The Enterprise AI Playbook

Lessons from 51 Successful Deployments

Elisa Pereira, Alvin Wang Graylin and Erik Brynjolfsson
Stanford Digital Economy Lab
Stanford University · April 2026

Foreword

There is no shortage of predictions and sentiment surveys about artificial intelligence today.

Every week brings new forecasts and debates about whether AI is useful, which jobs will disappear, which industries will transform, which companies will dominate. But when we speak with executives actually deploying AI inside their organizations, we hear a different set of questions. Not what might happen in five years, but what is happening right now. Practical realities, not abstract frameworks.

This report was born from a simple conviction: the most valuable insights about AI adoption are not in hypotheticals or predictions. They are in the patterns of those who have already walked the path.

We set out to build something empirical. To document real-world use cases that have actually delivered business value. To map the practices of organizations that are not just experimenting with AI but successfully deploying it at scale. We wanted depth. To understand the pitfalls that do not make it into press releases, the nuances that separate a successful pilot from a failed one, and the organizational realities that no vendor whitepaper will tell you.

Across 51 enterprise cases over 5 months, we found stories of transformation measured in weeks and others measured in years. Same technology, same use cases, vastly different outcomes. The difference was never the AI model. It was always the organization. Its readiness, its processes, its leadership, its willingness to change and fail.

Our ambition with this research is simple: to offer a practical window into what is actually happening inside companies as they create value with AI, including detailed company case studies. The future of work only makes sense when one first understands the present of work.

In the conclusion, we offer some forward-looking insights based on upcoming trends in the AI space. We hope these findings serve as both a mirror and a map. Reflecting where your organization might be and illuminating the paths on how you can move forward with confidence.

Elisa Pereira, Alvin Wang Graylin & Erik Brynjolfsson

The Research Team

Stanford Digital Economy Lab

Contributors

Elisa Pereira

Researcher, Stanford Digital Economy Lab · MSx Candidate, Stanford Graduate School of Business

Elisa Pereira is a researcher at Stanford's Digital Economy Lab and MSx candidate at the Stanford Graduate School of Business, with a background in venture capital and hands-on experience building dozens of enterprise AI solutions across Latin America. Her current research focuses on measuring the real-world impact of these deployments, identifying patterns behind successful implementations, and exploring how Latin America can establish technological sovereignty.

Alvin Wang Graylin

Digital Fellow, Stanford Digital Economy Lab · Stanford University

Alvin Wang Graylin is Digital Fellow at the Stanford Digital Economy Lab, and an author, serial entrepreneur and technology executive with over 35 years of experience in AI, XR, cybersecurity and semiconductor industries. He's currently the chairman of the Virtual World Society, Senior Fellow at the Asia Society Policy Institute CCA, lecturer at MIT and advises governments, organizations and corporations on technology transitions. His book, *Our Next Reality*, discusses how AI and immersive technology will reshape our world in the coming decade. His current research is focused on the economics of AI and the associated governmental policies needed to ensure a smooth transition to a post-labor economic model.

Erik Brynjolfsson

Director, Stanford Digital Economy Lab · Professor, Stanford University

Erik Brynjolfsson is the Director of the Stanford Digital Economy Lab and the Jerry Yang and Akiko Yamazaki Professor and Senior Fellow at the Stanford Institute for Human-Centered AI (HAI). He is also the Ralph Landau Senior Fellow at SIEPR, professor by courtesy at the Stanford Graduate School of Business and Department of Economics, and a research associate at the National Bureau of Economic Research (NBER). One of the most-cited authors on the economics of information, he has co-authored hundreds of articles and books, including The Second Machine Age and Machine, Platform, Crowd. He puts his academic insights to practical use via Workhelix, a company he co-founded to identify and measure the benefits of AI

Contents

The Macro Context

Why enterprise AI implementation matters now

General purpose technologies like AI enable and require significant complementary investments in process redesign, workforce development, and organizational restructuring. These investments are largely intangible and poorly measured in national accounts, which means productivity growth is systematically underestimated in the early years of a new technology and overestimated later, when the benefits are harvested. Brynjolfsson, Rock, and Syverson (2021) formalized these observations in a model called the "*Productivity J-Curve*".[1]

The macroeconomic outcome hinges not on the technology itself but on how organizations deploy it. We face a "productivity fork": AI can either augment workers and create new capabilities or primarily automate existing tasks and cut headcount. The path chosen will shape economic growth for decades.[2] In particular, automation displaces workers from existing tasks, but the creation of new tasks in which humans have a comparative advantage can reinstate labor demand. Whether AI leads to broad prosperity or concentrated gains depends on whether organizations generate enough new opportunities to offset labor displacement.[3]

Some employment effects are already surfacing. Analysis of high-frequency payroll data covering millions of U.S. workers finds that early-career workers in AI-exposed occupations have experienced a 16% relative decline in employment, with software developers aged 22 to 25 seeing a nearly 20% drop.[5] These "canaries in the coal mine" suggest that some of the labor market disruptions many anticipated are no longer hypothetical.

These measurement challenges are not merely academic. Standard metrics like GDP systematically fail to capture the welfare contributions of new and free digital goods. Their GDP-B framework, which measures consumer benefits rather than production costs, reveals substantial unmeasured value creation in the digital economy. If aggregate statistics undercount the gains from relatively simple digital services, they are likely to miss even more of the value that AI creates inside organizations—precisely the kind of value this report attempts to document.[6] One new non-monetary *benefit* that AI agent systems are delivering to software developers is "free time" to *think*. While AI agents autonomously build increasingly larger portions of the code, human coders are allotted more coffee breaks to ponder bigger picture issues. This won't show up in standard productivity measurements, but it is a real benefit that changes their daily work for the better.

The gap between these macro findings and what happens inside organizations is significant. The economic models describe aggregate effects. The firm-level experiments measure more controlled settings. Neither captures the messy operational reality of deploying AI across departments, overcoming resistance, and building the complementary infrastructure that the J-curve framework identifies as essential. This is the gap our research addresses.

Why this research

Despite billions in enterprise AI spending, a 2025 study from MIT's NANDA initiative concluded that 95% of generative AI pilot programs fail to produce measurable financial impact.[7] They argued that the failures stem not from model quality but from poor workflow integration and misaligned organizational incentives. This is the gap between what technology can do and what organizations manage to do with it.

In contrast, our objective was to understand the cases where AI was deployed successfully. We dove deep into companies and analyzed 51 cases where enterprise AI delivered measurable value. What did these organizations do differently? What did integration actually cost? Where did resistance come from, and most importantly, how was it overcome?

How we incorporated failure

While this report focuses on implementations that delivered measurable value, we did not study success in isolation. In every interview, we explicitly asked participants to describe the failures, false starts, and abandoned pilots that preceded their current results. We asked what they tried first, why it did not work, and what they changed.

What emerged is not a story of organizations that avoided difficulty. It is a story of organizations that failed iteratively and built systematic approaches to overcome initial setback. Two thirds of the companies we investigated had significant failed attempts prior to achieving value creation. The patterns in this report reflect what these organizations learned from the process as much as what they achieved through success.

We want to be transparent that this does carry a known limitation: selection bias toward positive outcomes. Our findings describe what success looks like and what it took to get there; we don't claim to provide representative data on how common success is across the broader economy.

"All happy families are alike; each unhappy family is unhappy in its own way." – Leo Tolstoy

Methodology

This research is based on in-depth interviews with executives and project leaders who have deployed AI solutions at scale. We focused exclusively on initiatives that have moved beyond pilot stage and are delivering measurable business value.

Research Sample Profile

Our 51 case studies draw from 41 organizations, 7 countries, 5 regions, representing over a million employees. (full list of anonymized companies in the appendix).

Figure 1. Research interview and analysis workflow

Selection Criteria

Four dimensions define the mature AI projects we selected for analysis:

- **Operational Stability**
  System is live, integrated into real workflows, and consistently used in production.
- **Sustained Business Adoption**
  Teams across functions actively rely on the AI system for decision-making over months (3+ months).
- **Quantified Value Creation**
  Clear business outcomes such as productivity gains, revenue growth, or customer satisfaction.
- **Scalability & Replicability**
  Can be extended or replicated across teams, geographies, or business units.

Technologies range from data science models (machine learning, deep learning) to agentic workflows.

