AI Operations

A Practitioner's Framework for Turning AI Capability into Business Value

Roberto Shimizu

April 2026

Executive Summary

Most enterprise AI programs do not fail because the model is weak. They fail because the surrounding work system is weak. Workflows are not redesigned, baselines are not established, ownership remains fragmented, and governance stays stuck at the policy level instead of being built into production behavior. The result is a familiar pattern: strong demos, active pilots, and disappointing operating results.

This paper argues that organizations need a distinct management discipline for turning AI capability into business value. I call that discipline *AI Operations*. It combines measurement discipline, workflow redesign, semantic structure, adoption design, and governance into one operating model. The goal is not more AI activity. It is dependable performance in production.

1 Introduction

Enterprise AI has a strange problem: capability is no longer scarce, but dependable business value still is. Most large organizations can now access strong foundation models, buy copilots off the shelf, and run pilots with relatively little friction. What remains difficult is turning that access into repeatable operating gains. McKinsey's 2025 global survey found that AI use is now widespread while meaningful enterprise-wide bottom-line impact remains concentrated in a small minority of companies. BCG described the same pattern in sharper operating terms: a small group is generating material value, a larger group is scaling with partial returns, and most companies are still seeing little measurable impact despite active investment.[9, 2]

That gap is usually described as an adoption problem, a governance problem, or a last-mile problem. All three descriptions are useful, but none is quite sufficient. Nigel Vaz makes a helpful move when he argues that AI should be understood less as a technology than as an operating system for how a business works.[12] What I keep seeing in the field is the same pattern in practice: organizations are trying to place AI into operating systems that were never redesigned to absorb it.

In IntelliDoctor, a hospital AI deployment I helped build, the technical signal was not the hardest part to establish. In one constrained clinical documentation workflow, AI-generated conduct quality was scored at 8.0/10 against a physician baseline of 3.5/10, and physicians accepted recommendations in 78.5% of cases. Yet the harder question began after the model performed well. Which workflow

should change first? Who owned the exceptions? What metric mattered most? How should leadership judge value with confidence rather than enthusiasm? The operating problem was visible even where the AI worked.

That is the central claim of this paper: most AI initiatives fail operationally before they fail technically. By operationally, I do not mean uptime alone. I mean the surrounding work system: baselines, workflow design, ownership, semantic consistency, adoption behavior, and governance in production. A model can be accurate and still destroy value if it adds rework, supervision burden, escalation cost, or organizational ambiguity.

This paper proposes *AI Operations* as the missing management discipline between AI ambition and business performance. The term is useful because the adjacent labels do not fully cover the job. Strategy language is too abstract. MLOps and LLMOps are too technically narrow. Governance is often too static. Change management alone does not say enough about instrumentation, semantics, or production control.

The argument proceeds in four moves. First, it explains why AI programs usually stall in the operating system around the model rather than in the model itself. Second, it defines AI Operations as a practical management discipline. Third, it argues that the semantic layer is not optional architecture but the infrastructure of reuse, control, and institutional learning. Finally, it translates the framework into a short set of executive tests for deciding whether an AI initiative is real or merely theatrical.

2 Why AI Fails Operationally Before It Fails Technically

Most AI programs do not break at the point of inference. They break at the point of operation. A convincing demo creates belief. A pilot proves local usefulness. Then scale exposes the real conditions of enterprise work: ambiguous handoffs, legacy roles, compliance constraints, fragmented systems, hidden exception paths, and managers who were never forced to define what success would actually look like. *Harvard Business Review* recently described this as the "last mile" problem of AI transformation and captured the pattern well: many organizations are becoming pilot-rich and transformation-poor.[7]

The decisive question is therefore not whether the model performs in isolation. It is whether the workflow performs under real operating conditions. That is a harder standard. It asks whether the system improves throughput, quality, compliance, economics, or judgment once it enters the mess of actual work. Many AI programs never reach that standard because they confuse prototype success with operating design.

The pilot trap

Pilots are useful. They reduce uncertainty, surface edge cases, and build internal confidence. But a pilot is not an operating model. It usually runs with hand-picked users, concentrated attention, and unspoken heroics from the team closest to the system. Those conditions do not survive scale.

This is why better models alone rarely solve the deployment problem. A stronger model may improve answer quality at the margin, but it does not decide who acts on the answer, when the answer is

trusted, how exceptions are handled, or which downstream process absorbs the change. McKinsey's 2025 survey found that organizations seeing stronger AI impact were much more likely to redesign workflows rather than merely insert tools into existing routines.[9] That is not a technical footnote. It is the difference between a feature and an operating capability.

