

A Science and Its Discipline

OPENING CASE

The discipline that changed its mission

DATED: FEBRUARY 2026. FIGURES AND THE QUOTATION CHECKED AGAINST THE FOUNDATION'S REPORT AND ANNOUNCEMENT, OCTOBER 2026.

In February 2026, the FinOps Foundation changed its mission statement. For years, it had described its purpose as “advancing the people who manage the value of cloud.” The new statement replaced one word: “advancing the people who manage the value of technology.”¹

The change accompanied the Foundation’s annual survey of its practitioners, the people inside companies who manage cloud spending for a living. The 2026 survey reached 1,192 respondents responsible for more than \$83 billion in annual cloud spending.² Its central finding concerned the expansion of FinOps into AI spending. Ninety-eight percent of respondents now managed AI spending, up from 31 percent two years earlier.³

FinOps was built to address a specific problem: cloud consumption varied with use, and the bills arrived after capacity had been used.⁴ The discipline developed to make engineering, finance, and business teams

¹. FinOps Foundation, “A One Word Change: How the FinOps Community Made Our Mission Evolution Inevitable,” February 19, 2026, accessed October 13, 2026. The mission statement before and after the change.

². FinOps Foundation, “State of FinOps 2026,” February 19, 2026, accessed October 13, 2026. Survey size and the cloud spending its respondents manage.

³. FinOps Foundation, “State of FinOps 2026,” February 19, 2026, accessed October 13, 2026. Share of respondents managing AI spend, 2026 survey against 2024 survey.

⁴. FinOps Foundation, “2024 Changes to the Definition of Cloud FinOps,” accessed October 13, 2026. Cloud consumption as variable and tracked after the fact.

jointly accountable for that spending.⁵ Its practitioners later took on AI spending, and by 2026 nearly all of the survey's respondents were managing it.

Asked whether AI was providing value, one practitioner stated what remained unknown. The report chose to quote the answer: "No one can answer that question yet."⁶

The quotation names the gap Chapters 1 and 2 described, in the words of someone asked to close it. When an organization pays for a metered resource, costs accrue with use, but someone must build the means to measure the value it receives. FinOps is a capable neighboring discipline that has expanded into work it was not built to govern: the economics of an organization's AI consumption.

That work requires both a body of knowledge about what is true and a practice that turns that knowledge into action. A neighboring discipline cannot supply either on its behalf. Chapter 14 explains how FinOps and that practice divide the work. Until both exist, the practitioner's question about whether AI is providing value remains unanswered.

3.1 Why name a science

Managers act on claims about how their business works. A budget assumes that costs follow usage. A pricing negotiation assumes that the provider has a reason to accept a lower price. A dashboard assumes that the information it displays will change someone's decisions. When those claims are wrong, careful execution does not repair the decisions built on them.

Claims like these can remain informal, held in experience, rules of thumb, or the slides of the last person to present. That knowledge is useful, but it has a cost: two managers can hold opposing beliefs about

⁵. FinOps Foundation, "What is FinOps?," accessed October 13, 2026. The Foundation's definition.

⁶. FinOps Foundation, "State of FinOps 2026," February 19, 2026, accessed October 13, 2026. Practitioner quotation as reproduced in the report.

the same deployment without being able to identify where they disagree. One might treat Chapter 1's Cursor licenses as a fixed software expense and the other as metered consumption. Without stating their claims, neither can identify the fact that would settle the disagreement, so the argument turns on confidence rather than evidence.

Chapters 1 and 2 each ended with an explicit result rather than an opinion. Chapter 1 stated the conditions under which deployed AI is a resource the organization consumes. Chapter 2 stated the conditions under which a scaled deployment cannot be controlled without a governing apparatus. Each result appeared in a panel with an identifier that points to the larger body of claims supporting it.

AI Business Economics is that body of knowledge. It is a science in a specific, limited sense: a set of claims about the business economics of consuming AI, each stating when it holds and expressed precisely enough to be checked. Here, science does not imply laboratory experiments or certainty. It means that every claim states when it applies, what it concludes, and which other claims it rests on.

The claims are kept in a registry, a structured record of every definition, assumption, and result in the science. Each result is linked to the claims that support it. A definition fixes what a term means, and an axiom states an assumption the science accepts without proof. A proposition states a single claim, accepted on the basis of the definitions, axioms, or earlier claims it names as support. A lemma combines propositions, and sometimes other lemmas, into an intermediate result. A theorem combines lemmas into a result with direct consequences for management.

DEFINITION

AI Business Economics

The science of the business economics of consuming AI: an ordered body of conditional claims, each stating when it applies, what it concludes, and which other claims it depends on.