Interview Approach

Each case study was developed through at least one structured 60-minute interview per company following a consistent discussion framework. We supplemented interview data with written documentation provided by participant companies, including internal metrics/reports, project plans/reviews, and financial updates. Interviews were conducted between Aug. '25 and Feb. '26.

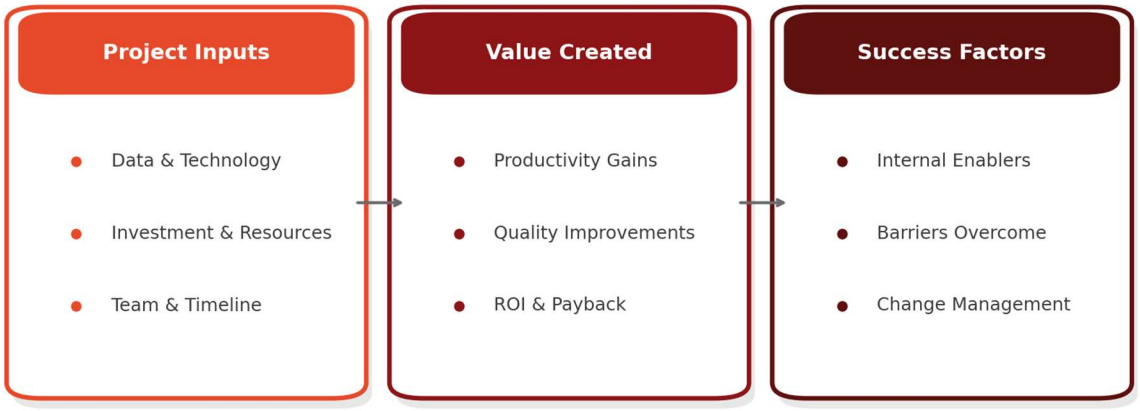


Figure 2. Interview approach and supplementary data sources

Scoring Criteria

Each dimension was scored based on documented evidence:
3 = Strong (all criteria met), 2 = Moderate (most criteria met), 1 = Weak (few criteria met).
We required evidence from systems, documentation, or named owners.

Sample Composition

In terms of business functions, our cases cover a wide range of applications. This diversity allows us to identify patterns that transcend specific use cases.

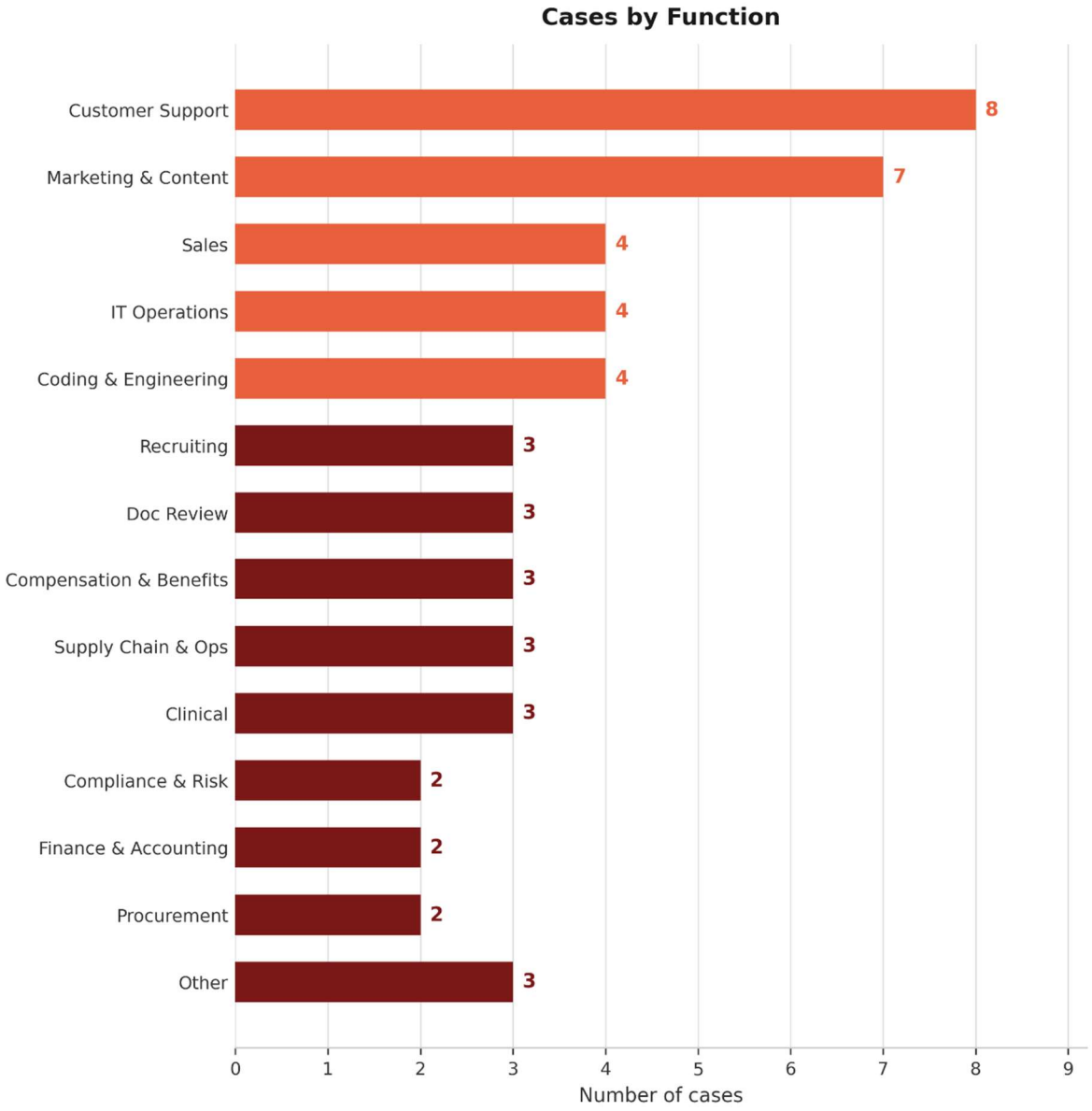


Figure 3. Sample composition by business function

Our sample spans 9 industries, with particular depth in manufacturing, financial services, and technology. The distribution reflects the current landscape of enterprise AI adoption.