Workflow misfit is where economics turn

A technically impressive assistant can still be economically useless. If it adds one more review step, one more uncertainty, or one more place where a human must translate between systems, the workflow may slow down even while the model appears fast. This is the hidden arithmetic of enterprise AI: generation speed is obvious; supervision cost is easier to miss.

In practice, AI often shifts work rather than simply removing it. It moves effort toward verification, escalation, and judgment. That can be beneficial when the workflow is redesigned around those new control points. It becomes destructive when the old process remains intact and AI is simply layered on top. Output volume rises. Hidden review work rises with it. Confidence falls. Tom Peperkamp makes the same point from a lean perspective: speeding the local act of production does not help much if the actual bottleneck sits elsewhere in the stream.[11]

Healthcare makes this visible quickly because workflow misfit does not stay theoretical for long. A draft recommendation is only useful if the surrounding clinical pathway knows when to adopt it, when to question it, and who is responsible if the case falls outside the expected path. In IntelliDoctor, that distinction appeared repeatedly. The issue was seldom whether the system could generate a clinically relevant suggestion. The issue was whether the suggestion landed inside a workflow designed to absorb it.

Capability is not the same as absorption

One of the most useful distinctions for executive teams is the difference between *capability* and *absorption*. Capability is what the model can generate. Absorption is what the organization can reliably take in, route, verify, and reuse.

Most firms are overestimating the first and underinvesting in the second. Anthropic's Economic Index and Peter McCrory's related discussion of adoption both point in the same direction: capability does not automatically become business value when the relevant organizational knowledge remains tacit, fragmented, or difficult to supply at the point of use.[1, 8] Critical operating knowledge still lives in people's heads: which specialist tends to take a certain referral, which diagnosis changes reimbursement exposure, which protocol matters for discharge safety, which exception requires escalation, and which output is safe enough to move without review. When that knowledge stays tacit, AI can only perform as a thin assistant around the edges. It cannot participate reliably in the real decision system.

That is why so many deployments look promising in the lab and fragile in production. The model is capable, but the organization has not made its own judgment logic legible. People then interpret the resulting failures as model immaturity when the deeper problem is institutional opacity.

Measurement vacuums create AI theater

The absence of measurement is one of the biggest reasons weak AI programs survive too long and promising ones struggle to earn sponsorship. If no one baselines the current workflow, almost any pilot can claim progress. If no one defines the economic metric, no one can say whether improved output quality mattered. If no one separates trial usage from durable behavior change, adoption gets overstated.

This is where AI theater begins. Organizations report number of pilots, number of users, model names, vendor contracts, and anecdotal enthusiasm. What they cannot report cleanly is whether the workflow improved and by how much.

The IntelliDoctor experience made this concrete for me. The technical gains were visible. What management still needed was a shared operating language for value. Did physician behavior change in a durable way? Did documentation quality improve in a way that reduced downstream loss? Did referral capture improve? Did protocol compliance become more visible? Which of those changes were repeatable, and under what conditions? Those are harder questions than "Does the model work?" but they are the only ones that matter once real money, risk, and accountability enter the picture.

Adoption and ownership determine whether value compounds

AI systems do not succeed because users are impressed. They succeed because people know when to rely, when to verify, and when to escalate. A user who is told only that the system is "smart" has not been given operational guidance. They have been given ambiguity.

Ownership matters just as much. In many organizations, no one owns the end-to-end behavior of the AI-enabled workflow. Technology owns the tool. A business function owns the nominal process. Risk owns the policy. No one owns performance in production. When results disappoint, the model gets blamed because it is the most visible artifact. The real failure was fragmented accountability.

A small clinical example illustrates the point. In one headache-related case, the relevant protocol was available and the system surfaced the right operational guidance, yet the recommendation was not followed and the patient returned within 24 hours. The lesson was not that the model needed another benchmark. The lesson was that adoption, workflow control, and escalation discipline were incomplete.

That is why most serious AI failure is operational before it is technical. If workflow design, measurement, tacit-knowledge capture, adoption rules, and ownership remain unmanaged, better models mostly magnify unmanaged complexity. That is the problem AI Operations is meant to solve.

3 AI Operations as a Management Discipline

If AI is going to matter operationally, it needs an operating discipline, not just a model strategy. AI Operations is that discipline: the practice of designing, measuring, governing, and continuously improving AI-enabled work so that technical capability becomes dependable business performance.