These links are the registry's most important feature. Every higher claim names the lower claims it depends on, which in turn name their own supporting claims. The result is a dependency graph: a map in which no claim can rest, directly or indirectly, on itself. A reader with the registry can start at any certified theorem and follow its supporting claims down to definitions and axioms. Appendix A reproduces every certified theorem and lemma, while propositions are cited by identifier. Readers can therefore trace any certified theorem to the lemma level using the book alone. Below that level, they work from the propositions supplied in a chapter or exercise.

For a manager, the graph resolves one practical concern: no conclusion rests on a circular chain of recorded support. A manager can therefore trace a doubt down to a smaller claim and examine it. Whether that smaller claim is true is a separate question the graph does not answer.

The registry also records which claims are complete. A claim is certified when every claim it depends on is certified and its supporting argument meets the registry's fixed conditions. The registry's validator computes that status at each version; no one enters the status by hand. The validator checks that every dependency is recorded and certified, that no chain of support loops back on itself, that every empirical claim has evidence attached, and that a reviewer has signed off on the claim's current wording.⁷ It checks the structure and record of an argument, not its truth. A claim with a gap in its supporting claims carries a record of what is missing. This book renders and cites only certified claims, because presenting an uncertified claim as a theorem would assert more than the science establishes.

The registry provides the book's justification, while the manager's learning needs determine its organization. Chapters follow the order in which a manager needs to learn, which differs from the order in which the science derives its results. The registry supports each formal result the

⁷ AI Business Economics registry, "dag.aiom, commit 9d7ee50," August 21, 2026. What the validator checks.

book presents, so that a reader who doubts one can trace its basis. Organizing the chapters by proof would leave a manager reading definitions for many pages before reaching a recognizable decision.

3.2 Tracing a result the reader already accepts

Once a deployment reaches scale, economic control requires a governing apparatus. Diligence from the people running the deployment is no longer enough. Chapter 2 ended with that result, which a reader who accepted its argument already believes. Tracing the result through the registry offers a clear way to learn how the registry works. The conclusion is familiar, so a reader can focus on how its supporting claims are arranged.

The registry states that result as THM-004:

THEOREM 2 · THM-004

Scaled AI Deployment Requires Cost Governance for Economic Control

Within the defined deployment and measurement boundary, if:

- (i) production scaling expands the usage surface of AI deployment;
- (ii) measurement enables visibility into measurable cost-bearing activity;
- (iii) visibility enables management mechanisms; and
- (iv) buyer total cost extends beyond access price;

then scaled AI deployment requires cost governance for economic control.

Each condition has a plain meaning a manager can check against a deployment. The first says the deployment has moved from a pilot into daily operations, multiplying the places where usage can occur. The second says that measuring usage makes the activity that costs money

visible. The third says that monitoring, limits, attribution, or other management mechanisms can be applied to visible activity. The fourth says the organization's total cost exceeds the price of access. When all four hold, an organization that wants economic control over the deployment needs cost governance.

THM-004 names four certified lemmas as its support, one for each condition. LEM-016, "Production Scaling Expands Usage Surface," supports the first. LEM-002, "Measurement Enables Visibility," supports the second. LEM-020, "Visibility Enables Management Mechanisms," supports the third. LEM-006, "Buyer Total Cost Extends Beyond Access Price," supports the fourth.

The registry also records the theorem's proof, which combines the lemmas in two pairs. The first and fourth establish that economic exposure grows with the deployment and cannot be determined from the access price alone. The second and third establish the only route to managing that exposure: measurement makes it visible, allowing management mechanisms to be applied. Economic control therefore requires an apparatus for both measurement and management, which the theorem calls cost governance.

This pairing explains Chapter 2's statement that THM-004 formalizes the economic consequence of the third flow's asymmetry. Costs accrue with use and grow with scale, as the first and fourth conditions describe. Control requires deliberate measurement and management, as the second and third conditions describe.

Each lemma, in turn, rests on propositions. LEM-020 offers a useful example of how the registry is built. Its statement reads: "If economic exposure or usage activity is visible to the relevant actor, then monitoring, limits, attribution, or management mechanisms can be applied within the relevant scope." It rests on seven propositions and one other lemma.

The propositions are small claims that appear nearly self-evident in isolation. PROP-043 states: “If activity is metered, records describing measured activity are generated or maintained.” PROP-053 states: “If economic exposure exists, monitoring mechanisms may be used to observe the sources or magnitude of that exposure.” PROP-085 states: “If usage information becomes visible, decisions regarding management, allocation, optimization, or monitoring may reference that information.” Their contribution becomes clearer when they are combined. PROP-043 establishes that metered activity leaves a record. PROP-085 allows management decisions to draw on visible usage, and PROP-053 establishes that monitoring can be used to observe economic exposure. None of the three alone says that controls can be applied to visible activity. LEM-020 reaches that conclusion by combining them with four further propositions and LEM-002, a combination the registry records as its proof.⁸

LEM-020 also rests on LEM-002, which directly supports THM-004’s second condition. Figure 3.1 shows the branch through LEM-020 and makes this shared support visible. The registry is a graph in which one claim can support several others, rather than a tree in which each claim has one parent.

