
EMBA-7450 · INFORMATION TECHNOLOGY AND STRATEGY · MODULE 5

AI Economics and Scaling Discipline

From Pilot to P&L

Gerald C. Kane, Ph.D.

Executive MBA course, Terry College of Business, University of Georgia
Fall 2026

This public edition omits the in-class activity specifications, which are designed to be experienced live in session.

© 2026 Gerald C. Kane. Licensed under CC BY-NC 4.0. You may share and adapt it for noncommercial use with credit to the author; you may not present it as your own. Full terms on the last page.

Module role: Economics and scaling-discipline module. Earlier modules established the 2026 AI landscape, executive AI literacy, technology choices, and governance obligations. This module asks what those choices are worth and what they cost once they leave the pilot environment. It also supplies the economic reasoning and measurement discipline for the capstone: value portfolio, prioritized use case, KPI value chain, ROI logic, full TCO and unit economics, benefits ownership, stage gates, and a 90-day plan. The boundaries hold: Module 3 owns detailed technology architecture; Module 4 owns responsible AI, security, regulatory obligations, and controls; Module 7 owns workflow redesign, talent, incentives, and change leadership. Module 5 touches those issues only where they affect value, cost, and scaling decisions.

Module Learning Objective

By the end of this module, you will be able to move an AI initiative from enthusiasm to investment discipline: classify the initiative by value logic, connect technical performance to business outcomes, identify the full cost of scale, assign benefits ownership, and recommend whether the next executive decision should be stop, pivot, or scale. Every student will complete the AI Business Case's Core fields: use case and portfolio category, baseline, KPI value chain with its weakest link marked, and a stop/pivot/scale call. At the Challenge depth, called the Full Case in this module, the same artifact extends to cost per successful task with the full harness loaded, key sensitivities as ranges, benefits ownership and Finance validation, risk constraints, and the next 90 days. For the Transform-path capstone, the tangible output is the complete one-page AI Business Case that a CFO, business unit leader, CIO, and risk leader could debate without you in the room to explain it.

If You Want to Know More

A useful way to prime the reading is to pick one AI initiative from your organization, live, in pilot, or under consideration, and hold two questions about it as you go: what is the cost of one successful task, with everything around the model loaded in, and what would have to be true for that task to show up in the P&L? The formal preparation sits in "Bring This to Class" at the end of Part 1.

The Pre-Class Read below makes the argument; the materials in the table carry the underlying data and provocation. These references are optional. Read what fits your time.

Source	Why It Matters For This Module
McKinsey, <i>The State of AI in 2026: On the Road to ROI</i> ¹	Enterprise scaling, the flat EBIT-impact share, operating-cost constraints, and what high performers do differently.
PwC, <i>2026 Global CEO Survey</i> ²	CEO-level evidence on uneven AI returns and the leader-laggard divide.
Stanford HAI, <i>2026 AI Index Report</i> ³	Macro adoption, capability acceleration, economic context, and responsible AI trends.
FinOps Foundation, <i>FinOps for AI</i> ⁴	AI spend complexity, allocation, forecasting, unit economics, and governance.
Gartner, <i>Why Half of GenAI Projects Fail: Avoid These 5 Mistakes</i> ⁵	A practitioner framework for examining business value, data readiness, TCO, responsible AI, and change management; do not treat the headline as a universal project failure rate.

Source	Why It Matters For This Module
Deloitte, <i>State of Generative AI in the Enterprise</i> Q4 ⁶	Quarterly practitioner survey on enterprise AI ROI realization and barriers.

BCG's *AI Radar 2026* is useful as optional background on enterprise AI investment patterns.⁷

Part 1: Pre-Class Read

MAIN POINT. AI adoption alone does not establish strategic value. McKinsey reports 44% of surveyed organizations scaling AI across the enterprise, while 37% report enterprise-level EBIT impact, unchanged from the prior edition.¹ The investment question is how AI activity produces an outcome worth its full cost. This module uses a chain from technical capability through adoption and workflow change to operating and financial results. A baseline, benefit owner, complete cost model, and decision gate help test each link. The survey identifies a gap; it does not establish where or why most individual pilots fail.

TERMS. Selected economics and measurement terms not already defined in Modules 1–4 appear in **bold** at their first substantive use in Part 1; each is defined in the Executive Glossary immediately before Part 2. Use the glossary as a lookup aid, not assigned vocabulary to memorize.

Opening Vignette: Aflac and the Unit Economics of Claims

Aflac, the supplemental insurance company headquartered in Columbus, Georgia, offers a useful setting for examining claims unit economics. Its executives have described both automated claims processing and the large share of manually handled claims with small payouts. Those observations invite a practical question: which additional claims could be processed faster and at lower total cost while maintaining quality?

Shelia Anderson, then Aflac's chief information officer, said in a 2023 MIT Sloan Management Review podcast that about 46% of claims were fully automated, primarily wellness claims, while 49% of manually handled claims resulted in payouts below \$200.³⁴ **Straight-through processing** completes a transaction without human intervention. In September 2024, Keith Farley, SVP of Individual Voluntary Benefits, described over 55% of Aflac's claims as automated.³⁴ These are executive-reported operating measures. A small payout limits one possible direct payment error; it does not bound the total cost of a mistake, including rework or customer harm.

The teaching question is what it takes to turn that operating progress into value. Compare the full cost of an automated claim, including exceptions and human review, with a comparable manual claim, and test whether quality holds as volume grows. That is the **benefits realization** discipline this module recommends; the interviews do not independently demonstrate the complete cost and financial-control system. Do you know your unit economics well enough to scale, and how will you measure them after launch?

Unit Economics Over Deal Size

This module recommends moving from deal-size thinking to unit-economics thinking. An AI business case may promise \$X million in savings, \$Y million in new revenue, or payback in Z years. Those totals need a defensible bridge from the demonstration to the income statement: the cost of

one successful task and the value of one captured outcome, multiplied by adoption that holds in daily work.

Asking about unit economics makes an investment proposal testable. The total investment still matters, but the team also needs to explain what changes as volume, quality, and adoption change.

The AI ROI Paradox

A company may have dozens of copilots, chatbots, and pilots while seeing little measurable effect on revenue, margin, working capital, retention, risk, or decision speed. A pilot can be technically successful and economically irrelevant. A tool can be popular with users and still fail to change the workflow that drives cost or revenue.

McKinsey's August 2026 survey reports a persistent gap between individual productivity gains and enterprise financial impact. Its approximately 6% of high performers meet two conditions: respondents attribute at least 5% of EBIT to AI and describe significant value from its use.¹ These are self-reported results, not audited or experimentally established returns. The detailed comparisons and operating-cost findings appear in the Current Horizon Scan.

PwC's 2026 CEO survey also reports uneven financial benefits.² In Deloitte's January 2025 enterprise survey, nearly three-quarters of respondents said their **most advanced GenAI initiative** met or exceeded ROI expectations. Yet more than two-thirds expected **30% or fewer of their experiments** to reach full scale within three to six months.⁶ Those are respondent-level assessments of different questions, not success and scaling rates for all projects. Gartner's practitioner framework points to unclear value, data readiness, TCO, responsible AI, and change management as risks to examine.⁵ Market spending forecasts are dated context and live in Part 2.

MIT CISR's research supplies another perspective: companies at different AI maturity stages had different financial performance. The largest adjacent-stage gap appeared between stages 2 and 3.⁸ This is an association, not a measured ROI gain caused by moving stages. Part 2 explains the study and its limits. For an individual investment, the practical question remains how technical capability becomes a business result.

text AI capability → user adoption → workflow change → operational KPI → financial impact → strategic value

Portfolio Recap: Same Map, New Evidence Standard

Portfolio Category	Module 5 Economic Question	Evidence Required Before Scale
Efficiency	Can the organization capture cost, time, capacity, or working-capital improvement?	Baseline volume, realized workflow change, quality controls, cost per successful outcome, and a path from gross time saved to capturable value.
Effectiveness	Does AI improve judgment, consistency, speed, or quality in a core workflow?	Sustained adoption, acceptance or override patterns, workflow KPI movement, and evidence that users trust the output in the moments that matter.

Portfolio Category	Module 5 Economic Question	Evidence Required Before Scale
Innovation	Does AI create a new revenue stream, customer experience, business model, or strategic option?	Validated customer need, early market signal, unit economics, risk acceptance, learning milestones, and staged funding gates.

Walmart's company-reported examples illustrate different portfolio questions: route optimization and fashion development on the efficiency side, supplier negotiation with Pactum on the effectiveness side, and customer-facing shopping assistance on the innovation side.³⁶ Each needs its own KPI chain. The examples illustrate a portfolio; they do not establish that a portfolio strategy causally outperformed a single-tool strategy. The case detail is in the Use-Case Gallery in Part 2.

What Makes AI Business Cases Different

AI business cases look familiar on the surface: investment, cost, benefit, risk, timeline, owner. Underneath, they behave differently from many traditional technology cases.

First, AI cost can rise with usage. A pilot is cheap because few users or transactions are involved; at scale, inference, context, retrieval, evals, monitoring, retries, human review, and vendor commitments change the economics.

Second, AI value depends on behavior in the workflow. A copilot can be installed without becoming part of the user's daily rhythm. An agent can automate a task while creating exception-handling work somewhere else.

Third, AI requires ongoing evaluation. Models drift, rules change, and agentic systems behave differently as tools, data, and permissions change. The cost of AI is run, observe, improve, govern, and eventually retire, not only build and launch.

Fourth, risk belongs in the cost model. IBM's 2026 breach study provides loss-severity benchmarks for organizations that experienced breaches.⁹ AI insurance can transfer selected risks under specified policy terms.⁴² Neither a study average nor an insurance premium is the total expected loss for your workflow. The council needs its own assumptions about incident probability, exposure, controls, and retained loss. The sources and a practical calculation are in Part 2.

Fifth, AI value may be indirect: reduced risk, faster decisions, protected customer experience, proprietary data assets, or future options. These benefits still need disciplined assumptions and stage gates.

The practical implication is that an AI business case is a living value case with assumptions, measurements, owners, and decision gates, not a static spreadsheet.

KPI Value Chain

The **KPI value chain** is the module's core tool. It prevents executives from leaping from technical performance to financial value. Make adoption explicit and test each link. The enterprise survey results motivate that inquiry; they do not identify the cause of failure in a particular project.

text Technical metric → Adoption metric → Workflow KPI → Financial KPI → P&L, balance sheet, or strategic impact

Technical metrics matter. Accuracy, retrieval quality, groundedness, latency, and output reliability all shape trust, but they are not business outcomes. A model can score well in a test environment and still fail to move the operating metric that matters. Adoption metrics show whether the capability is becoming part of work: activation, weekly use, workflow coverage, suggestion acceptance, override rates, time-to-proficiency, and sustained use after the novelty fades. The right metric depends on the use case. The principle is constant: if adoption is weak in daily work, the financial case is weak.

Work backward through the chain. If the financial KPI is not improving, did the operational KPI move? If not, did users adopt the tool in daily work? If so, did they trust the output, and was the technical performance good enough? If all of that holds, was the value hypothesis wrong? That diagnostic habit prevents a common executive error: blaming the model when the issue is workflow design, incentives, measurement, or a weak business hypothesis. The KPI Value Chain Examples table is in Part 2.

ROI Logic by Type of AI Investment

There is no single AI ROI formula that works for every initiative. The right logic depends on the type of investment.