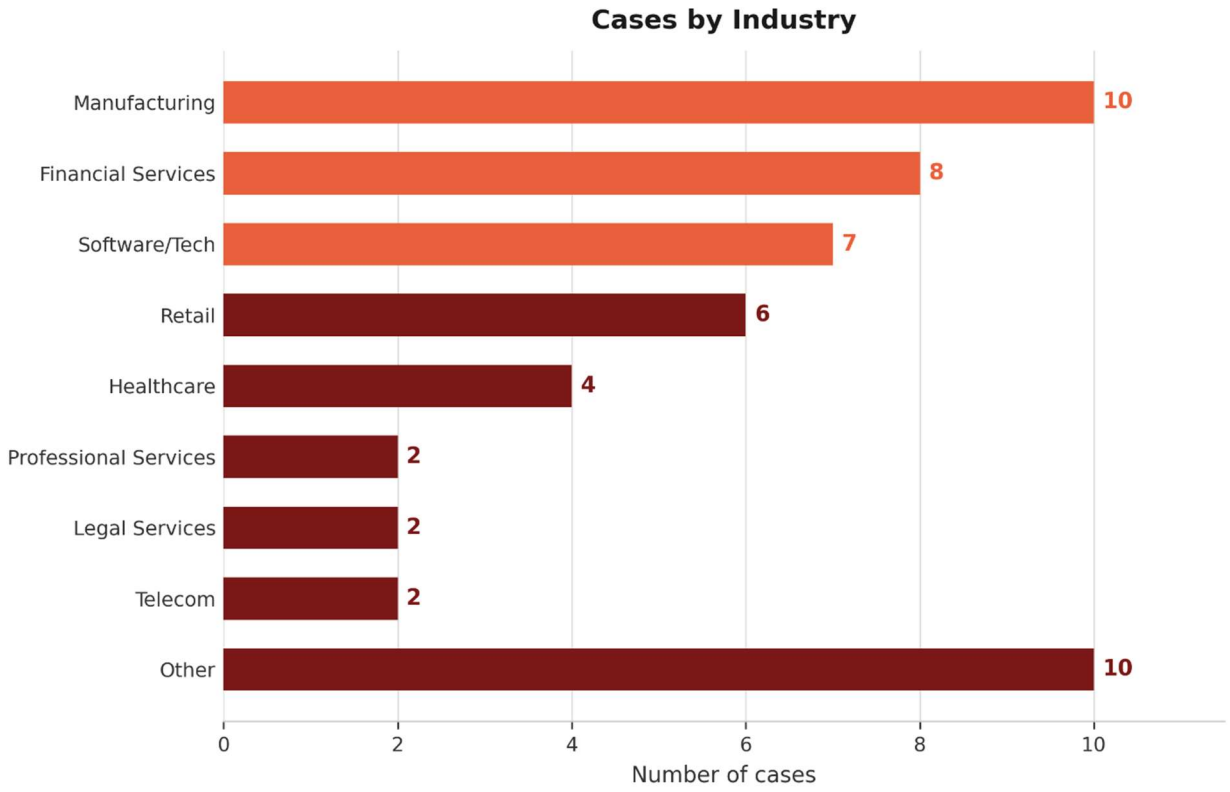


Figure 4. Sample composition by industry

Limitations

This research relies primarily on self-reported data from interview participants. While we triangulated information where possible and focused on mature initiatives with documented outcomes, readers should consider potential selection bias toward successful deployments.

Our sample, while diverse, is not representative of all enterprise AI initiatives. The concentration in technology and financial services reflects early adoption patterns rather than the full universe of AI deployment.

All data was anonymized and aggregated to protect proprietary information and follow subject company disclosure policies. Specific company names and identifying details have been removed or generalized.

Key Findings Summary

1. Technology is not the hardest part. 77% of the hardest challenges were invisible and intangible costs: change management, data quality, and process redesign. 61% of successful projects included at least one prior failure, whose costs never appear in the final ROI.

2. Timeline variance is organizational, not technical. Similar use cases took weeks at one company and years at another. The difference was executive sponsorship, existing organizational processes, and end user willingness.

3. Escalation-based models were associated with better results. Escalation-based models (AI handles 80%+ autonomously, humans review exceptions) delivered 71% median productivity gains versus 30% for approval models. This may, in part, reflect different types of tasks addressed.

4. Executive sponsorship is about actions, not approval. Effective sponsors clear blockers weekly, bridge business and technical teams, and tie AI adoption to corporate OKRs. Most critically, they create a culture that gives permission to fail.

5. Staff functions are the most frequent source of resistance, but some parts may become enablers after buy-in. Legal, HR, Risk, and Compliance were the most frequent source of resistance at 35%, ahead of internal end-users at 23%.

6. Headcount reduction is common but not inevitable. Headcount reduction was the largest outcome in 45% of the deployments, but alternatives (hiring avoided, redeployment, no reduction) accounted for 55%. Broader labor market data suggests entry level roles in AI exposed occupations are already declining.

7. Revenue from AI is real, but still rare, and follows three patterns. Personalization that converts, speed that wins deals, and internal tools repackaged as products. A small subset of cases also shows AI enabling work that was previously impossible.

8. Agentic AI works, but most firms have not used it, yet. Agentic implementations showed 71% median productivity gains versus 40% for high-automation but represented only 20% of cases. Agentic AI isn't a new UI; it's a redefinition of the role of humans and machines in the workflow.

9. Messy data is not a blocker if you design around it. LLMs fixed many of the data problems they were supposed to struggle with. Store everything, connect it, and let the models do the cleaning.

10. Security enables more than it blocks. Security was not a project killer in any of the cases we studied. Requirements that were initially barriers later enabled projects to handle sensitive data.

11. Model choice is a commodity for many use cases. For 42% of implementations, model choice was fully interchangeable. Companies don't always need the best available AI models. The durable advantage is in the orchestration layer, not the foundation model.

Chapter 1

Why do AI business cases underestimate real investment?

The hidden costs that determine success or failure

Published Findings

Scaling requires heavy non-model investment. McKinsey's research identifies that high-performing AI organizations (those attributing >5% of earnings before interest and taxes [EBIT] to AI) are significantly more likely to invest in "rewiring" business processes and data products rather than just model deployment.[8]

"Proof of Concept Factories" represent sunk costs. Accenture estimates that 80-85% of companies are stuck in a "Proof of Concept Factory" stage, where they conduct experiments but achieve low returns and low scaling success rates.[9]

Data foundations are a major line item. Strategic scalers are far more likely to possess a large, accurate data set (61% vs 38% for non-scalers) and invest heavily in data quality, management, and governance frameworks.

The Productivity J-Curve implies hidden investment. Earlier research found that for every $1 of tangible tech investment, companies spend up to $10 on intangibles (process redesign, reskilling, organizational transformation), initially depressing productivity before gains are realized.[1]

What We Found

77% of the hardest challenges practitioners faced were invisible costs: change management, data quality, and process redesign, not technical issues. Technology was consistently described as the easiest part. The true cost of a successful deployment usually includes at least one failed attempt (see Finding 2), and the bulk of investment goes to everything except the model.

Finding 1

77% of the hardest challenges are "invisible costs"

When we asked practitioners, "what was the hardest thing to fix?", the answers reveal where AI budgets actually go.

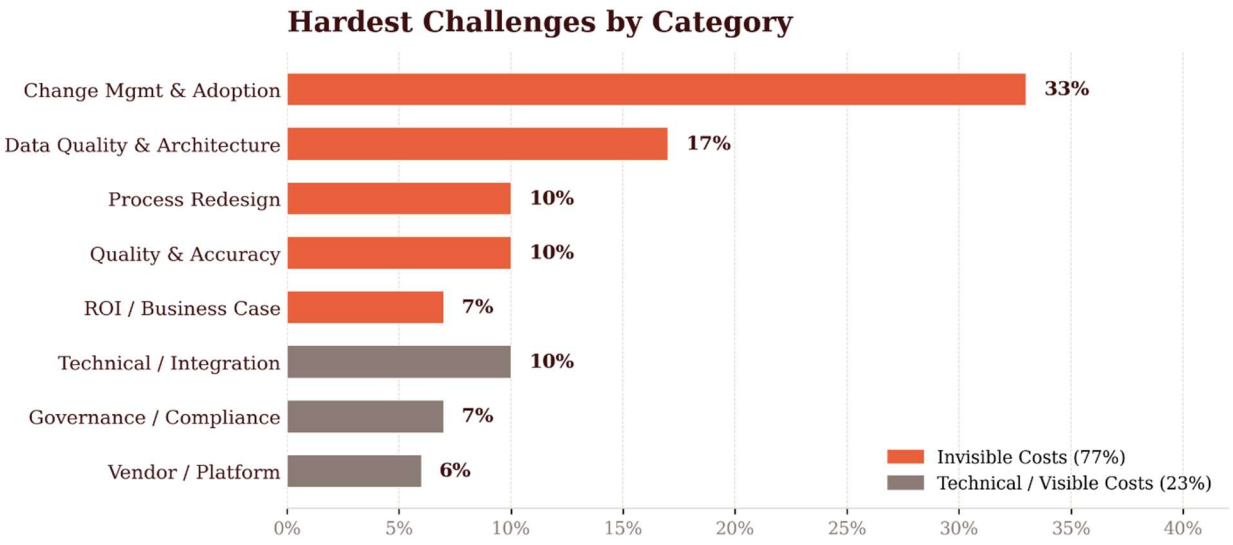


Figure 5. Hardest challenges in AI implementation

> *"All the hard work is in process documentation and data architecture. If you can do those two things, everything else is quite simple."*
>
> - Executive, Telecom Company

> *" Technology wasn't the bottleneck - organizational adoption was the failure point."*
>
> - Executive, Professional Services Company

Finding 2

61% had a failed AI project before their current success

These failed experiments represent sunk costs that never appear in the "successful" project's ROI:

Had previous AI failure(s)	**61%**
No previous failures	39%

These failed experiments are sunk costs that may never appear in the successful project's ROI but were often essential to it. The failures share a pattern: teams treated AI as a technology project instead of a process and change management project. First attempts failed when applied to broken workflows, when led by technical teams without business ownership, or when organizations assumed the model would fix problems that required redesigning the work itself.