⁸. AI Business Economics registry, “dag.aiom, commit 9d7ee50,” August 21, 2026. LEM-020’s proof record and its supporting propositions.

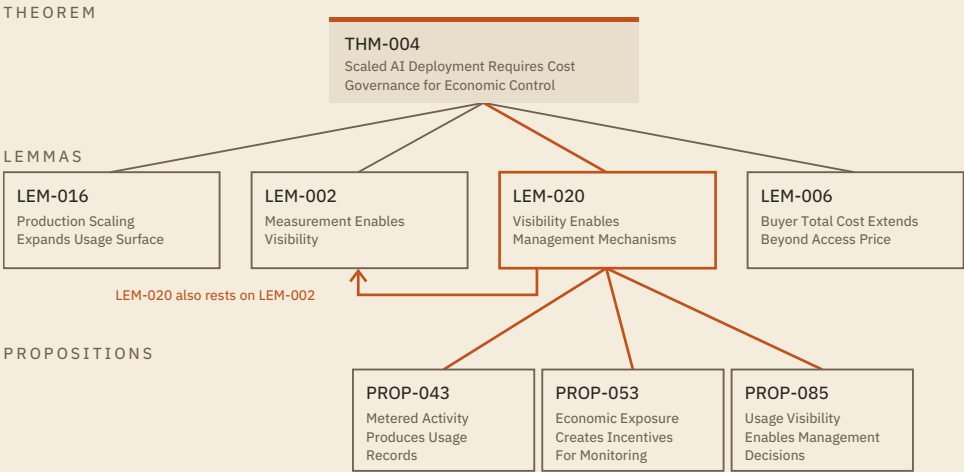


Figure 3.1 One branch of the trace. THM-004 rests on four lemmas, one for each condition. LEM-020 rests in turn on LEM-002, which also supports the theorem directly, and on seven propositions, three of which are shown.

Tracing the theorem changes how a manager can use it. Initially, THM-004 is a conclusion to accept or reject as a whole. The trace reveals a set of smaller claims that can be examined individually, allowing a manager to identify which supporting claim they doubt. A disagreement over an entire conclusion has no obvious way to end. A disagreement over a specific proposition concerns one stated claim, so the two sides can name the fact that would decide it.

DATED: AUGUST 2026

In the version used to write this book, the registry held 413 objects: 140 definitions, 35 axioms, 201 propositions, 21 lemmas, 11 theorems, and 5 records identifying evidence the science still needs. It recorded 1,777 dependency links among these objects, and 397 of the 413 were certified, including 9 of the 11 theorems and 20 of the 21 lemmas.⁹ The registry changes as the science develops, so these figures describe a single version rather than the science itself.

⁹. AI Business Economics registry, “dag.aiom, commit 9d7ee50,” August 21, 2026. Object counts, links and certification at the pinned registry version.

3.3 How to read a formal claim

A manager who cites a formal claim for more than it establishes bases a decision on something the claim never stated. To avoid that error, read the claim in four parts: its conditions, its scope, its non-claims, and the falsifying case that would defeat it. A manager who can state all four can use the claim without being misled by it.

The conditions are the “if” clauses, each of which is true or false for a given case. In THM-004, they are the four numbered clauses. If a deployment fails any one of them, the theorem says nothing about it. A pilot used by one team for a few weeks has not been scaled into daily operations, so it fails the first condition. The theorem neither requires governance for that pilot nor rules it out. Reading the conditions first prevents a basic misuse: applying a formal claim to a case it does not cover.

The scope defines which cases the claim covers. As presented in this book, THM-004 states its scope before the word “if”: one defined deployment and its measurement boundary. The registry adds that the theorem applies when an organization seeks economic control over the resulting cost.¹⁰ The claim concerns one deployment within one boundary, rather than an organization’s AI activity in general. Before applying it across a company, a manager must show that each deployment meets the conditions.

The non-claims identify conclusions the claim deliberately does not establish. THM-004 states that governance is necessary for control, but does not establish that governance delivers control or is costless. It does not specify the scale at which governance becomes necessary or the size of the economic exposure.¹¹ The registry records non-claims for every certified claim, but its lists are not exhaustive. THM-004 also does

¹⁰. AI Business Economics registry, “dag.aiom, commit 9d7ee50,” August 21, 2026. THM-004’s scope condition.

¹¹. AI Business Economics registry, “dag.aiom, commit 9d7ee50,” August 21, 2026. THM-004’s recorded exclusions.

not establish Chapter 2’s three flows; that exclusion comes from this book’s framework, rather than the registry. A reader with only the statement can identify most of them by asking whether it establishes that a requirement is necessary or sufficient, whether it specifies an amount, and whether it accounts for cost.