Efficiency ROI

Efficiency cases begin with the current process: how much work is performed, what it costs, how long it takes, and how much rework, exception handling, or vendor spend it creates. Use a consistent evaluation period:

```text ROI over the evaluation period = (total attributable benefits – total costs) / total costs

Total costs = implementation costs + operating costs over the same period ```

Keep cash savings, avoided future costs, and redeployable capacity separate, and do not count the same staff hour twice. Capacity becomes financial value only through a specified use, such as additional output or avoided hiring. For multiyear choices, compare discounted cash flows or NPV and disclose the discount-rate assumption. The formula is less important than a defensible account of the benefit, its timing, and the full cost of obtaining it.

Forrester's commissioned Total Economic Impact studies of Microsoft 365 Copilot illustrate how assumptions about licensing, adoption, and workflow gains produce a modeled financial result.<sup>10</sup> They are vendor-commissioned composite studies, not observed returns for a representative customer. Calibrate their assumptions against your own evidence. The figures are in the Use-Case Gallery, alongside Chick-fil-A's historical food-safety example, where the proposed unit of value is earlier detection rather than a productivity hour.

## Effectiveness ROI

Effectiveness cases are built around performance lift. The benefit may come from higher conversion, faster first response, better first-contact resolution, fewer defects, shorter review cycles, better forecast accuracy, or faster onboarding. A useful structure is:

text Estimated operating benefit = affected volume × adoption rate × attributable workflow KPI lift × financial value per unit – AI operating cost

Use the same period for volume, benefits, and costs. This structure makes adoption and the unit of value explicit. It estimates operating benefit; a full investment return also includes implementation costs and the timing of cash flows. A tool with a large theoretical benefit and 10% sustained adoption may be less valuable than a narrower tool with 80% adoption in a high-value workflow.

Vanguard's leaders report value from a portfolio spanning data, analytics, and AI.<sup>11</sup> Its coding-productivity estimate is narrower than a whole-job productivity measure; the gallery gives the distinction. For a portfolio like this, the discipline recommended here is to identify a business owner, baseline, and measurement cadence for each use case. Those are questions to test against company evidence, not details established for every Vanguard use case by the cited interview.

## Innovation ROI

Innovation cases should be governed through staged funding. The first business case may not be a case for full scale. It may be a case for learning whether a customer problem is real, whether the offering can be delivered safely, whether unit economics could work, or whether the company has a distinctive advantage.

The question to ask is "What must we learn next, what will that learning cost, and what decision will it enable?" rather than "Can we prove the full ROI today?" If the next stage does not increase or decrease confidence in a meaningful assumption, it is not a **stage gate**; it is continued exploration.

Build versus buy has its own economic profile in innovation cases. Buying can reduce development work while shifting talent needs toward integration, evaluation, security, procurement, and vendor management. Building carries continuing maintenance and engineering obligations, and can still depend on external models. Both can face **model depreciation**. Make the sourcing choice with unit economics and capability requirements in view. The Build vs. Buy table is in Part 2.

An AI investment ask also competes in the same capital allocation conversation as a buyback, a tuck-in acquisition, or a plant build, and it should be defensible against the firm's hurdle rate on the same terms. The Capital Allocation Frame in Part 2 carries the 2026 deal-market context.

## Ranges, Not Point Estimates

A number that looks precise is not the same as a number that is right. AI business cases are unusually sensitive to a few behavioral assumptions, and each is a guess at the pilot stage. Present the case as a range with named drivers: a base case built on the evidence in hand, a downside case in which adoption stalls or review costs stay high, and an upside case in which the workflow absorbs the tool. A council reviewing a point estimate can only accept or reject it. A council reviewing a range can direct the next 90 days of evidence-gathering at the assumption that moves the answer most.

Four levers move most cases: adoption, accuracy, oversight cost, and volume. The Worked Example in Part 2 explains each and runs the arithmetic across scenarios; the executive habit is to ask which lever the case is most exposed to and what evidence would narrow the range.

## AI FinOps and the Full Cost of Scale

---

AI FinOps, introduced in Module 3, connects AI consumption and spending to business value. In the FinOps Foundation's 2026 practitioner survey, 98% of respondents managed AI spend, versus 31% two years earlier, and AI cost management was a leading skill need.<sup>4</sup> FOCUS 1.3 added a Contract Commitment dataset in December 2025, providing a standard representation of **committed spend and capacity reservation** terms alongside cost and usage data. This addresses a tooling need; it does not imply that organizations previously had no way to track their contracts.

The executive lesson is to understand the unit economics of AI value rather than to make AI cheap. A low-cost model that produces unreliable answers can be expensive per successful outcome. A more expensive model that reduces rework, escalations, or risk may be cheaper in economic terms.

Two cost questions deserve explicit attention. **Capitalize vs. expense** depends on the reporting framework and the project's facts: FASB's 2025 software-accounting update addresses significant development uncertainty but is not a blanket instruction to expense AI development.<sup>12</sup> Model depreciation can require migration and renewed evaluation on a vendor's schedule; notice periods depend on the product and provider.<sup>43</sup> Part 2 separates accounting effective dates, consumer-product changes, API retirements, and planning assumptions.

A per-person tool budget is one control over consumption; it still needs a link to useful work. Compare spending per accepted outcome across workloads and investigate what drives variation. The Full TCO View and the spending-cap example in Part 2 turn that discipline into questions about budgets, quality, and delivery.

### Unit Economics

For executives, unit economics are often more useful than total AI spend. The relevant unit might be a resolved case, accepted recommendation, reviewed contract, forecast cycle, converted lead, avoided escalation, or successful answer. The metric that follows is **cost per successful task**:

text Cost per successful outcome = total AI operating cost / number of accepted, useful, or completed outcomes

The numerator should include the full harness, not only model calls:

text Total AI operating cost = model calls + context/retrieval + cache writes and reads + tool calls + evals + monitoring + human review + exception handling + rework + governance and support

Agentic tasks can involve repeated model calls and growing context. A 2026 preprint studying eight language models on coding tasks found roughly 1,000-fold greater token consumption than its code-chat comparison and up to 30-fold variation across runs of the same task.<sup>14</sup> These are results for the study's benchmark, not universal multipliers for enterprise agents. Measure the actual

distribution of cost per completed task before extrapolating a chatbot budget to an agentic workflow.

**KEY LESSON.** Track cost per successful task, not cost per prompt, token, or model call. The cheap-model fallacy, that lower per-call cost equals lower system cost, collapses once retries, retrieval failures, and human cleanup are loaded into the unit.

## Harness Economics

**Harness economics** are the cost consequences of making AI reliable enough to use repeatedly. They are easy to miss because many pilots price only the visible model call. Production systems also pay for the information environment and operating scaffolding around the model. The Harness Economics table is in Part 2.

Caching is a clear place to see the logic without getting technical. A recurring workflow often has a stable prefix: system prompt, policies, source hierarchy, examples, output format. If that prefix is paid in full on every call, the organization is renting the same context over and over. Designed well, caching cuts the marginal cost of repeated work materially; designed poorly, it raises cost and preserves stale instructions. The executive question is which context is stable enough to cache, who owns invalidation, and how you know the cache is saving money.

A system with short context, a high cache hit rate, good evals, and low rework may be cheaper than one with a cheaper model but weak retrieval, long prompts, retries, and heavy cleanup. Before approving scale, you should be able to say what the unit of value and the unit of cost are, how usage will change with adoption, and what cost signal would trigger a pivot in model, vendor, architecture, or workflow.

## Benefits Realization Discipline

---

A business case is a promise to change something. Benefits realization is the management system that keeps the promise from dissolving after approval.

Every AI initiative moving beyond proof of concept needs three anchors: a **baseline**, a **benefit owner**, and a measurement cadence. The baseline describes performance before AI. The benefit owner is the business leader accountable for changing the workflow and capturing value. Compensation linkage is one possible accountability mechanism; Sysco's gallery entry provides an example, while Chewy illustrates how leaders can constrain the financial interpretation of savings. The cadence determines when the KPI chain, adoption, cost, risk, and learning are reviewed, with Finance involved in validating measurement before a number enters budgets.

Baselines do not need to be elaborate, but they need to be real: average handle time before agent assist, forecast error before AI planning, contract review cycle time before a legal assistant, conversion before lead scoring, claims cost before automation. Without a baseline, teams measure improvement against memory, anecdote, or enthusiasm.

## Value Leakage

**Value leakage** occurs when the AI capability appears to work but the economic benefit is not captured. The tool may be used but outside the workflow. The workflow may change but incentives pull people back to the old process. Output may require so much review that savings disappear. Usage cost may rise faster than benefit. The claimed value may never be accepted by Finance. Risk or compliance issues may slow deployment until the economic window closes. Value can also leak through mismeasurement: an initiative that promised hours and delivered better decisions is scored as a failure at the investment council and gets killed for the wrong reason.

Shadow AI can create duplicate spending on tools outside enterprise procurement and expose data through workflows the organization does not oversee. Treat those as separate hypotheses to measure: reconcile tool usage and contracts, and review incident and access-control evidence. IBM's study provides breach-cost context, not an organization-specific estimate of either leakage channel.<sup>9</sup> The Measurement Cadence table is in Part 2.

## Stop, Pivot, or Scale

---

Executives need permission to stop AI projects, and you may need to give it before anyone asks. Stopping is not failure if the project produced useful learning before it consumed too much capital, credibility, or scarce talent. The failure is letting weak projects survive because no one created a decision gate.

The stop/pivot/scale discussion should be evidence-based, not personality-based. A project should stop when the **value pool** is too small, the business owner is missing, adoption remains weak after reasonable support, full **TCO** exceeds likely benefit, unresolved data or integration problems make scale unrealistic, risk constraints block production, or a simpler non-AI solution would solve the problem more economically.

Klarna illustrates a pivot in service design. In May 2025, Bloomberg reported CEO Sebastian Siemiatkowski's concern that cost reduction had gone too far and his plan to expand access to human service. Klarna continued using its AI assistant.<sup>37</sup> The case invites a joint test of cost and service quality; it does not establish that the AI program stopped or that the company's earlier claimed savings were independently validated. The detail is in the Cautionary Tales in Part 2.

A project should pivot when part of the capability is useful but the original scope is wrong: the target workflow is too broad, the tool needs to move inside the system where users already work, or context pruning, caching, model routing, process redesign, or a vendor change would improve the economics.

A project should scale only when the value chain is working: the use case is tied to a strategic priority, adoption is happening in daily work, the operational KPI is moving, unit economics are credible at production volume, a business leader owns the benefit, Finance accepts the measurement logic, and risk constraints are managed.

**KEY LESSON.** Scale when evidence supports the KPI value chain from technical capability through adoption and workflow change to operating and financial results. A demonstration is evidence about a limited setting. Production measurement tests whether those results hold in daily work.

The Stage Gates table is in Part 2.

## Portfolio Governance and the AI Investment Council

An **AI Investment Council** is not the same as an AI risk committee. A risk committee asks whether an initiative is safe, compliant, accountable, and appropriately controlled. An investment council asks whether the initiative deserves capital, talent, and management attention.

Its value is cadence and comparability: a place where Business, Finance, Technology, Data, Legal/Risk, HR or Change, and the operating unit look across the portfolio together. Its work can draw on NIST's GOVERN/MAP/MEASURE/MANAGE functions and the complementary ISO/IEC 42001 management-system standard, both covered in Module 4.<sup>1516</sup> ISO certification is distinct from using the voluntary NIST framework.

JPMorgan Chase illustrates the scale that ongoing oversight may need to support. A June 2025 interview described LLM Suite access for roughly 200,000 employees; a later interview described adding capabilities every few weeks.<sup>35</sup> Neither disclosure establishes the cadence of an investment council. The recommendation here is to review decisions often enough to respond to material changes in cost, capability, adoption, and risk. The gallery separates the platform disclosures from the bank's broader AI/ML value estimates.