> *"This was actually the second time they looked to AI for the recruiting process. It failed initially because they didn't account for bias, and they thought AI would just fix processes instead of requiring process redesign."*
>
> - AI Project Lead, Professional Services Company

The technology was consistently described as the easiest part.

> *"The more you invest in your data, the better you can get out of these AI solutions."*
>
> - Manager, Technology Company

> *"The problem isn't the models."*
>
> - Executive, Professional Services Company

The implication for budgeting: the true cost of a successful AI deployment usually includes at least one failed attempt, and the bulk of the investment goes to everything except the model.

CASE STUDY

Invoice Processing at a Logistics Company

How they overcame invisible costs

The Company

A $1B+ US-based logistics company managing a large fleet of refrigerated trailers. The company receives +100k invoices annually from vendors across the country performing maintenance on trailers - everything from tire changes to sensor replacements.

The Challenge

The volume and variation of invoices created a significant operational burden. Seven full-time employees were dedicated exclusively to this task: consolidating invoices, matching them to internal templates, validating the work, entering data into the enterprise resource planning (ERP) system, and generating client invoices.

> *"They get all these invoices in different channels, including fax. They might get phone calls. A lot of these repair shops, middle of nowhere, they just dial in and say, hey, we did this repair. So they might be phone calls, they might be emails, they might be all types of ways that they get this information."*
>
> - Senior Executive, Technology Services Company

The Invisible Work

Process Simplification: thousands of templates reduced to hundreds. Years of accumulated invoice templates were redundant and inconsistent. This cleanup was required before any AI could work.

> *"We very quickly realized that the 750 templates don't make any sense and most of them are repetitive. Nobody really did a review on this."*
>
> - Senior Executive, Technology Services Company

Data Annotation: Subject matter experts (SMEs) reviewed thousands of AI outputs. They validated AI-generated invoices on top of their daily work, explaining every mistake to improve the model.

Executive Sponsorship: President involved in weekly check-ins. This removed bottlenecks and ensured buy-in from the operations team.

> *"The president was checking in every week - what is the progress, where are we, what are the bottlenecks? Then the rest of the team also engaged."*
>
> - Senior Executive, Technology Services Company

Knowledge Transfer: Two junior IT staff embedded from day one. Daily stand-ups, weekly and monthly reviews. No black box - the company could operate the system independently.

The Solution

In simple terms, the company built a system that automatically reads invoices regardless of how they arrive, understands their content, and enters the data directly into the company's financial system, eliminating the need for manual processing.

The technical implementation used Azure Document Intelligence with Azure OpenAI Service, combining optical character recognition (OCR) parsing with large language model (LLM)-based semantic mapping. The system ingests invoices from multiple channels (email, fax, phone transcriptions), parses and extracts data using OCR, maps invoice content to the simplified template taxonomy, and writes validated data directly to MS Dynamics D365.

The Results

Headcount

7 → 2 full-time equivalents (FTEs)

Accuracy

85%

Processing time

< 24 hours

Time to production

8 weeks

Value created

> $1M

Key Lessons

> *"It always starts with the people. There are people, process, and technology - and I know it's in that order even though I'm representing a technology company. The technology was the easiest part. We basically used a lot of open-source and off the shelf stuff."*
>
> - Senior Executive, Technology Services Company

> *"Look guys, 80% is perfect for us. We can take these folks, we can just put them in the other bottleneck. I understand that you can keep improving and at one point the model is going to be 95%, but we don't care. What we care is immediate cost saving and getting rid of these backlogs."*
>
> - President, Logistics Company

Chapter 2

How to cross the valley of death between deployment and ROI?

What separates weeks from years in similar use cases

Published Findings

Intentional timelines beat "move fast." Accenture concludes that successful AI scalers are 65% more likely to set 1–2-year timelines to move from pilot to scale. Contrary to the "move fast" ethos, they are more intentional about the time required to scale responsibly.[9]

High performers redesign workflows, not just deploy tools. McKinsey reports that top performers are nearly three times more likely to fundamentally redesign workflows as part of their AI efforts. 55% of high performers redesigned workflows around AI versus only 20% of other companies.[11]

Most companies are stuck in pilot mode. While 88% of organizations use AI in at least one function, only one-third have begun to scale their AI programs at the enterprise level. Two-thirds remain in testing or proof of concept phase.

What We Found

Similar use cases can take weeks or years depending on the organization. We identified three factors that consistently accelerate projects - executive sponsorship, existing foundations, and end-user willingness - and four that slow them down. Every successful project in our sample used an iterative approach.

Finding 1

The range is dramatic: from weeks to years for similar use cases

A large fintech used an AI coding agent to migrate millions of lines of legacy extract, transform, load (ETL) code to a modern architecture. The project took weeks. A technology company redesigned their customer support system with AI and launched in six months. A major bank attempting the same customer support use case reports that projects take multiple years.

> *"Within weeks of the AI agent's launch, we identified a clear opportunity to accelerate the migration at a fraction of the engineering hours."*
>
> - Executive, Fintech

> *"It takes us multiple years just to even stand one of these things up."*
>
> - Executive, Financial Services

The same use case, the same AI models, vastly different timelines. The insight here is not a median or average. It is that **organizational context matters more than the technology itself.**

Finding 2

Three factors consistently accelerate time to value

Acceleration Factor	Frequency
Executive Sponsorship	**43%**
Building on Existing Foundation	32%
End User Willingness	25%

Building on Existing Foundation. Projects that leveraged existing infrastructure or platforms moved significantly faster. One technology company built their sales copilot in months because they had already developed an AI platform for customer support.

> *"We launched the first MVP [minimum viable product] in April. Because we finished the customer support project early, we went on to build this one."*
>
> - Executive, Technology Company

End User Willingness. When users genuinely want the solution, adoption friction disappears. In healthcare, hospital systems adopted ambient AI transcription despite unclear ROI simply because physicians were desperate for relief. With existing processes, after a full day of work, they were forced to spend hours documenting their daily activities.

> *"The state of current medical practice is so bad, and the doctors were so burnt out that the hospital systems were willing to try anything as a Hail Mary just to see if it made a difference."*
>
> - Executive, Healthcare AI Company

Finding 3

Four factors consistently slow projects down

Delay Factor	Frequency
Learning Curve and Iteration	25%
Data Quality and Preparation	21%
Regulatory and Compliance	21%
Process Documentation Gaps	21%

Data quality was a recurring theme. Regulatory constraints created structural delays in financial services, where compliance requirements extend timelines regardless of technical readiness.

> *"Majority of customers don't do a good job maintaining their knowledge bases."*
>
> - Executive, Software Company

Finding 4

Every successful project used an iterative approach

100%

used iterative approach

Of cases where we could identify the development methodology, all used an iterative approach. None used traditional waterfall planning. The pattern was consistent: start small, learn, expand.

> *"Think of it as like a layered cake. We built one process, documented it, then built that layer of the agent, then the second feature and the third feature on top of it."*
>
> - Executive, Logistics Company

> *"Probably 90% of the pilots and tests fail, but then we iterate on those until we find them and it grows and grows."*
>
> - Executive, Food Delivery Company

CASE STUDY

Recruiting at a Translation Services Company

How they crossed the valley of death

Professional Services | Recruiting | Mid Market

The Company

Their recruiting process had become their biggest cost sink and a strategic bottleneck to business scalability: slow candidate intake, high turnover, difficulty staffing niche languages and dialects, and inconsistent screening quality limited how fast the company could grow.

The First Attempt - and Why It Failed

This was the company's second attempt at AI for recruiting. The first failed for two reasons: they did not account for bias in their screening algorithms, and they assumed AI would fix broken processes without addressing the underlying workflow problems.

> *"They thought AI would just fix processes instead of also stepping back and making sure everything was working as expected."*
>
> - Executive, Professional Services

What Changed the Second Time

Three things were different:

First, the CEO took ownership rather than delegating to the CTO. The project had executive visibility and weekly check-ins that cleared bottlenecks quickly.

Second, they fixed the process before applying AI. They mapped the entire recruiting workflow and identified where the real pain points were.

Third, they targeted genuine pain. The recruiters were not mildly inconvenienced. They were burdened by a stream of applications that just overwhelmed the team each day, and it kept compounding.