A costly misreading confuses necessary with sufficient. A manager who interprets THM-004 as “governance produces control” might buy a dashboard and expect costs to come under control. The theorem promises no such result; it states only that control requires governance. LEM-020 makes the same limit explicit one level down: it does not establish that management mechanisms are automatically effective.¹²

DEFINITION

Non-claim

A conclusion a formal claim deliberately does not establish, though a reader might be tempted to draw it. Stating non-claims prevents a claim from being cited for more than it says.

The falsifying case describes what would show the claim to be wrong. For a conditional claim, the test is precise: a case that meets every condition but lacks the conclusion. The second and third conditions describe what measurement and management would make possible, rather than an apparatus already in place. A deployment with no instruments installed can therefore meet all four conditions.

For THM-004, the falsifying case would be a scaled deployment that meets all four conditions and whose organization controls its cost without any measurement and management apparatus. A flat-fee contract does not by itself provide such a case. If the fee covers every cost, the fourth condition fails and the theorem is silent. If it does not, a deployment under that contract is a falsifying case only if it meets all four conditions and the organization controls the costs beyond the fee with-

¹². AI Business Economics registry, “dag.aiom, commit 9d7ee50,” August 21, 2026. LEM-020’s recorded exclusions.

out measuring them. A claim that identifies no possible falsifying case cannot be shown to be wrong and tells a manager nothing.

These four parts answer an objection to the formal style: that the registry offers word games, dressing obvious points in numbered clauses. The common-sense version of THM-004 says that big deployments need cost controls, but gives no answer for the one-team pilot because it never defines “big.” The formal version does: the pilot fails the first condition, so the theorem is silent about it. The manager must first check whether the deployment has moved into daily operations. The common-sense version also allows readers to slide from “needs controls” to “controls deliver control,” a misreading the non-claims rule out. A formal statement fixes in advance what would count as the claim being wrong. Stating its conditions makes it possible to test and improve it.

3.4 The discipline that acts on the science

Knowledge alone does not change what an organization does. Medicine draws on physiology, but clinicians treat patients by applying that knowledge. Engineering draws on physics, but engineers build bridges using codes and procedures. In each case, a practice connects the knowledge to the result through its own methods, roles, and standards.

AI Operations Management is the practice of managing an organization’s AI consumption from week to week. An organization that practices it buys AI capacity, plans its use, records consumption, and assigns it to the work it serves. It decides what runs when capacity is limited and holds spending accountable for the value it returns. Chapter 1 described the problem as a category error. Wherever that chapter’s conditions

DEFINITION

AI Operations Management

The management discipline that acts on AI Business Economics: the practice by which an organization buys, plans, records, assigns, allocates, and holds accountable its consumption of AI.

hold, AI Operations Management addresses the error by managing AI as a resource the organization consumes.

The discipline has five functions that follow AI capacity through the organization: buying it, planning its use, recording and assigning consumption, routing it to work, and judging the value it returns. Each function is a capability an organization either has or lacks. Part III teaches all five, with metering and attribution covered in two chapters.

The first function, sourcing, matches the AI capacity an organization buys to the requirements of its work. It determines which models and providers serve which work, and on what terms, so that the organization neither pays for capability it does not need nor deprives work of capability it requires. A company that sends every request to its most capable model, including requests a smaller model would answer just as well, is paying for capability its work does not use.

The second function is planning and budgeting, whose absence proved costly in Chapter 2's opening case. The organization must state how much AI it expects to consume in the coming period and detect when actual consumption departs from that expectation. A budget is a commitment made in advance that makes deviations visible as they happen, before the money is gone.

The third function, metering and attribution, builds what Chapter 2 called the record flow. The organization must record what was consumed and assign each unit of consumption to the team, product, or purpose it served. Without that assignment, managers can see a cost but cannot trace it to the work that caused it.

The fourth function is allocation and routing. When capacity is limited, the organization must decide what runs first and send each piece of work to suitable capacity. These choices occur even when no one makes them deliberately. Consider a constructed example: two workflows share a provider account with a fixed rate limit. When their combined demand exceeds the limit, one of them waits, and no one has decided

which. Allocation and routing make that choice deliberate by applying a rule.

The fifth function is holding value to account. Consider a constructed example in which a support team adds an AI assistant to its help desk. At the end of a quarter, the organization should be able to say whether the assistant paid for itself, identifying what it cost and returned, over which months, and against which outcome. LEM-021, “ROI Evaluation Requires Measurement Boundary,” states the requirement: “If ROI is evaluated for AI adoption, the evaluation requires a defined boundary for relevant costs, value, time period, and observed or expected outcomes.”¹³ This book adds two requirements of practice that the lemma does not state: the boundary is set for each deployment before spending begins, and one named person is responsible for knowing the answer. This book calls it the value boundary.

Figure 3.2 shows the relationship between the two layers: the science provides the foundation for the discipline’s five functions.

¹³. AI Business Economics registry, “dag.aiom, commit 9d7ee50,” August 21, 2026. LEM-021’s statement.

