Correct.** Evidence-based gate decision.

✗ (C) Blame the vendor publicly — **Wrong.** Unprofessional.

✗ (D) Hide it and proceed — **Wrong.** Dishonest.

R6 Practice case and further reading

Practice case for self-study: Kessler Verpackung AG (fictional)

Repeat the full capstone on a new case without model answers. Use T7.1–T7.6 and the templates of Courses 1–6.

Element	Facts
Company	Folding-carton and flexible-packaging maker; €380m revenue; EBIT 4.8% (was 7%); 1,600 staff; plants in Germany and Poland.
The ask	Supervisory board wants “a generative AI strategy” after a customer asked about AI-generated artwork.
Ideas raised	AI artwork proofing; demand forecasting; print-defect vision; customer-service chatbot; contract review; predictive maintenance on presses; AI pricing; an “AI design studio” the CMO wants.
Data clues	Proofing loops average 4.2 rounds; 3% of jobs reprinted for colour; forecast error 31%; ERP master data 18% duplicates; sensors on 25% of presses; two customer complaints about data use.
Constraints	Budget up to €1.2m over 18 months; payback < 3 years; 8% discount rate; one data engineer; works council in both countries.

CONSULTANT'S TIP · Hints (read only when stuck)

The CMO’s “AI design studio” is the sponsor’s favourite — test its evidence early. Duplicated master data is the weakest link for forecasting and pricing. Customer data-use complaints shape the data boundary. Reprints for colour are a measurable baseline for vision and proofing. Expect at least three null or redirected recommendations.

Practice checklist

1. Complete Exercises 1–17 on Kessler in 25–30 hours.
2. Score yourself with the Exercise 17 rubric; ask a peer to score independently.
3. Present the two-minute story and the 12-slide deck to a colleague playing the CEO.
4. Add the result, labelled “capstone practice case”, to your portfolio.

Further reading

Resource	Why it is useful
The six earlier study books of this certificate	The frameworks and templates applied here — reread the cheat sheets
Barbara Minto — *The Pyramid Principle*	Governing thoughts, storylines and executive writing
EU AI Act — Reg. (EU) 2024/1689 as amended by Reg. (EU) 2026/1744	Current obligations and dates; check Commission guidance
NIST AI RMF 1.0 and ISO/IEC 42001:2023	Risk and management-system references for governance plans
Tetlock & Gardner — *Superforecasting*	Calibrated judgment and updating on evidence