> *"This was a painkiller for those guys. It wasn't 'Hey, this would be great.' It was 'I'm drowning.'"*
>
> - Executive, Professional Services

The Solution

The team built an AI powered recruiting pipeline with hyper-personalized screening by language and dialect, automated first-round video interviews with bias-mitigated evaluation, and a feedback loop connecting hiring outcomes back to screening criteria. The system learned which candidate signals predicted success.

The Results

Time to build **~1 month**	Time per role **3 hrs → 3 min**
Intake efficiency **+83%**	Screening efficiency **+79%**
Candidate conversion **+75%**	

Key Lessons

The same company, the same function, the same goal - but radically different outcomes. The first attempt failed. The second took one month and delivered 83% efficiency gains. The difference was not the technology.

Fix the process before applying AI. AI amplifies whatever process it is applied to. If the process is broken, AI makes it worse faster.

Target real pain. Adoption was easy because users were desperate for relief. The recruiting team did not need to be convinced. They needed to be rescued.

Chapter 3

How much human oversight is optimal?

Examining human involvement across AI implementations

Published Findings

Enterprise and individual usage patterns differ. Anthropic's Economic Index finds that 52% of individual Claude.ai usage involves human AI collaboration versus 45% full automation. Enterprise application programming interface (API) usage shows the inverse: 77% automation. This suggests enterprises deploy AI differently than individuals, but the optimal balance remains unclear.[10]

Structured human oversight correlates with success. McKinsey reports that 65% of AI high performers have defined human-in-the-loop processes to determine how and when model outputs need human validation, versus only 23% of other organizations — nearly a threefold difference.[11]

What We Found

Escalation-based operating models (AI handles 80%+ autonomously, humans review only exceptions) delivered the highest productivity gains with a median of 71%. This partly reflects task selection: the escalation model is typically applied to high volume, recoverable tasks, while approval and collaboration models serve regulated or high stakes work. The level of human oversight depends on error tolerance, regulatory requirements, and task complexity, and is often a strategic design choice rather than a limitation.

Finding 1

Moderate human oversight is associated with the highest productivity gains

We classified each case on a three-point scale based on the level of human involvement:

Human-in-the-Loop (HITL) Level	Description
Escalation	AI handles 80%+ autonomously; humans review only exceptions or sample ≤20%
Approval	AI does the work; human reviews and approves every output before action
Collaboration	Human and AI work together continuously on each task

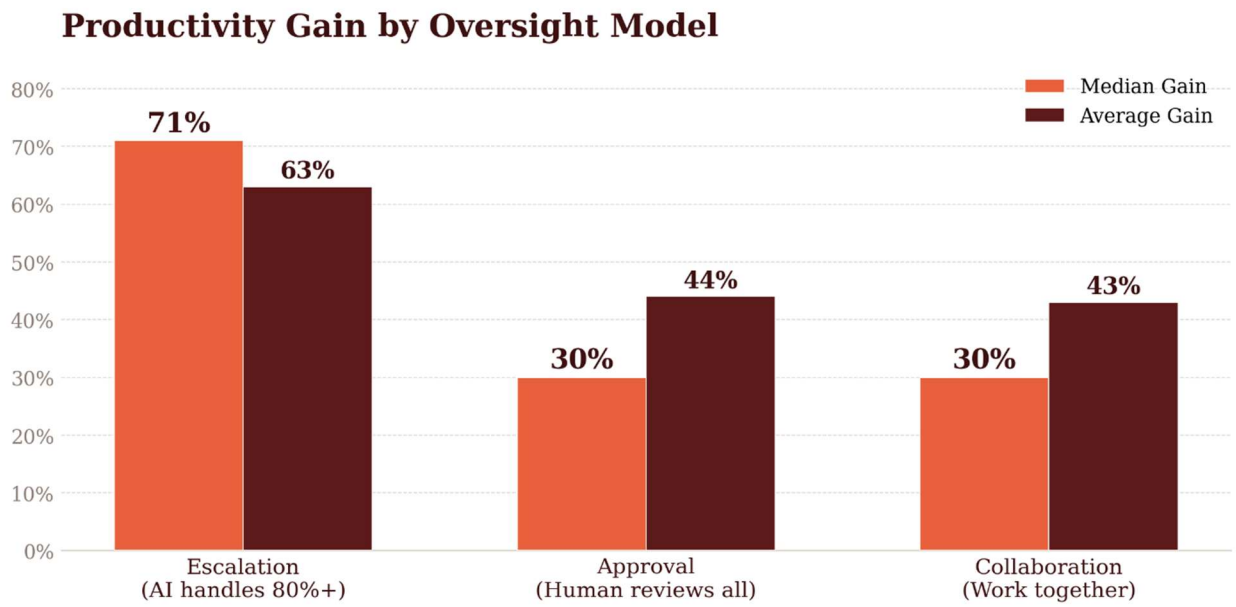


Figure 6. Human oversight models across AI deployments

> *"90 or 95% are now fully automated by an agent. If someone says their food didn't arrive or something went wrong with their order, 90 to 95% of those are completely automated."*
>
> - Head of AI, Food Delivery Company

Finding 2

The optimal oversight level varies by function

The appropriate level of human oversight (*HITL – human in the loop*) depends on error tolerance, regulatory requirements, and task complexity.

Function	Typical HITL Level	Avg. Gain
IT Operations	Escalation	90%
Customer Support	Escalation	71%
Claims Processing	Escalation	50%
Field Service	Approval	80%
Clinical Documentation	Approval	66%
Coding	Collaboration	54%

In customer support, a technology company achieved 82% ticket deflection by redesigning workflows around AI first resolution. In clinical documentation, physicians must approve every AI generated note because these are legal documents. In coding, engineers shifted from writing code to reviewing AI generated changes.

> *"Rather than engineers completing an entire migration task, they could just review the changes, make minor adjustments, then merge their PR [pull request]."*
>
> - Head of Engineering, Latin American Fintech

Finding 3

When human oversight is the obvious choice

Human oversight is not a sign of AI immaturity. In many contexts, it is the strategically correct design choice. Four patterns emerged where human involvement creates clear value:

Zero error tolerance. When a single mistake costs more than thousands of correct outputs, human review is essential. Marketing content for major brands, legal documents, and customer facing communications fall into this category.

> *"I cannot run a campaign with an error. I cannot run a large campaign that will reach millions of customers with uncertainty."*
>
> - Head of Strategy, Enterprise AI Company

Regulatory requirements. In healthcare, finance, and other regulated industries, human review is legally mandated regardless of AI capability. The question is not whether AI can do the work, but whether regulators will accept AI doing the work.

> *"The doctor reviews it, approves it, and then it gets sent back to the EMR [electronic medical record]. Doctors must still review every note due to legal requirements."*
>
> - Executive, Healthcare AI Company

Enterprise risk management. Large organizations prefer human-in-the-loop solutions even when full automation is technically feasible. The perceived risk of autonomous AI outweighs the efficiency gains.

Continuous improvement. Human reviewers identify patterns in AI errors that feed back into model improvement. This feedback loop accelerates learning in ways that fully automated systems cannot match.

CASE STUDY

Marketing Content at a Financial Services Company

How they calibrated human oversight

Financial Services | Marketing | Enterprise

The Company

A financial services company faced a content bottleneck. They had customer data enabling hyper personalization but could not generate content fast enough to leverage it. Traditional agency workflows took seven weeks per campaign.

The Solution

They deployed an AI platform that generates multi-channel content while maintaining brand consistency. The team chose an 80/20 model: AI handles 80% of the generation, humans provided 20% refinement and quality assurance. As the technology matures and learnings from experience grows, the percentage offloaded to AI will eventually go towards 100%.

How Human Oversight Enabled Success

This split was deliberate. Enterprise marketing cannot tolerate errors on customer facing content.

> *"To run at the enterprise level, you need 80% technology and 20% humans refining. The AI industry has not yet reached the level where you can nail that final 20%."*
>
> - Head of Strategy, Enterprise AI Company

The human layer served three functions:

Brand protection against errors that would damage years of brand building.

Edge case handling for unusual combinations requiring judgment.

Feedback loop where reviewers identify patterns that improve AI outputs.

The Results

Time to market

7 weeks → 6 hours

Click through rate

2x improvement

Production efficiency

>80% reduction in time

Key Lessons

Human oversight is not a tax on productivity. The 80/20 model delivered 97.6% reduction in time to market while maintaining zero error tolerance.