The council's review operates against the authority-and-evidence frame from Module 4: any initiative recommending an authority-level increase requires evidence that justifies it, and the council holds that line. It maintains the AI value portfolio, prioritizes use cases against strategy and readiness, approves funding by stage, requires KPI chains and baselines, reviews FinOps assumptions, tracks value leakage, and makes stop/pivot/scale decisions. Just as important, it notices dependencies that project teams may not see: shared data assets, platform gaps, governance bottlenecks, vendor concentration, change capacity, and scarce technical talent. A useful monthly conversation asks what has changed since the last review, which initiatives now have stronger or weaker evidence, which are consuming scarce talent without producing value or learning, which strategically important bets are underfunded because easier productivity tools are absorbing attention, and what should be stopped this month.

Two companions sit beside the council. NACD's board-level framework addresses AI strategy, capital allocation, AI risks, and technology competency.<sup>17</sup> Its 2025 survey found 47% of directors identified selecting AI tools that deliver ROI as a current challenge. Regulatory compliance is also a cost line for the council. Module 4 owns the obligations; the Horizon Scan and Regulatory Cost Reference in Part 2 connect selected obligations to investment planning.

## Bring This to Class

---

Pick one AI initiative from your organization, live, in pilot, or under consideration, and read this module through it. Every student should bring the AI Business Case's Foundation/Core fields:

1. **Use case and portfolio category.** Name the initiative, the business process it affects, and whether it is an efficiency, effectiveness, or innovation bet.
2. **Baseline.** State current performance before AI: cost, cycle time, quality, revenue, risk, or capacity. If you do not have it, write down who would own producing it.
3. **KPI value chain.** Draft the chain from technical metric to business impact and mark the link you are least able to evidence today.
4. **Stop / pivot / scale call.** State which decision you would make at the next investment review and what evidence would change it. Be ready to argue against your own recommendation.

For the Challenge depth, called the Full Case in this module, extend the same artifact with cost per successful task with the full harness loaded, as a range rather than a point; the two sensitivities the case is most exposed to; benefits ownership and the Finance partner who would validate the number; the risk constraints; and the next 90 days. This is deeper analysis of one artifact, not a separate graded pathway or a new assignment.

Optional, at either depth: ask your AI assistant to stress-test your value hypothesis with three alternative explanations for the effect you expect. Bring the strongest one. The prompt is in Part 2.

## A Final Word Before Class

---

We have given you a portfolio map, a value chain, three ROI lenses, a cost model, and a set of stage gates. These are useful tactics, and the capstone will use them. They should not obscure the higher-altitude point. AI economic discipline is the same discipline that makes any major capital allocation defensible: a clear unit of value, a clear unit of cost, an owner accountable for both, evidence that the assumptions still hold, and the willingness to stop something before it consumes more capital than it teaches. The tools in this module are how that discipline gets executed where AI cost rises with usage, value depends on workflow behavior, and the model itself is rarely the moat. The discipline is older than AI. Your job is to bring it.

## Executive Glossary

---

Use this alphabetized glossary as a lookup aid for Part 1. It includes only terms not already defined in Modules 1–4; definitions reflect how the terms are used here.

| Term                         | Plain-language definition                                                                                                                                                |
|------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>AI Investment Council</b> | The management-layer body that decides which AI initiatives deserve capital, talent, and attention, distinct from the risk committee that decides whether they are safe. |
| <b>Baseline</b>              | The measured performance of a workflow before AI is introduced, against which any claimed improvement is judged.                                                         |
| <b>Benefit owner</b>         | The business leader accountable for changing the workflow and capturing the value an AI initiative promises.                                                             |

| Term                                          | Plain-language definition                                                                                                                                                                             |
|-----------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Benefits realization</b>                   | The management system of baseline, owner, and review cadence that keeps a business case's promise from dissolving after approval.                                                                     |
| <b>Capitalize vs. expense</b>                 | Whether a development cost is recorded as an asset and written down over time or charged against earnings in the period incurred.                                                                     |
| <b>Committed spend / capacity reservation</b> | Contract structures in which a buyer commits to a spending volume or reserves compute capacity in advance, usually in exchange for a discount, taking on the risk of paying for what it does not use. |
| <b>Cost per successful task</b>               | Total AI operating cost, including the full harness, divided by the number of accepted, useful, or completed outcomes.                                                                                |
| <b>Harness economics</b>                      | The cost consequences of the scaffolding that makes AI reliable enough to use repeatedly: context, caching, retrieval, evals, monitoring, review, and retries.                                        |
| <b>KPI value chain</b>                        | The sequence by which technical performance becomes adoption, adoption becomes workflow change, workflow change moves an operational KPI, and the KPI produces financial or strategic impact.         |
| <b>Model depreciation</b>                     | A provider ending support or availability for a model, which may require migration and revalidation; retrieval reindexing may be needed if an embedding model or representation changes.              |
| <b>Stage gate</b>                             | A decision point at which an initiative must show specific evidence before receiving more funding, scope, or authority.                                                                               |
| <b>Straight-through processing</b>            | A transaction completed end to end by an automated system with no human touch.                                                                                                                        |
| <b>TCO (total cost of ownership)</b>          | The full cost of an AI capability over its life, including run, review, governance, migration, and retirement, not only build and launch.                                                             |
| <b>Value leakage</b>                          | The gap between the benefit an AI capability produces on paper and the benefit the organization captures in its results.                                                                              |
| <b>Value pool</b>                             | The estimated total benefit available in the defined workflow and period under stated assumptions. Adoption determines how much of that opportunity the initiative may capture.                       |

## Part 2: Field Reference

The Field Reference is the leave-behind. Use it during the course to scaffold board memos, investment-council reviews, and pilot stop/pivot/scale decisions; keep it after the course as an executive reference. Each component is self-contained — flip directly to a table, prompt list, or worksheet without reading surrounding prose.

### One-Page AI Business Case Template

A blank, paste-ready worksheet. Fill each field for one AI initiative; the completed version is the AI Business Case card format (see Part 3). Foundation/Core rows are the common starting point for every student; Challenge/Full Case rows extend the same artifact and are deeper analysis, not a separate graded pathway or a new assignment.

| Depth               | Field                                                                                                                                     | Your Entry |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------|------------|
| Foundation/Core     | Use case                                                                                                                                  |            |
| Foundation/Core     | Portfolio category (efficiency / effectiveness / innovation)                                                                              |            |
| Foundation/Core     | Strategic rationale                                                                                                                       |            |
| Foundation/Core     | Baseline (current cost, cycle time, quality, revenue, risk, or capacity)                                                                  |            |
| Foundation/Core     | Value hypothesis (one sentence)                                                                                                           |            |
| Foundation/Core     | KPI Value Chain (technical → adoption → workflow → financial → business impact), weakest link marked                                      |            |
| Foundation/Core     | Stop / pivot / scale criteria                                                                                                             |            |
| Challenge/Full Case | Full TCO assumptions (model usage, context/caching, integration, evals, monitoring, human review, training, governance)                   |            |
| Challenge/Full Case | Cost per successful task, as a range                                                                                                      |            |
| Challenge/Full Case | Key sensitivities (the two assumptions among adoption, accuracy, oversight cost, and volume the case is most exposed to, each as a range) |            |
| Challenge/Full Case | Benefits ownership (business owner, Finance partner, measurement cadence)                                                                 |            |
| Challenge/Full Case | Finance validation                                                                                                                        |            |
| Challenge/Full Case | Risk constraints                                                                                                                          |            |
| Challenge/Full Case | Next 90 days                                                                                                                              |            |

## Decision Tables

### KPI Value Chain Examples

| AI Initiative                 | Technical Metric                                                 | Adoption Metric                                               | Workflow KPI                                                 | Financial KPI                                             | Business Impact                                        |
|-------------------------------|------------------------------------------------------------------|---------------------------------------------------------------|--------------------------------------------------------------|-----------------------------------------------------------|--------------------------------------------------------|
| Demand forecasting            | Forecast error, bias, latency                                    | Planner use, override rate, forecast review cadence           | Inventory days, stockouts, expedite frequency                | Carrying cost, lost sales, working capital                | Lower working capital and fewer avoidable stockouts    |
| Customer service agent assist | Answer accuracy, groundedness, response latency                  | Agent active use, suggestion acceptance, escalation rate      | Average handle time, first-contact resolution, quality score | Cost to serve, retention, training cost                   | Lower operating expense and better customer experience |
| Sales opportunity copilot     | Lead-score precision, account insight quality                    | Seller weekly use, recommendation acceptance, CRM update rate | Activity quality, conversion rate, deal cycle time           | Revenue per seller, win rate, forecast accuracy           | Higher revenue productivity                            |
| Contract review assistant     | Clause extraction accuracy, hallucination rate, citation quality | Reviewer use, acceptance rate, rework rate                    | Review cycle time, issue detection, backlog                  | Outside counsel spend, internal capacity, deal cycle cost | Lower SG&A, faster contracting, reduced legal risk     |
| AI personalization            | Recommendation relevance, click prediction                       | Customer exposure, engagement, opt-out rate                   | Conversion, average order value, repeat purchase             | Customer lifetime value, revenue per visitor              | Higher revenue and stronger loyalty                    |
| Internal knowledge assistant  | Retrieval precision, citation quality, answer usefulness         | Active use, repeat use, abandonment rate                      | Search time, resolution time, duplicate work avoided         | Productivity capacity, onboarding cost, cycle time        | Faster decisions and better knowledge reuse            |

### Full TCO View

| Cost Area               | What To Include                                                                                                                | Why It Matters                                                                  |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|
| Model access            | API fees, subscriptions, enterprise licenses, model hosting, model commitments                                                 | Production economics can surprise Finance if the pilot used a narrow user base. |
| Tokens and inference    | Input tokens, output tokens, embeddings, reranking, multimodal inputs, agent loops, repeated calls                             | Usage growth can turn a low-cost tool into a high-variable-cost service.        |
| Context and caching     | Context-window size, stable prefixes, prompt or context caching, cache writes, cache reads, invalidation, long-context pricing | Repeatedly paying for the same context can quietly dominate unit economics.     |