The term earns its place because the adjacent labels leave important gaps. MLOps and LLMOps matter, but they do not fully address workflow redesign, adoption behavior, semantic consistency, business instrumentation, or executive control. "Transformation" is too abstract. "Automation" is too narrow. "Governance" is often too static and too far from production. AI Operations names the layer where executive intent either becomes operating reality or dissolves into pilot clutter.

It is also not built from scratch. If you came up through DMAIC, Lean, CAP, or large-scale change programs, some of this logic should feel familiar. It should. The durable disciplines of defining outcomes, establishing baselines, exposing bottlenecks, and building control into the operating system did not become obsolete when AI arrived. What changed is where that discipline now has to live: inside workflows, semantic structures, telemetry, and human review boundaries. Even the contemporary literature at the Lean Six Sigma edge is moving in this direction, treating AI not as a replacement for operating discipline but as a force multiplier for it. [6]

The clearest way to describe the discipline is through three moves: what it keeps, what it discards, and what it adds.

What AI Operations keeps

The strongest instincts from older operating disciplines still matter.

- Define the business outcome before celebrating local activity.
- Baseline performance before claiming improvement.
- Look for the bottleneck in the system, not just the defect in the artifact.
- Treat production learning as an operating asset.
- Expect control after deployment, not just effort before deployment.

None of that is nostalgic. It is simply still true.

What AI Operations discards

What should not be carried forward is just as important. I am not arguing for a revival of belt hierarchies, certification theater, or bureaucratic tollgates disguised as rigor. Too much of the old process world decayed into abstract maturity scores, ritualized documentation, and metric games detached from value creation. Those habits created deserved antibodies.

AI programs are already vulnerable to the same failure mode. It is easy to produce policy binders, evaluation templates, and governance committees that look serious while remaining disconnected from workflow behavior. AI Operations rejects that theater. Its job is to make AI deployments operationally legible and controllable.

What the AI era adds

AI changes the shape of operational control, so the discipline must add capabilities that older operating systems did not manage explicitly.

- **Task and prompt design as workflow design.** In many workflows, the prompt is not presentation. It is part of the work instruction.
- **Review boundaries.** The key question is not whether humans remain involved, but where their judgment adds value and where their review burden becomes waste.
- **Context quality.** AI performance is often constrained less by raw model strength than by whether the system is given the right context at the right moment.
- **Exception routing.** No serious workflow is average-case only. Edge cases need explicit paths, owners, and review triggers.
- **Semantic structure.** Shared meaning is not decoration. It is the layer that makes AI behavior reusable, auditable, and improvable across workflows.
- **Operational telemetry.** Accuracy in isolation is insufficient. Teams need visibility into overrides, escalations, latency, downstream rework, compliance exposure, and adoption quality.

The goal is not maximum autonomy. It is dependable performance under constraint.

Keep	Discard	Add
Baseline before action	Belt theater and certification rituals	AI telemetry and production observability
Sponsor accountability	Tollgate bureaucracy detached from value	Semantic structure as shared operating context
Root-cause discipline	Static control plans that never meet production	Human-in-the-loop review boundaries and exception routing
Behavior change as operating work	Metric formalism divorced from economics	Workflow redesign for AI-native decision support

The five requirements of AI Operations

For executive purposes, AI Operations can be reduced to five requirements. If one is missing, the initiative may still produce demos, but it is unlikely to produce durable value.

1. **A defined workflow.** The initiative must target a specific workflow, not a generic capability.
2. **A baseline and value metric.** Someone must know what performance looks like today and what should improve.
3. **A control design.** Human ownership, review boundaries, and exception paths must be explicit.
4. **A shared semantic structure.** The system must know what its key concepts mean across contexts and systems, and what sources ground those meanings.
5. **A production learning loop.** The organization must be able to observe behavior, learn from drift and exceptions, and improve the workflow in production.

This is the compact executive model. Model benchmarking can tell you something about capability. Governance policy can express principles. Neither can substitute for workflow definition, control design, semantic structure, or instrumentation in the work itself.

A practical operating method

The discipline becomes most useful when translated into an operating rhythm:

1. choose a workflow where pain and value are already visible,
2. baseline current performance, adoption behavior, and failure modes,
3. locate the bottleneck, the decision points, and the tacit knowledge the workflow depends on,
4. define the semantic structure: the key concepts, relationships, provenance rules, and exception boundaries,
5. redesign the workflow around AI strengths, human judgment, and explicit controls,
6. deploy, monitor, and improve in production.