The oversight level should match the stakes. The company views the 20% human component as transitional, expecting to reduce it as AI improves, but they started with what worked rather than waiting for *perfect* automation to arrive.

Chapter 4

What separates sponsors who drive results from those who just approve budgets?

The activities that define effective executive sponsorship

Published Findings

High correlation with performance. McKinsey reports that AI high performers are 3.0x more likely to agree that senior leaders demonstrate ownership and commitment to AI initiatives.[11]

Champion profile matters. Accenture finds that Strategic Scalers are typically championed by a Chief AI, Data, or Analytics Officer, while struggling firms rely on a lone champion within technology.[9]

Intentionality over presence. Scalers drive AI anchored in C-suite objectives; proof of concept factories lack connection to strategic imperatives.

The above research establishes correlation but does not address what sponsors actually do that makes the difference.

What We Found

Active Steering (weekly check-ins, proactive blocker removal) is the most common pattern among successful projects. But the seven cases that achieved organization-wide transformation all reached Strategic Integration: the sponsor made AI adoption a corporate Objective and Key Result (OKR) tied to bonuses, not just a project to support. When we looked beyond what sponsors did to how they led, a consistent pattern emerged: the most effective sponsors created conditions where teams could fail, learn, and try again without career consequences. The key wasn't the executive themselves, but that they created a corporate **culture** that encouraged experimentation, demanded collaboration, designed in accountability and nurtured a *safe* environment where initiative was not punished.

Finding 1

Active Steering is common, but Strategic Integration drives transformation

We classified sponsor engagement on a four-point scale:

Level	What It Means
1 Passive Approval	Approved budget, delegated entirely, little ongoing involvement
2 Periodic Oversight	Monthly reviews, removes blockers when escalated, reactive
3 Active Steering	Weekly check-ins, proactively removes blockers, involved in decisions
4 Strategic Integration	AI in corporate OKRs, incentives tied to adoption, culture change

Engagement Level	%
Periodic Oversight (Level 2)	12%
Active Steering (Level 3)	58%
Strategic Integration (Level 4)	**29%**

Active steering works for projects within a single function. But the seven cases that achieved organization-wide transformation all reached strategic integration: the sponsor made AI adoption a measure of organizational success, not just a project to support.

Finding 2

Four activities define what effective sponsors do

Activity	Cases	What It Looks Like
Resource Allocation	59%	Dedicated budget, people, infrastructure for AI
Strategic Integration	49%	AI connected to business objectives and OKRs
Org Communication	32%	Messaging AI importance across the organization
Blocker Removal	20%	Actively clearing obstacles before escalation

Resource allocation is table stakes. What separates effective sponsors is what they do beyond budgets: connecting AI to business objectives, communicating its importance across the organization, and most critically, actively clearing obstacles before teams had to escalate.

> *"The president was on top of it, checking in every week: what is the progress, where are we, what are the bottlenecks? Which was helpful because then the rest of the team also engaged."*
>
> - Senior Executive, Technology Services Company

Finding 3

Business plus Tech co-sponsorship unlocks cross-functional projects

Eight cases showed co-sponsorship between business and technical leaders was what made the difference. At a professional services company, the first AI-push attempt was CTO-led and just failed to gain traction. The second try succeeded when the CEO and the Head of Talent drove it together with the CTO:

> *"The org had to know this was a CEO-led thing, this wasn't just the CTO. When AI is tech-led and tech-first, it does not work or it rarely works."*
>
> - Executive, Professional Services Company

The CEO provided a strategic mandate. The Head of Talent defined incentives and success metrics. The CTO owned implementation. Each brought something the others lacked.

At a telecom company, success came from finding a leader who bridged both worlds:

> *"The biggest enabler is that we hired a senior vice president of AI who had a deep understanding of the process and would map it out in detail. But he also had a deep understanding of artificial intelligence. That's our number one issue: we lack people who understand the process AND understand the AI and can put the two together."*
>
> - Executive, Insurance Company

Finding 4

Effective sponsors give teams permission to fail

Chapter 1 reported that 61% of successful projects included a prior failure. But failure only converts into learning under specific conditions. When we examined how sponsors handled setbacks, three strategies separated organizations where failure accelerated the next attempt from those where it led to abandonment.

Sponsor continuity through failure. In every case where we could identify whether the same executive sponsored both the failed and successful attempts, the answer was yes. At a technology company, the executive who built and then scrapped the first platform personally led the redesign six months later. At a semiconductor manufacturer, early AI initiatives stalled because engineering built solutions without coordination. The same AI leader who oversaw those failures escalated to the CEO and drove the second wave to production. When sponsors change after failure, institutional memory walks out the door: what not to do, which stakeholders to involve, where the real bottlenecks are. And most importantly, it sends a message to everyone that failure is a career risk.

Controlled scope as a failure strategy. 73% of implementations started small deliberately, and 63% framed their pilots explicitly as experiments. This is not timidity. It is a political calculation. Small pilots fail cheaply. Cheap failures do not end careers. One professional services company had failed twice on prior technology implementations. The sponsor accepted 80% accuracy as good enough to move forward, treating imperfection as a starting point rather than a flaw. Starting with an achievable bar gave the team room to iterate without the pressure of delivering a finished product on the first attempt.

Feedback loops instead of launch dates. The most effective approach to failure was not tolerating it after the fact but designing to handle it in advance. At a semiconductor manufacturer, the shift between the failed and successful attempts was making continuous user feedback and iteration a first-class priority in the solution lifecycle, rather than treating each deployment as a finished product.

The common thread: in *none of the cases* we examined was anyone punished for a failed AI initiative.

CASE STUDY

Field Service at a Semiconductor Company

How they achieved effective executive sponsorship

Hardware Manufacturing | Field Service | Enterprise

The Company

A semiconductor manufacturer producing solid-state drives for enterprise customers. The company has multiple departments with different levels of technical requirements: Engineering and IT at the front, Operations and Finance in the middle, Legal and HR at the back.

The Problem

When enterprise customers reported issues, field service engineers needed to gather technical data before diagnosis. Product specs, test libraries, data sheets, engineering logs lived in five or six different repositories owned by different teams. The service-level agreement (SLA) for data gathering alone was 40 hours.

> *"Documents, different data sheets, different test libraries, and it all is not centralized. Each of [them] is owned by different teams. When you have a sighting, all this has to come together."*
>
> - AI Leader, Manufacturing Company

The First Attempts Failed

Earlier AI initiatives built LLM-based agents for data analysis. They worked in demos but not in production. The problem was not technical. Engineering built solutions for their own use cases without coordination. There were no shared standards, no accountability for adoption.

What the Sponsor Did

The AI leader recognized that departmental sponsorship was not enough. He escalated to the CEO through three specific actions:

1. Established AI champions in every department. Engineering and IT adopted quickly. But Legal, HR, and other non-technical departments lagged. The sponsor created champions in each department to drive peer-to-peer adoption.

> *"What I saw was within an organization there are different levels … So, we got AI champions in each department."*
>
> - AI Leader, Manufacturing Company

2. Made AI adoption a corporate OKR. When peer pressure was not enough, the sponsor escalated to the CEO and made AI adoption part of how the company measures success.

3. Created visible leadership commitment through AI demo days. CEOs present at demo days, giving recognition to teams driving adoption. This signaled that AI was a strategic priority, not an IT experiment.

> *"We had AI demo days with rewards being given and CEOs presenting prizes. That recognition and pride coming from the people who are doing the work are actually pushing the momentum forward."*
>
> - AI Leader, Manufacturing Company

The Solution

With organizational support in place, the team built a multi-agent framework for the field service bottleneck. When a customer issue came in, agents pulled data from all repositories automatically.

> *"What the agent framework does is it goes into five or six different areas that it needs to look for information tied to this customer, tied to this issue, tied to this engineering area, and then it pulls it in."*
>
> - AI Leader, Manufacturing Company

The Results

Data gathering time

40+ hours → < 1 hour

Issues with complete data

0% → 95%+

Product testing cycle

20% reduction

Key Lesson

Departmental AI initiatives hit a ceiling when they require cross-functional adoption. Tying AI adoption to corporate OKRs and bonuses broke through resistance that standard communication and training could not.