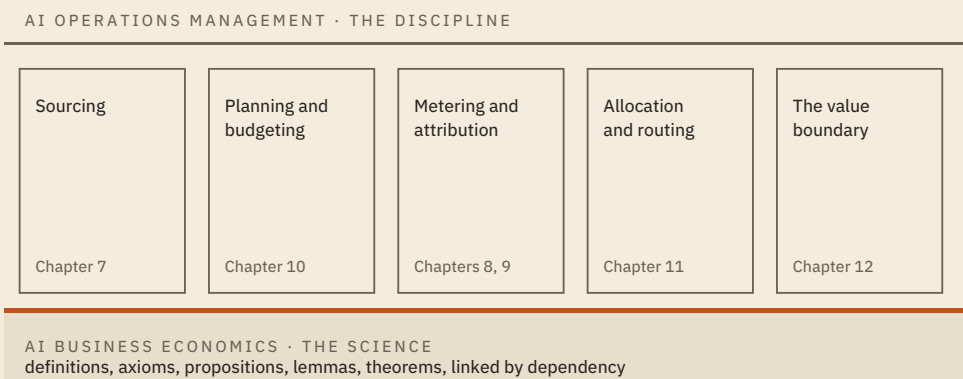


Figure 3.2 *The two layers. AI Business Economics states conditional claims about consuming AI and records what each one rests on. AI Operations Management acts on it through five functions, all taught in Part III.*

A result about governance changes nothing until an organization builds the apparatus it describes. The science therefore depends on the discipline to put its findings into practice. The discipline depends on the science because careful practice cannot repair a decision built on an incorrect claim.

3.5 Where the discipline ends

Chapter 1 set aside three related subjects: how models work, prompt engineering, and finding uses for AI. Each has its own purpose and literature, and each ends before AI Operations Management begins.

The first subject concerns how models work, including their architectures, training, and internal behavior. It belongs to computer science and machine learning engineering, and ends where a model becomes a service an organization can call. AI Operations Management begins where a model is a service an organization can buy and call.

The second is prompt engineering, the craft of writing inputs that produce better outputs. It improves individual uses of AI, while AI Operations Management governs their combined consumption, cost, and value. A better prompt can lower the cost of one request. Only the disci-

pline can determine whether the organization's requests, taken together, are worth what they cost.

The third subject is finding uses for AI: deciding which business problems it might solve. That work belongs to strategy and innovation, and ends when a use is chosen. AI Operations Management begins with buying capacity for the chosen use and manages its consumption from then on.

Four established disciplines sit closer to AI Operations Management. Each manages an aspect of work related to AI consumption, but each lacks something needed to govern its economics. Table 3.1 compares them.

NEIGHBOR	WHAT IT IS FOR	WHAT IT MANAGES	WHAT IT LACKS FOR THIS TERRITORY
AIOps	Applying big data and machine learning to automate IT operations, including event correlation, anomaly detection, and causality determination ¹⁴	The health and performance of IT systems	It uses AI to watch systems. It does not govern what consuming AI costs or returns
MLOps	Building, deploying, and monitoring machine learning models in production ¹⁵	The lifecycle of models an organization builds	It ships models. It does not govern the consumption of capacity an organization buys
FinOps	Maximizing the business value of technology through financial accountability and collaboration among engineering, finance, and business teams ¹⁶	Cloud and wider technology spending	It grew from managing cloud cost and usage. Its 2026 report quotes a practitioner saying that no one can yet answer whether AI is providing value
Regulatory AI governance	Ensuring that uses of AI are lawful, safe, and accountable to regulators	Risk and compliance	It asks whether a use is permitted, not whether it is economically managed

Table 3.1. The four neighboring disciplines. Each manages work related to AI consumption, but none was built to govern its economics.

FinOps is the closest neighboring discipline, making its boundary with AI Operations Management the hardest to define. It already manages spending, has expanded into AI, and defines its purpose in terms of

¹⁴. Gartner, “AIOps (Artificial Intelligence for IT Operations),” *Gartner Information Technology Glossary*, accessed October 1, 2026. Gartner’s definition of AIOps.

¹⁵. Dominik Kreuzberger, Niklas Kühl, and Sebastian Hirschl, “Machine Learning Operations (MLOps): Overview, Definition, and Architecture,” *IEEE Access* 11 (2023): 31866-31879. Definition and scope of MLOps.

¹⁶. FinOps Foundation, “What is FinOps?,” accessed October 13, 2026. The Foundation’s own definition of the practice.

maximizing the business value of technology.¹⁷ Chapter 14 explains how the two divide the work. Their distinction concerns what each was built to do, rather than its intent. FinOps was built to make cloud spending accountable when bills arrived after capacity had been used. AI Operations Management manages consumption metered by the request and judges it against a value boundary set before spending begins.

Regulatory governance determines whether an organization may use AI in a particular way. This book takes that determination as given and asks how the permitted use is managed.