This cycle is simple on purpose. Organizations want the acceleration of AI without the inconvenience of redesign. They want better decisions without making decisions about ownership, measurement, context, and exception handling. AI Operations insists that those choices be made explicitly. [6, 5]

The method should also produce artifacts, not just alignment meetings: a workflow map, a baseline scorecard, a semantic or ontology canvas, exception rules, an adoption view, and a review cadence for value and drift. Over time, those artifacts can be supported by reusable tools instead of being recreated for every initiative.

This is where the semantic layer becomes more than a technical component. It is part of the *Define* work because it makes the operating domain explicit, and part of the *Control* work because it provides a stable surface for review rules, provenance, and institutional memory. Once that layer exists, new AI use cases start looking less like fresh projects and more like new configurations built on shared meaning.

Why this is a management system

A management system defines what matters, how performance is seen, who owns the result, and how learning gets folded back into operations. That is exactly what most AI programs lack. Thomas Davenport and Laks Srinivasan's 2026 research is useful here because it shows that companies with stronger discipline around value assessment, aggregation, and reporting are more likely to report meaningful AI returns. They also found that when the finance function helps own or certify value, reported results are stronger.[4] The causal chain can be debated, but the managerial lesson is clear: value realization improves when someone has both the authority and the rigor to measure it.

This also clarifies what AI Operations is not. It is not model benchmarking, generic automation, or transformation theater. And it is not purely MLOps, because a perfectly managed model can still sit inside a badly managed operating system.

AI Operations is the missing management layer between technical possibility and operating value.

4 Shared Meaning as Operating Infrastructure

Enterprise AI becomes fragile the moment every team invents its own meaning in isolation. One prompt calls a case urgent. Another calls it complex. A dashboard groups referrals one way, an operations manager groups them another, and a model retrieves context from documents that were never written to support consistent interpretation. The result is familiar: duplicated logic, brittle integrations, inconsistent outputs, and endless debates about why the system produced a particular answer.

This is why the semantic layer matters. In practical terms, it is the shared structure of business meaning underneath AI-enabled work: common definitions, relationships, context boundaries, provenance, and reusable memory. That may sound architectural, but the consequences are managerial. Siddhant Pardeshi describes a similar problem in software-agent systems: flat text memory does not scale, while graph structure acts more like a GPS for navigation across complexity. Peter McCrory makes the organizational side of the same point: capability does not translate into adoption if the relevant operating knowledge remains tacit and unavailable at the point of use.[10, 8]

The semantic layer is therefore not just a glossary for AI. In the language of AI Operations, it does four jobs at once: it helps *define* the operating domain, *contextualize* work, *control* AI behavior, and help the organization *remember* by turning production lessons into reusable operating memory.

That is also why prompt craft alone cannot scale. Without shared meaning, every new use case becomes a custom integration project plus a new round of prompt improvisation. Logic gets duplicated, exceptions get handled differently across teams, and auditability gets weaker because the system has no stable representation of what it is reasoning about.

With a semantic layer, the economics of AI deployment change. New workflows can reuse shared definitions instead of inventing them. Retrieval improves because context can be assembled around meaningful relationships. Governance gains a control surface because the same concepts can be monitored and constrained across use cases. Improvement becomes cheaper because a clarification made once can propagate across many paths. Put differently: without the semantic layer, every AI use case looks like a new project. With it, more of them begin to look like new configurations on top of shared meaning.

A practical semantic layer does not begin with enterprise metaphysics. It begins with operational questions: What entities matter in this workflow? What decisions depend on them? Which relationships change the outcome? Which systems contain the evidence? What terms must mean the same thing across teams? What exceptions require special routing? Those questions produce practical artifacts: a concept map, a relationship model, provenance rules, retrieval boundaries, and exception logic. That is why the semantic layer belongs inside the methodology, not in an architecture appendix.

The IntelliDoctor case is the clearest example I know. The core problem was not lack of clinical data. It was fragmentation. Clinical and operational meaning was spread across five disconnected systems and multiple human routines. A conventional response would have been to wait for a large

centralized data program before attempting cross-system intelligence. We took a different path: build an operational ontology that could connect diagnoses, orders, protocols, referrals, specialties, billing implications, and patient context tightly enough for workflows to reason across them.

That semantic work unlocked capabilities that would otherwise have remained isolated features. It enabled a more unified clinical narrative across records. It supported referral scoring by connecting diagnosis, specialty suitability, capacity logic, and economic relevance. It made protocol-compliance visibility possible because discharge context, protocol rules, and return events could be interpreted together rather than as separate data fragments. It did not remove the need for engineering. It changed the economics of reuse and cross-system reasoning.