> *"AI is a mindset change, it's nothing more than that. It is actually completely change-management driven."*
>
> - AI Leader, Manufacturing Company

Chapter 5

Where does fatal resistance come from?

Understanding where pushback originates and how to overcome it

Published Findings

End user adoption is a major barrier. Accenture lists lack of employee adoption as one of the top challenges for AI implementations.[9]

Leadership engagement varies. McKinsey notes that 33% of high performers have senior leaders actively driving adoption, compared to significantly fewer in the general pool.[11]

Workforce composition matters. Anthropic found that US states with higher concentrations of tech workers have higher AI adoption, suggesting resistance may be higher in non-technical workforces.[10]

The above research does not isolate middle management as a distinct source of resistance, nor does it address the nature of resistance: is it fear of replacement, lack of skills, or poor tooling?

What We Found

Staff functions (Legal, HR, Risk, Compliance) were the most frequent source of resistance at 35%, not the AI end users. Each source resists for different reasons: C-level demands measurable proof of ROI, staff functions worry about process risks and blame, end users distrusted system inconsistency, and frontline workers feared replacement. Each group required a different solution.

Finding 1

Staff functions, not end users, are the most frequent source of resistance

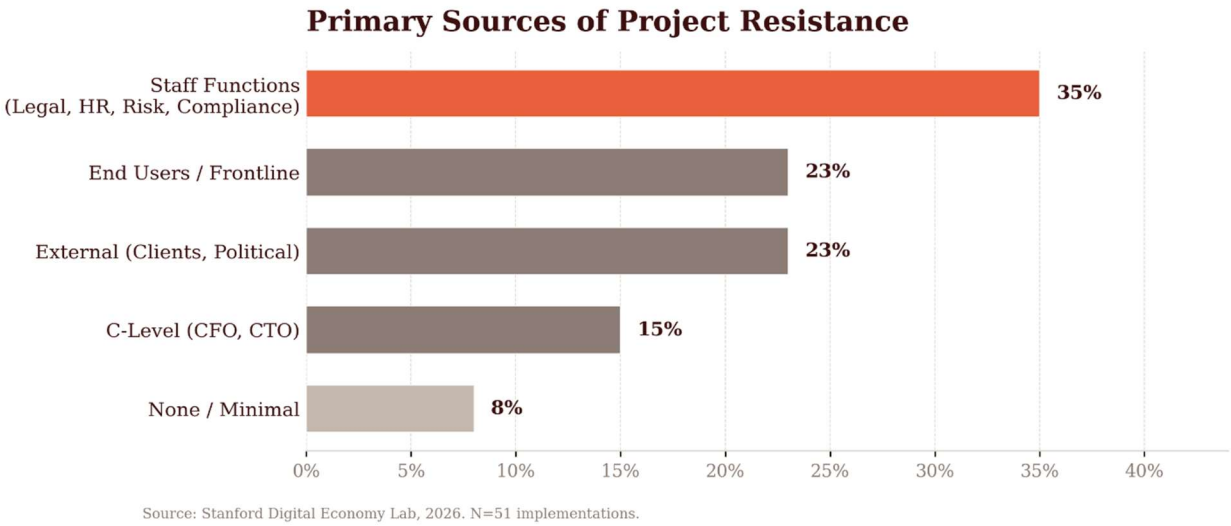


Figure 7. Sources of resistance to AI adoption

The conventional wisdom focuses on end user resistance, but staff functions were the most frequent blockers. Legal departments worried about liability. HR worried about change management. Risk and compliance teams worried about regulatory exposure. These functions have organizational authority to slow or stop projects regardless of executive support.

> *"What I saw was within an organization there are different levels of AI maturity. Engineering and IT want to push forward. Other organizations, maybe Legal, are holding back."*
>
> - AI Leader, Manufacturing Company

IT functions are a notable exception: rather than blocking, they more often serve as enablers, providing the platform infrastructure and data pipelines that allow business units to move faster.

> *"Middle management most resistant, while senior management and junior employees were more accepting."*
>
> - Executive, Retail Company

Finding 2

Each source resists for different reasons and requires different solutions

Staff functions worry about risk. At a large bank, past regulatory issues made risk teams extremely cautious. The solution is mandates, not persuasion. When AI adoption affected compensation through corporate OKRs and they don't need to take the blame for potential failure, Legal and HR found ways to enable rather than block. When given a role in governance rather than simply told to approve, staff functions frequently shifted from blocking to actively supporting deployment.

> *"I spent almost all of my time on risk and controls where everyone is very afraid to do anything."*
>
> - Executive, Large Bank

C-Level demands ROI proof. CFOs require clear financial justification before approving AI investments. The solution is measured pilots that demonstrate value before asking for broader investment.

> *"Hospital C-suite executives need direct line-item impact on balance sheet to justify software purchases."*
>
> - Executive, Healthcare AI Company

End users distrust inconsistency. Users accustomed to deterministic systems struggle with AI variability. The solution is expectation setting: users need to understand that AI outputs require review and that "good enough" performance on routine tasks frees time for higher-value work.

> *"We have to start by setting realistic expectations. Part of it is to change the thought process and shift the paradigm a little bit."*
>
> - Executive, Consulting Firm

Frontline workers fear replacement. This is the most discussed concern but appeared in only two cases. The fear is real but addressable. The solution is showing a concrete path forward: what work disappears, what work remains, and how roles evolve.

CASE STUDY

Security Operations at a Technology Services Company

How they overcame team resistance

Technology Services | Security Operations | Mid-Market

The Company

A technology services company with a six-person Security Operations Center (SOC) processing approximately 1,500 security alerts per month. The majority were false positives requiring manual triage.

The Problem

The team was drowning in alerts. With limited capacity, analysts could only investigate high-priority alerts thoroughly. Lower-priority alerts received minimal coverage. The work was mechanical: triage, classify, escalate or close. Analysts spent most of their time on repetitive tasks rather than the judgment-intensive investigation that required their expertise.

The Solution

The team deployed an AI system that automated alert triage. The AI handled initial classification and false positive filtering, processing alerts in seconds rather than hours. It escalated only alerts requiring human judgment to analysts.

Resistance

When leadership proposed the AI solution, **the expected concern was job security.** With a six-person team and AI capable of replacing most of the workload, the risk of resistance was real.

How the Sponsor Overcame Resistance

The Head of Technology had full mandate and no dependencies on other teams. He bought into the solution and ran the implementation as a dedicated program.

First, the context did most of the work. The team was already overwhelmed and failing to cover lower-priority alerts. This was not a team performing well that AI would disrupt. This was a team that could not keep up. AI was positioned as relief, not replacement.

Second, the division of work was intuitive. AI took the mechanical triage that consumed most of their time: classification, false positive filtering, routine escalation. Analysts kept the judgment-intensive work that required expertise.

Third, the sponsor framed freed capacity as a path up, not out. The extra bandwidth would go to higher-value activities that the team had never had time to pursue. The message was specific: AI replaces the hiring the company would otherwise need to do, not the people already there.

> *"You have to have a roadmap for the people. What's in it for the individual? They should see their life gets easier. And because it gets easier, that extra bandwidth is now employed for other activities which skill them up."*
>
> - Executive, Technology Services Company

The reframe was specific: AI replaces the hiring the company would otherwise need to do, not the people already there.

> *"AI is not replacing the person you have. AI is replacing the person **you don't need to hire**. The person you have can now do two or three or four people's work."*
>
> - Executive, Technology Services Company

The Results

Alerts processed **1,500 → 40,000/mo**	Alert coverage **High-priority → 100%**
Team capacity required **6 → 1.5 FTEs**	Freed capacity redeployed **4.5 FTEs**

No one was laid off. The 4.5 FTEs of freed capacity were redeployed to threat hunting, security architecture, and capability development.

Key Lesson

Fear of replacement dissolves when the path forward is concrete. The sponsor showed exactly what work would disappear (mechanical triage), what work would remain (expert investigation), and what new work would emerge (capability building). The team moved from resistance to advocacy once they saw AI as liberation from drudgery rather than threat to employment.

Chapter 6

When productivity gains are high, what happens to headcount?

Firing, reallocating, or freezing hiring?