3.6 Five questions the discipline exists to answer

Each of the five questions corresponds to one of the discipline's functions and is answered in the chapter that teaches that function. Part III covers them in a different order, for reasons explained there.

An organization that manages its AI consumption can answer five questions from its own records:

Question 1. “Are you paying for capability you do not need, or starving work that needs more?” This is the sourcing question, answered in Chapter 7.

Question 2. “What do you expect the AI flow to consume next period, and how will you know when it deviates?” This is the planning and budgeting question, answered in Chapter 10.

Question 3. “Who consumed what last month, in service of which work?” This is the metering and attribution question, answered across Chapters 8 and 9.

Question 4. “When capacity is constrained, who decides what runs first, and by what rule?” This is the allocation question, answered in Chapter 11.

¹⁷. FinOps Foundation, “What is FinOps?,” accessed October 13, 2026. The Foundation’s definition, which names business value.

Question 5. “Where, exactly, does this AI flow pay for itself, and who is on the hook for knowing?” This is the value boundary question, answered in Chapter 12.

These are the Founding Questions, which an organization must answer “from its own records.” It can always produce an answer by asking the provider, estimating, or repeating the business case, but none of those methods produces knowledge the organization holds and can check.

At the end of Part I, readers do not yet have a method for answering these questions from records. Parts II through IV provide it: Part II presents the certified claims the functions rest on, Part III builds the five functions, and Part IV turns them into an institution.

CRAFT SECTION

The trace procedure

The **trace procedure** helps a manager examine a formal claim before relying on it. It breaks a theorem into smaller supporting claims that can be checked individually, allowing the manager to assess more than the conclusion as a whole. The procedure has six steps.

Step 1. Locate the claim in the registry by its identifier and record its name exactly as the registry gives it. A claim’s name is a label; only its statement says what the claim establishes.

Step 2. State its scope, then restate each condition in ordinary business language. Retain every antecedent and add none.

Step 3. List the claims it depends on by identifier and name.

Step 4. Follow the dependencies one level down. Restate each in one sentence and note any condition it adds that the claim above did not state.

Step 5. State what the claim establishes and, separately, what it does not establish.

Step 6. Name the falsifying case: one that meets every condition but lacks the conclusion.

Steps 2 and 6 do most of the work. Dropping an antecedent in step 2 produces a claim the registry does not make. Step 6 tests whether the reader has understood the claim: a reader who cannot describe the case that would defeat it cannot say what it rules out.

Applied to THM-004, the procedure reveals a requirement of authority at step 4 that the theorem's wording does not state. The following trace shows how.

Step 1 records THM-004, "Scaled AI Deployment Requires Cost Governance for Economic Control."

Step 2 restates the conditions. Within one deployment and its measurement boundary, the theorem applies when four conditions hold: scaling expands where usage can occur; measurement makes the activity that costs money visible; management mechanisms can be applied to visible activity; and total cost exceeds the access price.

Step 3 lists the four dependencies: LEM-016, LEM-002, LEM-020, and LEM-006, one for each condition. The registry marks all four as required. Some claims also have contextual dependencies, which bear on a claim without being conditions of it. Problem P2 provides an example. Step 3 records whether each dependency is required or contextual.

Step 4 restates each dependency. LEM-016 says that moving from isolated use into repeated operations can increase the number of usage events and demand points. LEM-002 says that measured and recorded activity becomes observable enough to support analysis, attribution, monitoring, or management. LEM-020 says that monitoring, limits, attribution, or other management mechanisms can be applied to visible usage or exposure. LEM-006 says that a buyer who incurs several kinds of cost pays more in total than the access price.

Step 4 also examines what each dependency adds. LEM-020's scope requires reliable visibility and the authority to apply management mechanisms.¹⁸ That requirement belongs to the scope, rather than the third condition. A team that can see its costs but lacks authority to cap them falls outside LEM-020's scope. The support for THM-004's third condition therefore does not extend to that team. The trace reveals where the lemma's support stops, a limit not stated in the theorem's own wording.

Step 5 separates what the claim establishes from what a reader might infer. THM-004 establishes that cost governance is necessary for economic control of a scaled deployment. It does not establish that governance is sufficient or costless, specify the scale at which governance becomes necessary or the size of the exposure, or establish that a deployment has three flows.

Step 6 names the falsifying case: a scaled deployment that meets all four conditions and whose organization controls its cost without a measurement and management apparatus. Finding such a case would show the theorem to be false.

The procedure is deliberately slow. Tracing a theorem takes longer than reading it and agreeing with it, but the time is worth spending once for each claim on which a decision depends. The trace shows a manager exactly what supports the decision.

Chapter summary

AI Business Economics is the science of consuming AI: an ordered body of conditional claims, each stating when it applies, what it concludes, and which other claims it depends on. The claims are kept in a registry and linked by dependency. This book renders and cites only certified claims.