| Data and integration    | Data cleaning, connectors, RAG systems, vector stores, metadata, access controls, pipelines                                    | The model may work in isolation and fail inside the live workflow.              |
| Workflow implementation | Product design, systems integration, process redesign, exception handling                                                      | Availability is not the same as adoption.                                       |
| Human review            | Quality assurance, approvals, escalation, audit review                                                                         | Automation can simply move work into hidden review queues.                      |

| Cost Area                       | What To Include                                                                                                                   | Why It Matters                                                                                                                                               |
|---------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Monitoring and evaluation       | Evals, test sets, monitoring, drift detection, usage analytics, incident review                                                   | Performance can degrade silently after launch.                                                                                                               |
| Governance and compliance       | Legal review, security review, privacy controls, documentation, audit trails                                                      | Unresolved risk can block scale even when the tool works.                                                                                                    |
| Change and training             | User training, manager enablement, communications, support, adoption analytics                                                    | Initial excitement often fades without operating support.                                                                                                    |
| Vendor and procurement          | Contract minimums, capacity reservations, lock-in, SLAs, vendor risk                                                              | The organization can overcommit before economics are clear.                                                                                                  |
| Model deprecation and migration | Replacement-model cost changes; revalidation of evaluations, prompts, and contracts; retrieval reindexing where embeddings change | Use vendor terms and a migration scenario to budget. One major migration in 18 months can be a stress-test assumption, not an established industry interval. |
| Ongoing operations              | Product owner, support, maintenance, prompt and retrieval updates, continuous improvement                                         | No one should assume the system owns itself after launch.                                                                                                    |

## Harness Economics

| Economic Driver   | Why It Matters                                                                                                                                 | Metric To Ask For                                                                   |
|-------------------|------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Context length    | Longer prompts, retrieved documents, tool outputs, and history increase input cost and can reduce quality when signal is diluted               | Average input tokens per successful task and unsupported-claim rate by context size |
| Caching           | Stable system prompts, policies, examples, and reference packets can often be reused at lower cost; stale cache can also preserve old mistakes | Cache-hit rate, cache-read savings, cache-write cost, and invalidation events       |
| Retrieval quality | Poor retrieval causes wrong answers, extra calls, human cleanup, or lost trust                                                                 | Retrieval precision/recall, stale-source rate, missing-source rate                  |
| Evals             | Test sets, graders, red-team runs, and regression checks are recurring costs, not launch costs                                                 | Eval cost per release and regression failure rate                                   |
| Monitoring        | Logs, traces, dashboards, alerts, and incident review are part of run cost                                                                     | Monitoring cost per production workflow and time to detect failure                  |
| Human review      | Review queues may absorb work that automation was supposed to remove                                                                           | Review minutes per output, reversal rate, and queue backlog                         |
| Retries and loops | Agents can repeat calls, try tools, fail, retry, or hand off between steps                                                                     | Calls per completed task, timeout rate, and loop-abort rate                         |

## Worked Example: Cost Per Successful Task

A hypothetical contract-review assistant handles 10,000 eligible contracts a month. 70% of eligible volume is routed to the tool (adoption rate). Of those routed, 90% complete without a retry and 10% require one retry (retry rate). Of completed runs, 20% still require human review before the output is accepted (human-review rate). The math:

- Eligible volume routed to AI:  $10,000 \times 70\% = 7,000$  contracts
- Retries:  $7,000 \times 10\% = 700$  additional AI calls
- Total AI calls (model + retrieval + tool calls): 7,700, at a blended cost of \$0.80 per call = **\$6,160**

- Human review:  $7,000 \times 20\% = 1,400$  contracts reviewed at 15 minutes each, at a fully loaded reviewer cost of \$60/hour = **\$21,000**
- Monitoring and evals allocated to this workflow: **\$1,200/month**
- Total AI operating cost:  $\$6,160 + \$21,000 + \$1,200 = \mathbf{\$28,360}$
- Successful outcomes (contracts completed to acceptance, including reviewed ones): 7,000
- **Cost per successful task:  $\$28,360 / 7,000 \approx \$4.05$**

Against the assumed manual cost of \$9.50, the operating-cost comparison remains favorable after human review is included. The full \$4.05 cost is about 4.6 times the \$0.88 model-only cost. Human review is the largest cost component, so pricing only model calls would materially understate the expense. This illustration assumes all 7,000 routed contracts eventually meet the same acceptance standard as manual work. A full investment decision must also include implementation, change, residual risk, and any other costs outside this monthly operating example.

### Sensitivity: How the Four Levers Move the Number

Four levers move most AI business cases. Adoption determines how much of the eligible volume the economics apply to. Accuracy determines the retry rate and the share of output that needs human review, and through them most of the harness cost. Oversight cost prices the review queue itself. Volume amortizes the fixed harness of evals, monitoring, and governance across more successful tasks.

The \$4.05 figure is a base case, not a fact. The same arithmetic, run under different assumptions on adoption, accuracy, oversight cost, and volume, shows where the case is exposed:

| Scenario             | What Changes                                                  | Cost per Successful Task | Case vs. \$9.50 Manual Baseline                                 |
|----------------------|---------------------------------------------------------------|--------------------------|-----------------------------------------------------------------|
| Base case            | 70% adoption, 10% retry, 20% review at 15 min / \$60 per hour | $\approx \$4.05$         | Clears comfortably                                              |
| Accuracy slips       | Retry rate 25%, review rate 35%                               | $\approx \$6.42$         | Still below baseline; savings per task fall about 43.5%         |
| Oversight cost rises | Review at 25 minutes and \$75 per hour                        | $\approx \$7.30$         | Approaching breakeven                                           |
| Adoption stalls      | 40% of eligible volume routed                                 | $\approx \$4.18$         | Routed volume falls 42.9%; the eligible value pool is unchanged |
| Volume doubles       | 20,000 eligible contracts, same rates                         | $\approx \$3.97$         | Fixed harness amortizes; case strengthens                       |

Read the table for the pattern, not only the decimals. Accuracy and oversight assumptions affect unit cost; adoption affects captured volume; eligible volume affects the potential value pool. In the adoption scenario, monthly savings on routed work fall from \$38,140 to \$21,280 under the stated assumptions. A case that clears only in its base scenario warrants more evidence before scale, and its stop/pivot/scale criteria should name the relevant boundary.

## Stage Gates

| Gate              | Question                                         | Evidence Required                                                              | Possible Decision                      |
|-------------------|--------------------------------------------------|--------------------------------------------------------------------------------|----------------------------------------|
| Strategic fit     | Is this use case worth exploring?                | Clear business problem, portfolio fit, executive sponsor, rough value pool     | Approve discovery, reject, or redirect |
| Proof of value    | Can AI produce useful output for this problem?   | Technical feasibility, sample outputs, early user feedback, risk screen        | Continue, narrow, or stop              |
| Pilot             | Does it work in a real workflow with real users? | Adoption, workflow KPI movement, quality, exception rates, early TCO           | Pivot, move to production, or stop     |
| Production        | Can it run safely, reliably, and economically?   | Monitoring, support model, cost controls, governance acceptance, benefit owner | Scale, limit scope, or stop            |
| Scale and harvest | Is enterprise value being realized?              | Financial validation, sustained adoption, operating metrics, unit economics    | Expand, optimize, sunset, or replace   |

## Measurement Cadence

| Stage            | Review Rhythm        | Main Question                                                   |
|------------------|----------------------|-----------------------------------------------------------------|
| Proof of value   | Weekly or biweekly   | Does the capability produce useful outputs in a narrow setting? |
| Pilot            | Biweekly or monthly  | Are users adopting it, and is the workflow KPI moving?          |
| Production       | Monthly              | Are cost, quality, risk, and operational KPIs within range?     |
| Scale            | Monthly or quarterly | Is realized value showing up in financial or strategic metrics? |
| Mature operation | Quarterly            | Should the system be optimized, expanded, retired, or replaced? |

## AI Mode → ROI Lens

| AI Mode    | Efficiency Lens                                                            | Effectiveness Lens                                                  | Innovation Lens                                                      |
|------------|----------------------------------------------------------------------------|---------------------------------------------------------------------|----------------------------------------------------------------------|
| Analytical | Operating cost reduction through better forecasting, scoring, or detection | Decision quality lift in pricing, risk, churn, demand, or fraud     | New analytical products or insight-as-a-service offerings            |
| Generative | Time savings in drafting, summarizing, translating, and synthesizing       | Quality lift through better drafts, faster review, broader coverage | New content products, multilingual reach, or new client-facing tools |
| Agentic    | Workflow automation that reduces handoffs and exception cost               | End-to-end task completion that improves cycle time and consistency | New service models where agents act on the customer's behalf         |

## Build vs. Buy

| Axis                 | Build                                                 | Buy                                                              |
|----------------------|-------------------------------------------------------|------------------------------------------------------------------|
| Data sensitivity     | High when proprietary data is the differentiator      | Acceptable when data flows can be controlled in vendor contracts |
| Workflow specificity | High when the workflow is the moat                    | Acceptable when the workflow is largely standardized             |
| Talent availability  | Requires sustained AI engineering and platform talent | Shifts talent need toward integration and vendor management      |

| Axis                        | Build                                                                                          | Buy                                                                                                                                                              |
|-----------------------------|------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Vendor concentration risk   | Depends on model, cloud, and infrastructure dependencies even when application code is owned   | Depends on the vendor portfolio, portability, and contractual exit rights                                                                                        |
| Cost profile                | Heavy upfront investment plus a long post-deployment cost tail (maintenance, evals, migration) | Subscription or consumption pricing; faster to value but ongoing                                                                                                 |
| Speed to value              | Slower; full capability cycle                                                                  | Faster; productized capability already exists                                                                                                                    |
| Differentiation potential   | High when the build matches a unique problem                                                   | Lower; competitors can buy the same thing                                                                                                                        |
| Outcome-based contract test | Not applicable; you carry the delivery risk yourself                                           | Apply Gartner analyst Tom Coshow's question before accepting outcome pricing: if the vendor is not taking on the risk, why is the contract outcome-based at all? |