That distinction matters. The semantic layer is often misunderstood in two opposite ways. One camp treats it as academic ontology work with little operating consequence. The other treats it as magic infrastructure that removes the need for integration work. Both views are wrong. A semantic layer does not eliminate data quality problems, workflow design, or engineering effort. It concentrates interpretation logic into a reusable layer so that each new AI deployment does not have to rebuild the same business meaning from scratch.

For executives, the practical consequences are straightforward:

- **Control:** shared meaning makes it easier to define what the system is allowed to infer, recommend, or trigger;
- **Reuse:** once business concepts are explicit, workflows can build on each other instead of remaining standalone tools;
- **Speed:** new use cases can be assembled faster because the organization is reusing meaning, not renegotiating it every time;
- **Institutional learning:** production lessons can be encoded once and reused many times.

In healthcare, ambiguity is expensive because it shows up as poor handoffs, weak compliance visibility, lost revenue, and patient risk. But the same logic applies well beyond hospitals. Any enterprise trying to deploy AI across multiple functions will eventually face the same problem: scale requires shared meaning. If AI Operations is the management discipline, the semantic layer is one of its core operating mechanisms.

5 Governance, Control, and Executive Readiness

Governance that lives only in policy documents will fail on contact with production. The executive question is not whether the organization has an AI policy. It is whether the organization can control AI in operation: where it is used, what it is allowed to influence, how value is measured, and who intervenes when behavior drifts.

That sounds basic, but it is where many leadership teams are weakest. They can approve pilots, negotiate licenses, and ask for responsible-AI principles. They often cannot tell the difference between technical promise and operational credibility. Davenport and Srinivasan make the point

indirectly in their 2026 research: the differentiator is not simply which AI tools a company deploy, but how deliberately the organization measures, manages, and reports the value those tools create.[4]

A useful governance standard, therefore, is not elegance of policy but quality of operational answers. Senior leaders should be able to ask a short set of questions and get clear, testable responses.

1. **What workflow is being changed?**
2. **What metric should move, and what is the baseline?**
3. **What decisions remain human-owned?**
4. **What exceptions trigger review or escalation?**
5. **What semantic and data controls constrain the system?**
6. **Who owns production performance end to end?**

These questions are intentionally unglamorous. That is their strength. Boards do not need deeper enthusiasm about models. They need better filters for separating AI theater from operational substance.

AI theater versus operational credibility

AI theater celebrates demos, vendor relationships, pilot counts, and model names. It reports activity, not control. It treats adoption as logins, not behavior change. It speaks confidently about capability while remaining vague about workflow economics.

Operational credibility looks different.

- It shows measured improvement in a defined workflow.
- It names the baseline and the failure modes.
- It makes ownership visible.
- It explains the human review boundary.
- It can trace how data and semantic rules shape system behavior.
- It has a review cadence for drift, exceptions, and policy change.

This is the language leadership teams need because AI cuts across too many silos to be governed function by function. McKinsey's 2025 survey found that organizations seeing stronger impact were more likely both to redesign workflows and to place senior leaders in meaningful governance roles.[9] Without executive sponsorship, organizations buy the software but avoid the redesign.

Control must be designed into the workflow

Real control does not begin with a policy memo. It begins when the workflow itself contains measurable checkpoints, escalation rules, and accountable interventions. A model recommendation

that no one is obligated to review is not governed. A human sign-off step that exists only to absorb anxiety is not governed either. Both are forms of ambiguity.

The stronger design principle is to place review where it changes outcomes. Low-risk, high-volume decisions may justify more automation and sampling. High-risk or low-frequency cases may require explicit human ownership. Many workflows will need tiered control, where AI handles the standard path and people focus on exceptions, disputes, and borderline cases. Governance becomes real when those choices are visible in the work itself.

Executive readiness is a capital-allocation decision

Many executives say they want AI value while funding only the most visible layer of the stack. They approve licenses, experiments, and innovation budgets. They underfund instrumentation, semantic assets, workflow redesign, user training, and operating review. Then they are surprised when the program remains stuck between enthusiasm and proof. That gap is not hypothetical. Accenture's 2025 research found that executive conviction about generative AI is extremely high even while a large share of leaders admit they do not yet have the expertise to lead these transformations well.[3]

Executive readiness means funding the less glamorous parts of deployment. It means accepting that value realization may depend less on another model upgrade than on better telemetry, clearer ownership, and more disciplined process change. It means asking for before-and-after evidence, not narrative confidence.