¹⁸. AI Business Economics registry, "dag.aiom, commit 9d7ee50," August 21, 2026. LEM-020's scope condition.

The registry provides the book's justification, while the manager's learning needs determine its organization. The registry supports each formal result the book presents, so that a reader who doubts one can trace its basis.

A trace follows a claim through the registry to its supporting lemmas and propositions. It breaks a conclusion into smaller claims that can be examined individually.

A formal claim must be read in four parts: its conditions, its scope, its non-claims, and the falsifying case that would defeat it. Stating a claim formally fixes in advance what would count as being wrong.

AI Operations Management applies the science through five functions: sourcing, planning and budgeting, metering and attribution, allocation and routing, and holding value to account at a boundary. It begins when a model becomes a service an organization consumes. It shares boundaries with AIOps, MLOps, FinOps, and regulatory governance, while remaining a distinct discipline.

The Founding Questions are the five questions the discipline exists to answer from an organization's own records. By the end of Part I, readers understand the questions but do not yet have a method for answering them.

Key terms

AI Business Economics

The science of the business economics of consuming AI: an ordered body of conditional claims, each stating when it applies, what it concludes, and which other claims it depends on.

AI Operations Management

The management discipline that acts on AI Business Economics: the practice by which an organization buys, plans, records, assigns, allocates, and holds accountable its consumption of AI.

Registry

The structured record of every definition, assumption, and result in AI Business Economics, with each result linked to the claims that support it.

Theorem

A registry result that combines lemmas into a conclusion with direct consequences for management. The book presents each theorem it relies on in a numbered panel.

Lemma

An intermediate registry result that combines propositions, and sometimes other lemmas, to support one or more theorems.

Proposition

A single registry claim, accepted because of the definitions, axioms, or earlier claims it names as its support. Lemmas combine propositions.

Certified claim

A registry claim that meets the validator's fixed conditions and whose dependencies are all certified. The book renders and cites only certified claims.

Trace

The process of following a registry claim through its supporting claims, from a theorem to its lemmas and from a lemma to its propositions.

Dependency

A claim that another claim rests on. A required dependency is one without which the claim cannot hold. A contextual dependency bears on the claim without being one of its conditions.

Condition

A requirement a deployment must meet for a formal claim to apply to it. A claim says nothing about a case that fails any of its conditions.

Scope

The part of a formal claim that defines which cases it covers, such as one deployment within one boundary. A case outside that scope is not covered, regardless of whether it meets the conditions.

Non-claim

A conclusion a formal claim deliberately does not establish, though a reader might be tempted to draw it. Stating non-claims prevents a claim from being cited for more than it says.

Falsifying case

A case that meets every condition of a formal claim but lacks its conclusion. Finding such a case would show the claim to be wrong.

Value boundary

The boundary within which an organization judges whether a deployment pays for itself, defined by the costs and value it counts, the period it covers, and the outcomes it measures against. LEM-021 requires such a boundary whenever return on AI is evaluated. Setting one for each deployment before spending begins is this book's practice.

Founding Questions

The five questions AI Operations Management exists to answer from an organization's own records, one for each of its five functions.

Trace procedure

The six-step method for examining a formal claim before relying on it: locate the claim; state its scope and restate its conditions; list its dependencies; follow them one level down; separate what the claim establishes from what it does not; and name the case that would defeat it.

Discussion questions and problems

DISCUSSION QUESTIONS

- 1 The book states that the registry provides its justification but does not determine its organization. Describe what a reader would lose if the chapters followed the registry's order of derivation.
- 2 Of the four borders in Table 3.1, which is hardest to defend, and why? State what evidence would move that border.
- 3 Restate one Founding Question in language your chief financial officer would use, preserving all its content. Then identify the part you were most tempted to omit.
- 4 Chapter 2 diagnosed a deployment through its three flows. Match each flow to the functions in Section 3.4 that act on it, then explain which flow draws on the most functions and why.

PROBLEMS

P1 · WORKED

A trace of LEM-020

The craft section traced a theorem. Use the same six steps to trace one of its lemmas, LEM-020, as a model for Problem P2.

Step 1. LEM-020, "Visibility Enables Management Mechanisms." Its statement reads: "If economic exposure or usage activity is visible to the relevant actor, then monitoring, limits, attribution, or management mechanisms can be applied within the relevant scope."

Step 2. Scope: the relevant actor has reliable visibility and the authority to apply management mechanisms. Within that scope, when the people responsible for a deployment can see its usage or cost, they can apply monitoring, limits, attribution, or other management mechanisms to it.

Step 3. Eight dependencies, all required: the lemma LEM-002 and seven propositions, PROP-032, PROP-039, PROP-041, PROP-042, PROP-043, PROP-053, and PROP-085.