## Contract Structures

| Structure                                                            | What You Pay For                                      | Cost-Risk Profile                                             | Finance Treatment Notes                                                                                                                                         |
|----------------------------------------------------------------------|-------------------------------------------------------|---------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Committed-spend (cloud / hyperscaler enterprise discount programs)   | Multi-year volume commitment in exchange for discount | Risk of overcommitment and unused minimum spend               | Service commitments generally create operating costs; assess prepayments and other contract components separately. FOCUS 1.3 supports commitment tracking.      |
| Capacity reservation (provisioned throughput, dedicated instances)   | Reserved compute capacity at fixed rate               | Risk if utilization is below forecast                         | A reservation alone does not establish capitalization. Assess the actual service, lease, prepayment, and development components under the applicable framework. |
| Token / consumption pricing                                          | Pay-per-use for inference and API calls               | Cost rises with adoption; exposure to vendor price changes    | Opex; harder to forecast without unit-economics model                                                                                                           |
| Outcome / output-based pricing (per resolved case, per conversation) | Pay per successful unit of work                       | Cost transparency; vendor and buyer must agree on definitions | Opex; depends on resolution definitions and dispute mechanism                                                                                                   |
| Hybrid (base + consumption + outcome)                                | Mix of the above                                      | Designed to balance predictability and elasticity             | Opex; requires more sophisticated FinOps allocation                                                                                                             |

## Four Cost-Management Artifacts and One Named Role

The FinOps discipline in the Pre-Class Read becomes operational through four artifacts and one named role proposed below. Consumption and outcome pricing can coexist with seat pricing, so the team needs a view of the actual contracts and workloads. Engineering, Finance, and Procurement should set a review cadence that matches the rate of spending and operational change.

| Artifact      | What it is                                                                                                                                                                                         | Cadence   | Owner        |
|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|--------------|
| Subsidy Audit | Identifies which workloads would become economically untenable if current pricing were repriced to reflect underlying cost; names exposed contracts, ranks them by repricing risk, assigns owners. | Quarterly | CFO with CIO |

| Artifact                  | What it is                                                                                                                                                                                                              | Cadence                                       | Owner                   |
|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------|-------------------------|
| Cost Scoreboard           | An operational view of AI consumption — tokens, dollars, jobs, hits per workload — that sits next to engineering metrics rather than the P&L. An observability tool, not a cost-cutting tool.                           | Weekly                                        | Engineering with FinOps |
| Cheap-Model Bake-off      | Tests cheaper, smaller, or open-weight models against frontier defaults on the firm's actual production workloads, to keep the model-selection decision live rather than drifting into lock-in.                         | Quarterly                                     | Model Sommelier         |
| Escape-Hatch Architecture | Routes routine work to cheaper models by default with an explicit escalation path to frontier models or human review; a design discipline built into the routing layer before it is needed, not an emergency procedure. | Standing design                               | Engineering             |
| Model Sommelier (role)    | A designated owner of model selection, vendor price negotiation, and architectural routing — a procurement-and-architecture function, not a research one.                                                               | Weekly with engineering; monthly with finance | Named individual        |

The Model Sommelier is the module's proposed name for an owner of model selection, vendor economics, and routing. The responsibility is to keep choices explicit as terms and capabilities change, with the authority to recommend a different model or architecture when evidence warrants it. The title is optional; the accountable function matters. A multi-model posture still needs evaluations, portability, and a business reason for the additional complexity.

The Subsidy Audit connects to Module 6's question about software contracts exposed to consumption repricing. The 98% figure in the FinOps practitioner survey shows how widely AI spending features in those respondents' work; it does not establish how many organizations lack an accountable owner.<sup>4</sup>

**KEY LESSON.** The discipline is not to make AI cheap. It is to keep the model-selection and routing decision live, owned, and repriceable — so cost tracks value as vendor pricing moves.

## Do/Don't Quick Reference

| Do                                                                                                           | Don't                                                                     |
|--------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|
| Target pilots at value pools where the operating metric is named, owned, and measured against a baseline     | Target pilots at visible tasks because they are easy to demo              |
| Load the full harness into TCO — context, caching, retrieval, evals, monitoring, review, retries, governance | Price only the model call and discover the rest at production             |
| Assign the benefit owner in the business before approving the build                                          | Run the program from IT or innovation alone with no accountable P&L owner |
| Track cost per successful task as the unit economic                                                          | Track cost per prompt, token, or model call and call it unit economics    |
| Present investment estimates as ranges with named sensitivity drivers                                        | Present a single-point ROI figure and let precision stand in for evidence |
| Build a stop/pivot/scale gate before the next investment ask                                                 | Let weak projects survive because no one made a decision gate             |
| Match the AI Investment Council cadence to the rate of model and vendor change                               | Review the AI portfolio annually and let decisions be made elsewhere      |

| Do                                                                            | Don't                                                                   |
|-------------------------------------------------------------------------------|-------------------------------------------------------------------------|
| Price regulatory compliance and incident risk as TCO lines                    | Treat risk and compliance as soft constraints discovered at scale       |
| Budget for vendor-specific retirement notices and a migration stress scenario | Treat a fixed 18-month migration interval as an observed universal rule |

## Evidence Quality Ladder

Students and executives will bring uneven evidence into a business case. This ladder describes progress toward financial realization. At every rung, separately assess causal attribution: whether AI caused the change, whether comparison groups are credible, and whether other changes explain the result. A later rung does not automatically provide stronger causal evidence.

| Rung                       | What It Looks Like                                                                             | What It Can Support                                                                         |
|----------------------------|------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| Anecdote                   | "A few people said it worked well."                                                            | A discovery conversation, not a funding ask.                                                |
| Pilot sample               | A small, non-representative group tried it; results are directional.                           | Proceeding to a structured pilot with a named baseline.                                     |
| Workflow telemetry         | Usage and outcome data from the live workflow, not a lab setting.                              | A pilot-to-production decision, if adoption and workflow KPI both show signal.              |
| Finance-validated baseline | Finance has reviewed and accepted the before-state numbers the case is measured against.       | A production investment ask; the council can trust the denominator.                         |
| Production run-rate        | Sustained performance data from live production, not a launch window.                          | A scale decision.                                                                           |
| Budget-realized value      | A benefit is reflected in actual financial results and the budget, with assumptions reconciled | Evidence of financial capture; attribution and the full cost basis still require validation |

A weak case presents early evidence with more confidence than it warrants. Naming the evidence level accurately gives the council a clearer next decision and identifies the measurement work still needed.

## The Pilot-to-P&L Step

MIT CISR's December 2024 briefing compares four maturity stages: Experiment and Prepare; Build Pilots and Capabilities; Develop AI Ways of Working; and Become AI Future Ready. The quantitative base is a 2022 survey of 721 companies, supplemented by 2024 interviews, primarily concerning analytical AI.<sup>8</sup> The largest adjacent-stage financial-performance gap appeared between stages 2 and 3. This comparison is cross-sectional: it neither measures a firm's ROI from moving stages nor isolates architecture as the cause. The management question it raises is whether a firm has the shared capabilities and working practices needed to sustain value across use cases.

## The Capital Allocation Frame

An AI investment competes with other uses of capital, including acquisitions, buybacks, and operating assets. Compare alternatives on their risk, timing, expected cash flows, and strategic purpose. Bain's 2026 midyear report provides market context: global M&A value reached \$2.4

trillion in the first five months, up 41% year over year.<sup>18</sup> That is a measure of the overall deal market, not AI-only transactions or proof of returns to an individual AI project. The council still needs a case for the next dollar invested.

## Cautionary Tales

---

### Klarna and the Effectiveness Gap (May 2025)

In May 2025, Bloomberg reported that Klarna's CEO considered the cost-cutting push in customer service excessive and was planning additional human-service capacity. The company's later annual filing continued to describe its AI assistant.<sup>37</sup> This supports reading the episode as a service-model pivot, not the abandonment of AI. **The lesson to test:** monitor service quality alongside cost and specify what would trigger a change in the balance between automation and human support. The public accounts do not establish the complete causal or financial story.

Several sources suggest that workforce substitution deserves its own downside scenario. Orgvue's April 2025 survey found 39% of respondents' organizations had made AI-related redundancies, and 55% of that subgroup said they had made wrong decisions. Robert Half separately reported 32% of U.S. hiring managers who eliminated a role primarily because of AI had rehired for the same or a similar role. Gartner forecast that by 2027, half of organizations that cut customer-service headcount because of AI would rehire for similar functions.<sup>20</sup> These have different populations, questions, and dates; they cannot be pooled into an observed reversal rate. For a headcount-reduction case, estimate a scenario for reacquiring capability, including recruitment, onboarding, and lost knowledge, using the organization's own evidence.

### Ford and the Role of Experienced Judgment (2026)

Ford said it added 350 experienced technical specialists through internal promotions and external hiring following a 2023 reorganization, to complement its use of AI. That statement does not establish that all 350 were rehired after being replaced by AI. Ford also reported leading mainstream brands in the 2026 J.D. Power Initial Quality Study, its first such result in sixteen years.<sup>21</sup> These observations do not isolate the contribution of AI, training data, experienced engineers, or other process changes to the quality result.

**The lesson to test:** a staffing change should account for the expert judgment a workflow needs. Before booking savings, determine which knowledge has been captured, which decisions still require experienced people, and what evidence will show that quality is holding. Treat loss of institutional knowledge as a risk to investigate, not a proven diagnosis of this case.

### McDonald's, IBM, and the Drive-Thru Pullback (June 2024)

McDonald's announced the end of its IBM voice-ordering pilot in June 2024.<sup>40</sup> **The lesson:** evaluate whether performance, exception handling, service quality, and operating cost justify a production decision. A successful demonstration alone cannot answer those questions. Module 1 carries the broader case discussion.

## SEC Presto Automation Enforcement (January 14, 2025)

The SEC issued a settled cease-and-desist order against Presto Automation over misleading statements about its Presto Voice product and use of AI.<sup>38</sup> Presto accepted the order without admitting or denying the findings. **The lesson:** statements about product availability, automation, and human involvement need support before they enter investor communications. The case provides a specific enforcement example; it does not make every capability claim material in every context.

## DOJ Charges Concerning Nate, Inc. (April 9, 2025)

The U.S. Attorney's Office for the Southern District of New York announced fraud charges against Nate's founder. The indictment alleged that he raised more than \$40 million while misrepresenting the app's AI capabilities, with overseas workers and later some bots performing transactions presented as AI automation.<sup>39</sup> These are allegations in the charging announcement, not a conviction. **The lesson:** substantiate what is automated, what humans do, and what investors are being told about both.