It also means restoring a role that seasoned operators will recognize immediately: the champion. Every serious AI deployment needs one. Not just someone to approve budget, but someone senior enough to remove blockers across silos, insist on measurement discipline, force workflow redesign, and make behavior change part of the operating plan rather than an afterthought. That is old transformation logic, but it is still correct.

The IntelliDoctor experience was instructive here too. The barrier was not simply whether the AI could produce useful outputs. It was whether leadership had a robust language for evaluating a working system across adoption, financial impact, protocol control, and workflow change. That is why AI Operations matters at the executive level.

A simple standard follows from this. If a board or executive team cannot answer where AI is acting, what outcome it is meant to change, how the system is constrained, and who is accountable for production behavior, then it is not governing AI yet. It is still funding experimentation and hoping that value compounds.

6 Conclusion

The organizations that win with AI will not be the ones with the most demos. They will be the ones with the best operating discipline.

In most enterprises, AI value is constrained less by model weakness than by operational weakness: unclear baselines, unchanged workflows, missing semantic structure, weak adoption design, and governance detached from production. AI Operations is the management layer that brings those elements together.

The next move does not need to be grand. It does need to be serious. Choose one high-friction workflow. Baseline it. Define the key terms and control points. Build the semantic layer just far enough to make the workflow legible. Redesign the workflow around AI plus human judgment. Measure what changes. Then improve it in production.

Over time, that work should become a practical method, not a one-off intervention: a playbook, a semantic canvas, a baseline scorecard, a control design, and supporting tools that make the next deployment easier than the last.

That is the real executive test. Not whether the organization is "doing AI," but whether it is building the discipline required to turn AI capability into durable operating value.

References

[1] Anthropic. Anthropic economic index report: Learning curves. https://www.anthropic.com/research/economic-index-march-2026-report, March 2026. Primary research report on Claude usage patterns.

[2] Boston Consulting Group. The widening AI value gap. https://www.bcg.com/publications/2025/are-you-generating-value-from-ai-the-widening-gap, 2025. BCG study on enterprise AI value realization.

[3] Karalee Close and Kestas Sereiva. Reinventing enterprise models in the age of generative AI. https://www.accenture.com/us-en/insights/consulting/gen-ai-reinventing-enterprise-models, March 2025. Accenture research report.

[4] Thomas H. Davenport and Laks Srinivasan. 7 factors that drive returns on AI investments, according to a new survey. *Harvard Business Review*, March 2026. Based on a survey of senior executives and enterprise leader interviews.

[5] Eric Ethington. Automation, AI, and the risk of doing things efficiently that should not be done at all. https://www.lean.org/the-lean-post/articles/automation-ai-and-the-risk-of-doing-things-efficiently-that-should-not-be-done-at-all/, March 2026. Lean Enterprise Institute.

[6] Matthias Holweg, Thomas H. Davenport, and Ken Snyder. How AI fits into lean six sigma. *Harvard Business Review*, November 2023.

[7] Karim R. Lakhani, Jared Spataro, and Jen Stave. The "last mile" problem slowing AI transformation. *Harvard Business Review*, March 2026. Seven-friction framework for AI scaling.

[8] Peter McCrory and Azeem Azhar. Anthropic's head of economics on AI adoption data, claude code, the burden of knowledge, and the next generation of experts. https://www.exponentialview.co/p/anthropics-head-of-economics-on-ai, 2026. Exponential View interview.

[9] McKinsey & Company. The state of AI: Global survey 2025. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai, 2025. McKinsey Global Survey on AI.

[10] Siddhant Pardeshi and Sam Charrington. Agent swarms and knowledge graphs for autonomous software development with siddhant pardeshi. https://twimlai.com/podcast/twimlai/agent-swarms-knowledge-graphs-autonomous-software-development, March 2026. The TWIML AI Podcast, episode 763.

[11] Tom Peperkamp. Lean in the age of AI: Optimizing the stream, not the coder. https://www.lean.org/the-lean-post/articles/lean-in-the-age-of-ai-optimizing-the-stream-not-the-coder/, February 2026. Lean Enterprise Institute.

[12] Nigel Vaz and Amy Bernstein. Strategy summit 2026: Why AI transformation needs a human touch. https://hbr.org/podcast/2026/03/

strategy-summit-2026-why-ai-transformation-needs-a-human-touch, March 2026. HBR IdeaCast podcast episode.