Step 4. LEM-002 says that recorded activity becomes observable enough to support analysis, attribution, monitoring, or management. Its scope adds a requirement: the record must be detailed enough for the relevant person to observe the activity.¹⁹ PROP-032 says compute use can be measured. PROP-039 says measured compute use can inform methods for assigning cost. PROP-041 says that metering a resource or activity records one or more measurements of it. PROP-042 says measured activity can be distinguished by quantity. PROP-043 says metering produces records of what was metered. PROP-053 says monitoring can be used to observe economic exposure, adding the condition that exposure exists. PROP-085 says management decisions can draw on visible usage. Together, these claims explain how visibility makes management possible.

Step 5. LEM-020 establishes that visibility makes management mechanisms available. It does not establish that the mechanisms work automatically or that visibility alone determines the right policy or price.²⁰

Step 6. The falsifying case is an organization that can see a deployment's usage and has authority to manage it, but still cannot apply any monitoring, limit, or attribution to it.

P2 · COMPLETION

A trace with two steps left open

THM-002, "AI Adoption Cost Is Underestimated When Access Price Is Treated as Total Cost," states a result taught in Part II. Its statement reads: "If buyer AI adoption cost includes cost categories beyond access price, and ROI evaluation

¹⁹. AI Business Economics registry, "dag.aiom, commit 9d7ee50," August 21, 2026. LEM-002's scope condition.

²⁰. AI Business Economics registry, "dag.aiom, commit 9d7ee50," August 21, 2026. LEM-020's recorded exclusions.

requires a defined measurement boundary for cost and value, then treating access price as total adoption cost creates underestimation risk within the evaluation boundary.” Steps 1, 2, 3, and 6 are worked below. Complete steps 4 and 5.

Step 1. THM-002, as named above.

Step 2. When a buyer’s cost of adopting AI includes more than the access price, and evaluating the return requires a boundary that includes both cost and value, treating the access price as the whole cost risks understating it.

Step 3. The theorem has four dependencies. LEM-006 and LEM-021 are required, while LEM-002 and LEM-020 are contextual. LEM-006 appears in the craft section. LEM-021, “ROI Evaluation Requires Measurement Boundary,” reads: “If ROI is evaluated for AI adoption, the evaluation requires a defined boundary for relevant costs, value, time period, and observed or expected outcomes.” LEM-020 concerns what a buyer can do about an understated cost once that cost is visible. It is contextual: the claim that treating access price as total cost creates underestimation risk holds without it.²¹

Step 4. Restate each dependency in one sentence and note any condition it adds.

Step 5. State what THM-002 establishes and what it does not. Then use the treatment of LEM-020 in step 3 as a model to explain what LEM-002 contributes, given that the theorem does not depend on it as a condition.

Step 6. The falsifying case is a buyer whose adoption cost includes categories beyond the access price and who evaluates the return within a boundary covering cost and value. The buyer treats the access price as the total cost, yet doing so creates no risk of understating the cost within the evaluation boundary.

P3 · INDEPENDENT

A reply to the word-games objection

A senior colleague reads THM-004 and says: “This is common sense in numbered clauses. Anyone who has run a large deployment knows you need controls. Formalizing it adds nothing.” Write a one-page reply using the four parts of a formal claim from Section 3.3. Identify at least one thing the formal state-

²¹. AI Business Economics registry, “dag.aiom, commit 9d7ee50,” August 21, 2026. THM-002’s dependencies and their recorded roles.

ment lets the colleague do that the common-sense version does not. Do not argue that the colleague's conclusion is wrong.

Interleaving: question 4 requires Chapter 2's three flows, and the cumulative case below requires both earlier chapters.

PART I CUMULATIVE CASE

CUMULATIVE CASE · PART I

One announcement, three readings

DATED: FEBRUARY 2024. THE COMPANY'S OWN ANNOUNCEMENT, USING ITS OWN FIGURES. ONLY THAT ANNOUNCEMENT IS USED HERE.

In February 2024, the payments company Klarna announced results from the first month of an AI assistant powered by OpenAI. The company reported that the assistant had handled 2.3 million conversations, accounting for two-thirds of its customer-service chats. It estimated that the assistant was doing the work of 700 full-time agents. The company reported that, in that first month, customers resolved their questions in under two minutes, compared with eleven minutes previously. Klarna estimated that the assistant would improve its profit by \$40 million in 2024.²²

Work from this announcement alone.

First, apply Chapter 1's consumption-event inventory. Identify the consumption events described in the announcement and where each would be metered.

Second, apply Chapter 2's three-flow mapping. Diagnose the usage, record, and cost-and-value flows, identifying the evidence behind each diagnosis.

Third, apply each of the five Founding Questions to the announcement. For each question, state whether the announcement provides an answer from the organization's own records and what record would be needed to answer it.

Keep your answers. The case returns in Chapter 6.

²². Klarna, "Klarna AI assistant handles two-thirds of customer service chats in its first month," February 27, 2024, accessed October 13, 2026. All figures in this case are the company's own, as announced.