## Uber and Tool-Spending Limits (June 2026)

Bloomberg reported Uber's \$1,500 monthly cap per employee per agentic coding tool.<sup>13</sup> **The lesson:** a spending limit and an outcome measure serve different purposes. The first constrains exposure; the second helps determine whether the work is worth its cost. Pair a budget with cost per accepted task, quality, and evidence of useful delivery. High adoption alone is insufficient to justify the bill.

## Air Canada and the Liable Chatbot (February 2024)

In *Moffatt v. Air Canada* (February 2024), the tribunal awarded compensation after the airline's chatbot supplied misleading fare information, as summarized in legal analysis of the decision.<sup>41</sup> It was a damages award, not an order directing the airline to perform a fare promise. **The lesson:** include the costs of keeping customer-facing information accurate and handling corrections or disputes. The case does not establish that omitted human review inevitably creates an equal amount of legal liability. Module 1 carries the broader discussion.

## Cost of Getting AI Wrong

---

IBM's 2026 *Cost of a Data Breach Report*, released July 29, studied 602 organizations that experienced breaches. The publisher's announcement supplies two useful severity benchmarks:<sup>9</sup>

- **\$4.99 million** — global average breach cost in the study, up 12% year over year.
- **Approximately \$6 million** — average cost of AI-enabled malicious breaches; these represented one-quarter of malicious breaches in the study.

These figures describe the studied breaches. They do not give an annual breach probability for every company, an AI-specific causal premium, or an expected loss for a particular workflow.

Insurance can transfer selected risks. Munich Re's aiSure offers AI-performance coverage, and HSB announced AI liability insurance for small businesses in March 2026.<sup>42</sup> AI exclusions described in

the cited HSB materials concern commercial general liability policies; coverage must be checked against the actual policy and endorsements. These products do not establish universal coverage or exclusions for cyber and errors-and-omissions policies.

For an investment case, estimate scenario probability × financial impact over a stated period, with explicit uncertainty. Then separate the costs of controls, insurance premiums, deductibles, and losses the organization retains. A premium reflects a specified transfer of covered risk, not the price of all residual AI risk. Avoid counting the same loss both as insured exposure and retained loss.

## Capex vs. Opex Reference

---

The capex-vs-opex question is now on board and audit-committee agendas because AI development cost is increasingly material and the accounting guidance has moved.

**FASB ASU 2025-06**, issued September 18, 2025, updates ASC 350-40 internal-use software accounting. It is effective for annual periods beginning after December 15, 2027, with early adoption permitted.<sup>12</sup> Capitalization depends on authorization, probable completion, and resolving significant development uncertainty where present. FASB expects little change for most software, with possible decreases in capitalization for some cloud-delivered software. This is not a categorical rule that AI costs must be expensed. Model the treatment applicable to the entity, project, and adoption date.

**IFRS / IAS 38** provides a different framework for recognizing intangible assets, including qualifying development costs.<sup>22</sup> The reporting entity's applicable accounting framework determines which rules govern; the location of the development team does not. Where an organization prepares reports under more than one framework, Finance should reconcile the treatments and separate accounting presentation from the underlying cash flows.

For the investment council, distinguish cash outlay from the timing of expense recognition and amortization. The accounting treatment may change reported earnings and budget presentation even when the project's underlying cash costs are the same.

## Sectoral Standards

---

Industry-specific frameworks govern AI investment in regulated sectors. A short reference for the regulated members of the cohort:

**Banking — SR 26-2 (Federal Reserve, OCC, FDIC).** The revised interagency guidance, issued April 17, 2026, supersedes SR 11-7 and SR 21-8 and takes an approach tailored to model risk and materiality. Generative and agentic AI are outside this guidance's specific scope. Its footnote 3 nevertheless directs banks to existing risk-management and governance practices when developing appropriate controls.<sup>23</sup> Budget for that work; the exclusion does not remove governance responsibilities.

**Healthcare — FDA AI/ML guidance.** The original joint Good Machine Learning Practice principles date to 2021; IMDRF finalized its GMLP principles in January 2025. Related FDA materials include 2024 transparency principles and predetermined change control plan guidance.

<sup>44</sup> Identify the applicable medical-device pathway and document before costing a clinical AI deployment; these materials have different purposes and should not be treated as one universal approval requirement.

**Insurance — NAIC Model Bulletin on AI (December 2023).** The NAIC's April 1, 2026 reference list identifies 24 states plus D.C. as adopters and lists other insurance-specific state guidance separately.<sup>45</sup> The model bulletin is not uniform federal law. Determine the relevant jurisdiction's adopted requirements for insurer use of AI, and price the program for those obligations. Module 4 supplies the broader sectoral framework.

## Regulatory Cost Reference

---

Regulatory compliance is a cost line on the AI Investment Council's spreadsheet, not a separate workstream. Module 4 owns the obligations; the Current Horizon Scan below carries the current posture. The cost references executives most often ask for:

- **EU AI Act high-risk compliance.** A 2021 study supporting the Commission's impact assessment estimated quality-management-system setup costs of about €193,000–330,000 and annual maintenance of about €71,400 where no QMS already existed.<sup>24</sup> These are proposal-era scenario estimates, not present-day quotations, provider floors, or deployer ceilings. Price the actual role, system, applicable duties, and controls already in place.
- **U.S. state requirements.** Colorado's SB 26-189, signed May 14, 2026 and effective January 1, 2027, replaces the earlier state framework and includes duties concerning notice, explanations, correction, and human review for covered decisions.<sup>31</sup> For Colorado or any other state, identify the exact provision, covered activity, responsible entity, and effective date before assigning a compliance cost.
- **Federal enforcement.** The FTC's September 2024 Operation AI Comply sweep applied existing consumer-protection law to alleged deceptive AI claims and practices.<sup>25</sup> The SEC's Presto order and the DOJ's Nate charges illustrate separate investor-disclosure and criminal-enforcement settings. Identify the actual agency action rather than treating them as a single AI-specific law.
- **Substantiating AI claims.** The cited 2026 legal commentary discusses risks in availability claims, demonstrations, and productivity assertions.<sup>25</sup> It is analysis, not a new statute. Distinguish FTC or SEC enforcement from the National Advertising Division's advertising self-regulation. A user-perception survey may support a perception claim while failing to establish an objective time-saving or financial claim. The cost line is the evidence needed for the claim actually being made: contemporaneous tests, definitions, limitations, and documentation that can travel with a pitch or disclosure.
- **Sectoral overlays.** Banks, insurers, and clinical deployers may need additional controls; use the Sectoral Standards reference above to identify the applicable framework.

The amount depends on scope and existing controls. Estimate it before approving scale.

## Model Deprecation Reference

---

Notice periods vary by product and provider. OpenAI's January 29, 2026 announcement retired certain models from ChatGPT on February 13, a 15-day interval, while explicitly leaving API availability unchanged at that announcement. Anthropic's published policy provides at least 60 days' notice for retirement of publicly released API models.<sup>43</sup> Neither supports a universal 90–180-day window. Migration planning should include evaluations on the replacement, prompt and contract review, and retrieval reindexing if the embedding representation changes. Use actual vendor terms and dependency exposure. A scenario with one major migration in 18 months is a planning assumption to stress-test, not an observed industry rule; replacement prices can rise or fall.

## Current Horizon Scan (As of September 14, 2026)

---

This section holds dated market forecasts, survey findings, company reports, and accounting or regulatory developments. Each entry identifies the evidence type. **Unresolved** names what the evidence does not settle; **Strategic implication** is the module's interpretation for an investment decision. A published forecast is evidence of an expectation, and a company disclosure is evidence of what management reported. Neither is automatically a realized or independently validated result. Recheck time-sensitive entries before using them for an investment or compliance decision.

### Market and Spending Signals

- **Enterprise AI spending forecasts. Forecast:** Gartner estimated worldwide 2026 AI spending at \$2.52 trillion in January, up 44%, then revised the estimate to \$2.59 trillion and 47% in May. A separate July forecast put spending on AI platforms and models at \$64 billion, up about 63%.<sup>26</sup> These are different market scopes, not interchangeable measures of enterprise AI operating budgets. **Unresolved:** Full-year actuals and the allocation of costs relevant to an individual buyer. **Strategic implication:** Large market spending does not establish the economics of a particular workflow.
- **The enterprise EBIT gap. Survey evidence:** McKinsey's August 25, 2026 report surveyed 1,719 respondents. It reports 44% scaling enterprise AI versus 38% previously, 37% reporting enterprise EBIT impact, and approximately 6% meeting its high-performer definition. Eighty percent reported improved personal productivity; one in five cited operating-cost constraints; 32% had declined a software purchase after using agentic coding tools. Among organizations above \$1 billion revenue, 40% were scaling agents versus 27% previously. Nearly three-quarters of high performers reported workflow redesign, versus about one-quarter of others.<sup>1</sup> **Unresolved:** Causal attribution and independently measured returns. **Strategic implication:** Test the adoption-to-value chain; the differences do not prove which practice caused the financial result.
- **Infrastructure spending and power demand. Analyst estimate and forecast:** FactSet's July 2026 analysis describes more than \$690 billion in hyperscaler FY2026 capex and rising use of external finance.<sup>19</sup> Its fiscal-period figure should not be combined with calendar-year estimates that count leases or customer prepayments differently. IEA's base case projects global data-center electricity use at 945 TWh in 2030, more than double the 2024 level; this covers all

data centers, with AI a major driver.<sup>27</sup> **Unresolved:** Realized demand, financing, local capacity, and supplier pricing. **Strategic implication:** Stress-test procurement costs and availability. Power demand is one input to supplier economics, not a formula for future token prices.

- **Deal-market context. Market report:** Bain's June 29, 2026 midyear report put global M&A value at \$2.4 trillion for the first five months, up 41%, with the count of deals above \$10 billion up 52%.<sup>18</sup> These are overall M&A figures. **Strategic implication:** Evaluate an AI investment alongside other capital uses on comparable risk and return assumptions.
- **Outcome-based pricing. Executive and analyst interviews:** CIO Dive's August 31 report describes Zendesk charging for fully automated resolutions and Pegasystems charging per completed case while managing its model costs. These arrangements coexist with seat pricing.<sup>28</sup> Gartner's analyst reported use by 19% of services buyers and in 13% of seller-side agreements; he forecast fewer than 25% of tech CEO services contracts using outcome pricing through 2031. HP's executive described options expected for early adopters in mid-to-late 2027. **Unresolved:** Adoption, contract definitions, and the allocation of delivery risk. **Strategic implication:** Specify what counts as a completed outcome, how partial completion is treated, and which risk the vendor actually assumes.
- **Where survey respondents see value. Vendor-sponsored survey:** SAP and Oxford Economics surveyed 2,600 business leaders across 13 countries. Sixty-nine percent were satisfied with current AI ROI, while reported benefits emphasized insight and customer interaction more than some originally expected savings. Twelve percent described the specified skills or processes/frameworks as fully ready for governance; 38% lacked human-in-the-loop processes for agent workflows, 37% lacked agent permission/access controls, and 44% had an agent registry.<sup>29</sup> **Unresolved:** Objective financial returns and causal explanations for the reported gaps. **Strategic implication:** Investigate benefits that differ from the original hypothesis, but validate their value before relabeling a missed savings target as a success.

## Accounting, Cost, and Risk Signals

- **FASB ASU 2025-06. Issued accounting standard:** The September 18, 2025 update is effective for annual periods beginning after December 15, 2027, with early adoption permitted. It changes capitalization criteria, including treatment of significant development uncertainty.<sup>12</sup> **Unresolved:** The treatment of a particular project under the entity's adoption policy and facts. **Strategic implication:** Finance should reconcile the cash-flow case, capitalization policy, and earnings presentation before the investment reaches the audit committee.
- **IBM breach-cost study. Study evidence:** IBM's July 29, 2026 announcement reports a \$4.99 million average breach cost in its 602-organization sample and roughly \$6 million for AI-enabled malicious breaches.<sup>9</sup> **Unresolved:** The frequency and severity distribution relevant to the buyer's own workflows. **Strategic implication:** Use the study as severity context and build a separate, explicit risk scenario; averages from breached organizations are not expected annual losses for all organizations.
- **AI insurance. Documented products:** Munich Re's AI-performance coverage and HSB's March 2026 small-business AI liability launch illustrate options for transferring specified risks.<sup>42</sup> General-liability AI exclusions make actual policy wording important. **Unresolved:** Availability, underwriting, limits, exclusions, and price for a particular buyer. **Strategic**

**implication:** Compare covered and retained risk; an insurance premium does not price every AI exposure.

- **FinOps practice. Practitioner survey and specification:** The 2026 State of FinOps survey had 1,192 respondents and reports 98% managing AI spend versus 31% two years earlier.

FOCUS 1.3 added the Contract Commitment dataset in December 2025.<sup>4</sup> **Strategic**

**implication:** Use the four artifacts and named role above to turn spending data into decisions. The survey population should not be generalized to all enterprises.

## Regulatory Signals

- **EU AI Act application dates. Enacted law:** Regulation (EU) 2026/1744 was published July 24, 2026 and entered into force July 27. It changes the relevant high-risk application dates to December 2, 2027 for Annex III and August 2, 2028 for Annex I product-related systems. Article 50 applies from August 2, 2026, while GPAI provisions have their own application and transition rules.<sup>30</sup> **Unresolved for a particular deployment:** The system's classification, provider/deployer role, applicable exceptions, and implementation requirements. **Strategic implication:** Map the actual obligation. Article 50 includes specified interaction-notice and synthetic-content duties; it is not a single identical disclosure rule for every generative feature.
- **Colorado SB 26-189. Enacted law:** Signed May 14, 2026 and effective January 1, 2027, the act replaces the earlier Colorado framework for covered automated decision-making. Its duties include notice, adverse-outcome explanations, correction, and human review.<sup>31</sup> **Unresolved for a deployment:** Coverage, implementation details, and any subsequent authoritative change. **Strategic implication:** Cost the enacted duties and check official updates; do not rely on an unsourced promise about future enforcement.
- **U.S. banking model-risk guidance. Issued guidance:** The April 17, 2026 interagency revision supersedes SR 11-7 and SR 21-8. It excludes generative and agentic models from its specific scope but directs banks to existing risk-management and governance practices for appropriate controls.<sup>23</sup> **Unresolved:** How those practices should be implemented for a particular use case and whether further guidance changes them. **Strategic implication:** Include the required control work in TCO; an exclusion from this document is not a governance exemption.

## Frameworks Glossary

---

**NIST AI Risk Management Framework (AI RMF 1.0, January 2023) and Generative AI Profile (NIST AI 600-1, July 2024).** Voluntary framework organized around GOVERN, MAP, MEASURE, and MANAGE. The GenAI profile identifies twelve risk categories. Use it to organize risk work and evidence.<sup>15</sup>

**ISO/IEC 42001:2023.** A certifiable AI management-system standard, complementary to security and quality management standards. Certification concerns the management system within its defined scope; it does not certify every model output or financial claim.<sup>16</sup>

**ISO/IEC 5338:2023 (AI System Lifecycle Processes).** Defines lifecycle processes for AI systems. Use its actual requirements and scope when specifying engineering and management work; it is not a substitute for a project-specific control plan.<sup>49</sup>

**OECD AI Principles (2019, updated May 2024).** International principles including accountability, transparency, and robustness. They provide policy guidance; binding duties come from applicable law and regulation.<sup>49</sup>

**FinOps Foundation FOCUS Specification and State of FinOps 2026.** FOCUS standardizes technology cost and usage data. Version 1.2 was ratified in May 2025 and announced in June, including virtual-currency reporting; version 1.3 was ratified in December 2025 and added the Contract Commitment dataset. Version 1.4 was available by this review date. These version references identify features, not universal provider implementation.<sup>4</sup>

**FASB ASU 2025-06 (September 2025).** Updates internal-use software capitalization guidance, including development-uncertainty criteria. Treatment depends on the facts; it is not a blanket AI-expensing rule.<sup>12</sup>

**NACD 4-Pillar Framework for AI Oversight (2025).** AI strategy, capital allocation, AI risks, and technology competency. A board-level framework that complements the management-layer AI Investment Council.<sup>17</sup>

**MIT CISR Enterprise AI Maturity Model (December 2024).** Four stages associated with different financial performance in the study. The largest adjacent-stage gap was between stages 2 and 3; it is not a demonstrated causal ROI effect.<sup>8</sup>

## Workbook Prompt

---

Run this as a 45-minute exercise with three to five colleagues from across your organization (operations, technology, finance, risk, and the relevant business unit). Steps 1, 2, and 5 are the Foundation/Core pass and can be run in twenty-five minutes; steps 3 and 4 complete the Challenge/Full Case. The two passes differ in depth, not grading status.

1. **Pick one in-flight AI initiative.** Write a one-sentence description that names the business problem, not the technology.
2. **Build the KPI Value Chain on the spot.** Technical metric → adoption metric → workflow KPI → financial KPI → business impact. Mark the link you cannot evidence today.
3. **Estimate unit economics.** Cost per successful task with the full harness loaded — model calls, context, retrieval, evals, monitoring, human review, retries. Estimate the value per successful outcome. If the math does not survive a CFO's first question, mark which assumption breaks it.
4. **Identify the weakest link in the chain.** Is the gap technical, adoption, workflow, financial, or strategic? Each requires a different intervention.
5. **Write a three-sentence stop/pivot/scale recommendation.** What would the next investment review decide, and what evidence would change it?

Repeat the exercise for two more initiatives. The pattern across three cases often surfaces more about your organization's AI investment discipline than any single case examined alone.

## Prompt Examples

---

A small executive prompt library for use in business-case stress tests, vendor evaluations, and council reviews. Adapt to your context.

### **Stress-test a value hypothesis.**

"I am evaluating an AI initiative with the following hypothesis: [paste]. Identify three plausible alternative explanations for the observed effect, and three external factors that could erode the value before scale. For each, name the evidence I would need to rule it out."

### **Surface harness costs that pilots typically miss.**

"I am building a TCO for an AI initiative in [function]. The pilot priced model access and inference. List the harness costs I should add — context and caching, retrieval, evals, monitoring, human review, exception handling, governance, change and training, vendor commitments, and ongoing operations. For each, give me a question I should ask the team to estimate it."

### **Probe a vendor on unit economics.**

"I am meeting with a vendor pitching an agentic AI product. Draft five questions that would force the vendor to disclose cost per successful task — including how they define 'successful,' what falls outside that definition, how cost scales with volume, what happens if the underlying model is deprecated, and what their typical customer's harness cost looks like in production."

### **Draft a stop/pivot/scale recommendation.**

"Here is the evidence we have on an AI initiative: [paste]. Draft three short paragraphs — one arguing to stop, one to pivot, one to scale — with the strongest case for each. Identify the single piece of evidence that would tip the recommendation from one path to another."

### **Translate cost-per-call into cost-per-successful-task.**

"A vendor is quoting \$X per million tokens. Help me translate that into cost per successful task for a use case where: average input is [N] tokens, average output is [M] tokens, [P]% of attempts require a retry, [Q]% require human review, and the human review cost is \$[R] per minute. Show the math."

## Use-Case Gallery

---

Extended named cases that illustrate range and depth without crowding the Pre-Class Read. Each carries the relevant ROI lens and any disclosure caveat.

## JPMorgan Chase LLM Suite

JPMorgan described LLM Suite access for about 200,000 employees in a June 2025 interview, and described adding capabilities every few weeks in an October interview. Separately, at a September 2024 conference, the bank reported broad AI/ML value of \$1–1.5 billion in 2023 and said 2024 was heading toward approximately \$2 billion.<sup>35</sup> These are different disclosures: the value estimate is not an isolated LLM Suite return. **Lens:** efficiency, effectiveness, and risk reduction. **Caveat:** company-reported access and value do not establish adoption intensity, net ROI, or a particular investment-council cadence.

## Walmart

Walmart reports that route optimization eliminated 30 million unnecessary miles and that an AI-assisted fashion process can shorten development by as much as 18 weeks. Its customer-shopping assistant Sparky and business-partner assistant Marty serve different users. Pactum provides a separate supplier-negotiation system.<sup>36</sup> **Lens:** efficiency, effectiveness, and innovation. **Caveat:** these are separate company or vendor accounts with different metrics and periods; they do not constitute a single measured portfolio return.

## Vanguard

In Davenport and Bean's October 2025 interview, Vanguard leaders described approximately \$500 million of value across data, analytics, and AI. They distinguished roughly 25% productivity improvement in coding from 10–15% across the full software development life cycle.<sup>11</sup> **Lens:** effectiveness and efficiency. **Caveat:** these are company-reported estimates spanning multiple activities, not an independently audited GenAI ROI. Ask how each benefit is defined, attributed, costed, and reconciled with financial results.

## Salesforce Agentforce

Salesforce's December 3, 2025 Q3 FY2026 release reports nearly \$1.4 billion in combined Agentforce and Data 360 ARR, up 114% year over year, and paid Agentforce deals up 50% quarter over quarter. A separate July 2025 company account reports 85% resolution on its own help site.<sup>47</sup> **Lens:** vendor growth and a self-service operating example. **Caveat:** ARR is vendor revenue, not customer ROI; the resolution figure concerns Salesforce's own site and a different reporting period. Neither establishes the unit economics of a buyer's deployment.

## Sysco and Chewy: Two Answers to What AI Savings Mean (September 2026)

On September 9, 2026, Sysco and Chewy put AI savings expectations on the record and emphasized different aspects of how the benefits would be managed.

Sysco introduced a program targeting at least \$500 million in cost reduction by FY2029, including \$100 million of net cost savings in FY2027, and added structural cost-out to its long-term equity performance program. Its release specifies standalone Sysco guidance. An August announcement describes monthly board oversight through the renamed Artificial Intelligence Transformation & Technology Committee.<sup>32</sup> The combination provides concrete mechanisms to examine: a target, a compensation measure, and a review forum. It does not establish that the target has been achieved or that those mechanisms caused a financial result.

Chewy's CEO described expected AI savings in the low tens of millions in FY2026, rising to about \$50 million annualized in FY2027. The call reported approximately 30% self-service chat resolution among the selected mobile-app customers using its new assistant, plus tools in customer care, pharmacy, and selected veterinary sites. Management cautioned against adding the savings mechanically to margins because efficiencies would also offset costs and support reinvestment.<sup>33</sup>

**Lens:** efficiency and benefits realization. **Caveat:** the full savings targets are forward-looking company guidance. Management describes early operating benefits, but the targets are not independently validated returns. Sysco identifies standalone-company scope; a reviewer should still request the detailed net-cost bridge. Chewy describes a contribution to broader productivity rather than a separate increment to profit.

**What to take from the pair.** Sysco illustrates tying a cost target to compensation and oversight. Chewy illustrates defining how expected savings enter the earnings model. Both are mechanisms a council can examine, not proof of a universally strongest approach. Agree on the benefit definition and accountability before the number enters an investment commitment.

### Chick-fil-A

Chick-fil-A's Davis Addy described machine learning for foodborne-illness surveillance using social-media signals at a 2019 conference.<sup>48</sup> The case offers a historical risk-avoidance illustration: the proposed value is earlier investigation of a signal, while cost includes running the system and reviewing false positives. A business case would need evidence that alerts change detection or response outcomes and a defensible estimate of avoided loss. **Lens:** efficiency through risk mitigation. **Caveat:** the talk does not establish current performance, a realized financial return, or the company's cost of capital.

### PDI Technologies

PDI Technologies, a private-equity-backed technology provider serving convenience retail and fuel businesses, announced PDIQ in March 2026. Its launch described 600 agents in internal operations, more than 400 AI-certified engineers, and access for selected customer beta programs.<sup>46</sup> **Lens:** efficiency and effectiveness. **Caveat:** internal agent counts and customer beta access are different deployment measures. Embedding AI may spread some fixed costs across transactions, while variable usage and support costs remain; the announcement does not establish realized ROI or PDI's investment hurdle rate.

### Microsoft 365 Copilot (Forrester TEI)

Forrester's commissioned enterprise Total Economic Impact study models 116% three-year ROI and \$19.7 million NPV for a composite organization. The separate SMB study projects up to 353% three-year ROI.<sup>10</sup> **Lens:** efficiency. **Caveat:** composite and projected results are conditional on the study's assumptions, not realized returns for a representative buyer. Replace its license, adoption, time-saving, implementation, and benefit-capture assumptions with your own. Time saved is an outcome measure; adoption is the proportion of eligible users or work actually using the tool.

## Part 3: Card Specification

### The AI Business Case

The AI Business Case card is drafted in the Investment Committee simulation for a student-selected use case. A strong card is written as an executive investment brief, not as a class reflection; there is no post-class submission.

| Depth                  | Section                         | Expected Content                                                                                                                                           |
|------------------------|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Foundation/<br>Core    | Use case                        | Name the AI initiative and the business process it affects.                                                                                                |
| Foundation/<br>Core    | Portfolio category              | Classify it as efficiency, effectiveness, innovation, or a combination.                                                                                    |
| Foundation/<br>Core    | Strategic rationale             | Explain why this use case matters now.                                                                                                                     |
| Foundation/<br>Core    | Baseline                        | Define current performance: cost, cycle time, quality, revenue, risk, capacity, or customer outcome.                                                       |
| Foundation/<br>Core    | Value hypothesis                | State the expected business change in one sentence.                                                                                                        |
| Foundation/<br>Core    | KPI Value Chain                 | Connect technical metric, adoption metric, workflow KPI, financial KPI, and business impact; mark the weakest link.                                        |
| Foundation/<br>Core    | Stop/pivot/scale criteria       | Define what evidence would trigger each decision.                                                                                                          |
| Challenge/Full<br>Case | Full TCO and FinOps assumptions | Identify the major cost drivers, including model usage, context length, caching, integration, evals, monitoring, human review, training, and governance.   |
| Challenge/Full<br>Case | Cost per successful task        | Define the unit of success and estimate the full harness cost per accepted, useful, or completed outcome, as a range.                                      |
| Challenge/Full<br>Case | Key sensitivities               | Name the two assumptions (adoption, accuracy, oversight cost, volume) the case is most exposed to, each expressed as a range rather than a point estimate. |
| Challenge/Full<br>Case | Benefits ownership              | Name the business owner, Finance partner, and measurement cadence.                                                                                         |
| Challenge/Full<br>Case | Risk constraints                | Name the production constraint most likely to block scale and whether it is priced as a TCO line.                                                          |
| Challenge/Full<br>Case | Next 90 days                    | Identify the next three actions needed to improve evidence quality and decision readiness.                                                                 |

Example value hypothesis:

If claims adjusters use AI-assisted summarization and evidence retrieval in 70% of eligible claims, average handling time will fall by 15% while quality scores remain stable.

**Capstone build (Transform path):** the card becomes the value-logic section of the final memo. It should give a CFO and business owner enough evidence to debate whether the use case should be funded, narrowed, stopped, or scaled.

**Next-module handoff:** Module 6 will put the economic case under board-level pressure by asking whether the use case still makes sense when a competitor, customer agent, infrastructure constraint, or frontier risk changes the strategic context.

---

## Works Cited and Sources

---

Sources below identify the evidence used in this September 14, 2026 revision. Survey findings, executive reports, forecasts, studies, and legal instruments have different evidentiary roles. A citation does not turn an estimate or allegation into a realized result.

1. McKinsey & Company, The State of AI in 2026: On the Road to ROI, August 25, 2026. Online survey of 1,719 respondents in 97 countries, May 4–June 8, 2026; comparisons refer to the prior survey edition. High performers report both significant value and at least 5% of EBIT attributable to AI.
2. PwC, 2026 Global CEO Survey. CEO self-reports of financial benefits; “both cost and revenue benefits” is a narrower measure than any benefit.
3. Stanford HAI, The 2026 AI Index Report. Optional background.
4. FinOps Foundation, State of FinOps 2026 (1,192 practitioner respondents); FinOps for AI; FOCUS 1.2 announcement, June 2025; FOCUS 1.3 announcement, December 11, 2025, following December 5 ratification; FOCUS specification overview, including version 1.4. Adoption statistics concern the survey's practitioners; specification features do not establish provider implementation.
5. Gartner, Why Half of GenAI Projects Fail: Avoid These 5 Mistakes. Used as a practitioner framework for failure risks, not a universal measured failure rate.
6. Deloitte AI Institute, State of Generative AI in the Enterprise, fourth-quarter report release, January 21, 2025. ROI question concerns each respondent's most advanced initiative; scaling question concerns the share of experiments expected to scale within three to six months.
7. BCG, AI Radar 2026: As AI Investments Surge, CEOs Take the Lead. Optional background.
8. Peter Weill, Stephanie L. Woerner, and Ina M. Sebastian, Building Enterprise AI Maturity, MIT CISR, December 2024. Cross-sectional financial comparisons; the briefing describes the 2022 quantitative survey and subsequent interviews.
9. IBM, IBM Study: One in Four Malicious Breaches Are AI-Enabled, Costing Companies \$6 Million on Average, July 29, 2026; Cost of a Data Breach report page. The cited headline findings concern 602 breached organizations studied during March 2025–February 2026. They are severity benchmarks, not an annual incident probability for all organizations.
10. Forrester Research, The Total Economic Impact of Microsoft 365 Copilot, commissioned enterprise composite study, 2025; Microsoft, Summary of the projected SMB Total Economic Impact study, October 17, 2024. Modeled and projected returns depend on study assumptions; neither is a representative customer's realized ROI.
11. Thomas H. Davenport and Randy Bean, Investing in AI Payoffs at Vanguard, MIT Sloan Management Review, October 27, 2025. Interview-based company estimates spanning data, analytics, and AI; coding and full-development-cycle productivity are distinct measures.
12. FASB, ASU 2025-06: Targeted Improvements to the Accounting for Internal-Use Software, issued September 18, 2025. Consult the operative amendments, effective date, and basis for conclusions, including BC4; the standard is not a blanket rule for AI expensing.
13. Bloomberg, June 2, 2026, report carried by the Los Angeles Times. Spokesperson-attributed account.

14. Longju Bai et al., How Do AI Agents Spend Your Money? Analyzing and Predicting Token Consumption in Agentic Coding Tasks, arXiv:2604.22750, April 2026. Preprint; coding-benchmark results with eight models, not universal enterprise-agent cost multipliers.
15. NIST, AI Risk Management Framework, January 2023; Generative AI Profile, NIST AI 600-1, July 2024. Voluntary framework and companion profile.
16. ISO, ISO/IEC 42001 explained. AI management-system standard; distinct from NIST framework use and from validation of an individual AI output.
17. NACD, Director Essentials: Implementing AI Governance, September 2025. The four sections are AI Strategy, Capital Allocation, AI Risks, and Technology Competency. Survey sample sizes are question-specific; see the relevant chart caption.
18. Bain & Company, 2026 M&A Midyear Report release, June 29, 2026. Overall global M&A market, first five months of 2026.
19. FactSet Insight, Hyperscalers Tap External Financing as AI Capex Outruns Cash Flow, July 23, 2026. Analyst aggregation; distinguish fiscal-period capex from calendar-year estimates including leases or customer prepayments.
20. Orgvue, 55% of businesses admit wrong decisions in making employees redundant when bringing AI into the workforce, April 29, 2025; Robert Half, Is the UK set for a wave of AI correction recruitment?, reporting the separate U.S. hiring-manager result; Gartner, Prediction concerning customer-service rehiring by 2027, February 2026. Different populations and measures; the Gartner figure is a forecast.
21. Ford statement appended to Motor1's reporting, July 2026, concerning 350 experienced technical specialists added through internal promotion and hiring; Ford, 2026 J.D. Power Initial Quality Study result. Neither source isolates the causal contribution of AI or staffing changes to the quality result.
22. IFRS Foundation, IAS 38 Intangible Assets. Apply the reporting entity's governing framework and the relevant recognition criteria.
23. Federal Reserve, SR 26-2: Revised Guidance on Model Risk Management, April 17, 2026; attached interagency guidance, especially footnote 3 on generative and agentic AI and existing governance practices.
24. CEPS, Clarifying the costs for the EU's AI Act, September 24, 2021, explaining the CEPS/ICF/Wavestone study supporting the Commission's impact assessment. Historical proposal-era QMS cost scenarios, not universal bounds or current quotations.
25. FTC, Operation AI Comply announcement, September 25, 2024; Madison Fleischaker, Operation AI Comply: Continued Enforcement Against Misleading Claims, Holland & Knight, August 18, 2026. The second item is legal commentary; NAD self-regulation is distinct from government enforcement.
26. Gartner, January 15, 2026 worldwide AI spending forecast; May 19, 2026 revised worldwide forecast; July 20, 2026 AI platforms and models forecast. Distinct dated forecasts and market scopes.
27. IEA, Energy and AI — Executive Summary, 2025. Base-case electricity projection for all data centers; AI is a major driver, not the entire category.
28. Patrick Thibodeau, Agentic AI is shifting the pricing models CIOs rely on, CIO Dive, August 31, 2026. Original executive and analyst interviews; commercial practices and forecasts retain their attribution.
29. SAP and Oxford Economics, Value of AI Report 2026 release, July 15, 2026. Vendor-sponsored survey of 2,600 business leaders across 13 countries. Satisfaction, readiness, and controls are respondent reports; projected future ROI is not used as a realized result.
30. European Parliament and Council, Regulation (EU) 2026/1744, Official Journal, July 24, 2026; AI Act, Regulation (EU) 2024/1689, read with the amendments; European Commission, Navigating the AI Act. Apply the relevant provisions, exceptions, and transition rules to the actual system and role.
31. Colorado General Assembly, SB 26-189, signed May 14, 2026; effective January 1, 2027. The enacted text and subsequent official actions govern implementation.
32. Sysco, September 9, 2026 AI efficiency-program and guidance release; August 20, 2026 board and AI-transformation governance announcement. Company targets, compensation design, and oversight; the September guidance explicitly concerns standalone Sysco.

33. Chewy, Q2 FY2026 earnings release and published earnings-call transcript, September 9, 2026. Original executive statements in a third-party transcript; savings targets are forward-looking and early chat results concern selected customers.
34. Shelia Anderson, Continuous Learning With AI: Aflac's Shelia Anderson — podcast transcript, 2023; Keith Farley, Human First: How Aflac Combines AI With Authentic Connections, CMSWire, September 18, **2024**. Farley's title at the interview was SVP of Individual Voluntary Benefits. Claims-automation figures are executive self-reports.
35. JPMorgan Chase, Barclays conference transcript, September 2024, discussing 2023 value and the 2024 outlook; McKinsey interviews with Mark Birkhead, June 5, 2025, and Derek Waldron, October 2025. Access, release activity, and broad AI/ML value are different disclosures.
36. Walmart, AI-powered logistics announcement, March 14, 2024; Agentic-strategy announcement, May 29, 2025; Walmart Connect, AI-powered retail media and Marty; Pactum, Walmart customer account. Company and vendor accounts; supplier negotiation is distinct from Marty's partner-assistant role.
37. Bloomberg News, Klarna Slows AI-Driven Job Cuts With Call for Real People, May 8, 2025; Klarna, 2025 annual filing. A change in human-service capacity does not imply the AI assistant was shut down.
38. SEC, Presto Automation administrative proceeding, Release 33-11352, January 14, 2025. Settled cease-and-desist order, without admission or denial of the findings.
39. U.S. Attorney's Office, Southern District of New York, Nate founder fraud-charge announcement, April 9, 2025. The announcement alleges more than \$40 million raised and describes manual workers and later bots; it expressly identifies allegations, not a conviction.
40. CBS News, June 17, 2024 report, with McDonald's statement.
41. Legal analysis of Moffatt v. Air Canada, February 2024. Secondary legal account of the tribunal's damages award; primary decision not retrieved in this review.
42. Munich Re, AI-performance insurance; HSB, AI liability insurance announcement, March 18, 2026, and product explanation. Coverage and exclusions depend on the actual policy.
43. OpenAI, Retiring GPT-4o and other older models, January 29, 2026, concerning February 13 ChatGPT changes and explicitly distinguishing API availability; Anthropic, Model deprecations. Product-specific retirement events and policy, not a uniform industry timetable.
44. FDA, Good Machine Learning Practice guiding principles and AI software medical-device overview. Distinguish 2021 joint principles, 2024 related guidance, and January 2025 IMDRF principles.
45. NAIC, AI Model Bulletin implementation map and state citations, status April 1, 2026. The adoption list and the separate category of other state regulation/guidance are not interchangeable.
46. PDI Technologies, PDIQ launch announcement, March 26, 2026; Genstar/TA investment announcement. Launch metrics concern internal agents, engineering certification, and selected customer beta programs; no majority ownership share or investment hurdle is asserted.
47. Salesforce, Q3 FY2026 earnings release, December 3, 2025; company help-site account, July 2025. Different periods, scopes, and metrics; vendor growth is distinct from customer ROI.
48. Restaurant Dive, Chick-fil-A taps AI to identify food-safety issues, May 28, 2019, reporting Davis Addy's conference presentation. Historical operating illustration.
49. ISO, ISO/IEC 5338:2023; OECD, AI Principles. Standards and policy principles have different status from binding jurisdiction-specific requirements.

## Copyright, reuse, and citation

© 2026 Gerald C. Kane. This module is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), <https://creativecommons.org/licenses/by-nc/4.0/>.

You may share and adapt it for noncommercial purposes, provided you credit Gerald C. Kane as the author, link to the license, and indicate any changes you made. You may not present this material, or adaptations of it, as your own work, and you may not sell it or use it in paid training or consulting without written permission.

**Cite as:** Kane, G. C. (2026). *Module 5: AI Economics and Scaling Discipline*. EMBA-7450, Information Technology and Strategy, Executive MBA, Terry College of Business, University of Georgia. <https://www.profkane.com/course>  
Permissions and commercial use: [jerry@profkane.com](mailto:jerry@profkane.com)