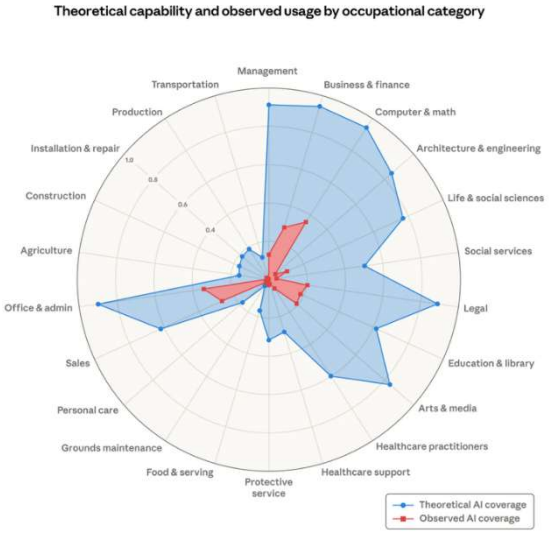
Published Findings

Expectations of decrease. McKinsey found that 32% of respondents expect their organization's workforce to decrease in the next year due to AI, while 43% expect little change and 13% expect an increase.[11]

Service Operations hit hardest. In the past year, 39% of respondents in Service Operations and 30% in Manufacturing reported a decrease in employees due to AI.

Deskilling versus upskilling. Anthropic's analysis suggests AI covers higher-education tasks, potentially leading to deskilling for some roles and upskilling for others.[10]

A recent Anthropic paper investigated theoretical AI job exposure to actual observed AI coverage adopted in the workplace, and showed that although in some fields, the exposure can be up to 90%, but actual current adoption is significantly lower. [23] (*see figure*) This would help to explain why the current impact on unemployment may not be as high as many had feared yet. But as adoption expands, the outlook may become much bleaker.



The above research captures expectations and aggregate trends but does not link specific high-productivity projects to actual headcount decisions.

What We Found

Reduction is the most common outcome at 45%, but not the majority. The combined alternatives (hiring avoided, no reduction, redeployment) account for 55% of cases in aggregate. Three distinct strategies emerge: accelerate rather than cut, redeploy to higher-value work, or reduce headcount directly. The technology does not dictate the outcome. Revenue-generating applications more often led to redeployment or acceleration, while cost-reduction applications more often led to direct cuts.

Finding 1

Reduction is the most common outcome, but not the majority

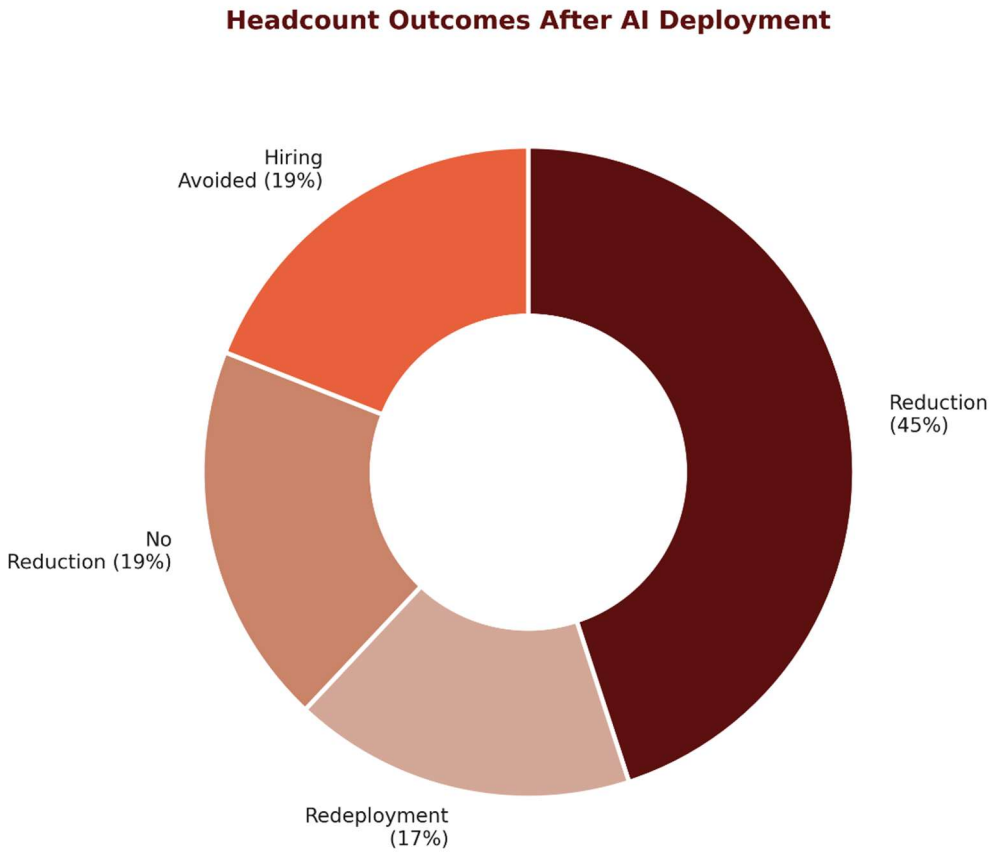


Figure 8. Headcount impact of AI deployments

Reduction is the largest single category but represents less than half of outcomes. Today, companies are still finding ways to capture AI productivity without eliminating positions. This may shift over time as productivity gains grow and social norms evolve.

Finding 2

Three distinct strategies emerge

Strategy 1: Accelerate rather than cut. Some companies explicitly chose to reinvest productivity gains into growth rather than cost reduction.

> *"There was debate on whether AI should reduce headcount. The CEO and COO leaned toward cost reduction; I pushed to use gains to accelerate the roadmap due to a large backlog."*
>
> - CTO, Education Technology Company

The productivity gains went into shipping more features faster, not into reducing engineering headcount.

Strategy 2: Redeploy to higher-value work. Other companies moved people from automated tasks to work that required human judgment. At a technology consulting company, AI automated 80% of invoice processing. Rather than cutting the team, they moved people to the next bottleneck:

Strategy 3: Reduce headcount directly. Some companies cut staff. At a private equity (PE)-owned company, an 88% productivity gain in coding led to reducing the development team from seven to three.

The choice depends on strategic context. Growth-stage companies tend toward acceleration. Cost-focused ownership (PE, turnaround) tends toward reduction. Technology does not dictate the outcome.

CASE STUDY

Engineering at an Education Technology Company

How they chose acceleration over headcount cuts

Education Technology | Engineering & Content | Enterprise

The Company

An education technology company with thousands of employees, including +200 in technology and +100 engineers. The company provides continuing education and professional certification courses across regulated industries.

The AI Implementation

The CTO implemented a three-pillar AI strategy: productivity tools for engineering, customer experience improvements, and AI-differentiated products. The engineering team ran a six-month pilot with GitHub Copilot and Cursor.

> *"Outcome: 20 to 30% reduction in time and effort for engineers early on, with upside expected as prompts and confidence improved."*
>
> - CTO, Education Technology Company

On the content side, the company re-architected production so that subject matter experts became human-in-the-loop reviewers rather than drafters. AI drafted content; SMEs refined it. This generated millions in cost savings.

The Headcount Debate

With documented productivity gains and cost savings, the leadership team faced a decision: use the gains to reduce headcount or reinvest them.

The positions were clear. The CEO and COO, under PE pressure to show returns, leaned toward cost reduction. The CFO was initially unconvinced that AI would generate net savings. The CTO argued for acceleration.

The Decision

For engineering, the company chose acceleration over cuts. The rationale was strategic: the company had a large product backlog. Shipping features faster would generate more revenue than cutting the team that ships features...for now.

> *"Savings were used to accelerate the roadmap, not reduce staff."*
>
> - CTO, Education Technology Company

For content production, the savings were captured and reinvested:

> *"We captured real savings in 2025 budgeting by reducing certain departmental budgets and reinvesting in AI. The biggest savings driver: SME content production re-engineering."*
>
> - CTO, Education Technology Company

The Results

Engineering productivity

20-30% time saved

Development costs

Millions in savings

Engineering headcount

No reduction

Savings reinvested in

AI development

Key Lesson

Productivity gains create a strategic choice, not an automatic outcome. The same gains that could justify headcount cuts can also justify accelerating the roadmap. The decision depends on whether the company prioritizes near-term cost reduction or long-term growth.

A Forward-Looking Caveat

The findings above are drawn from backward-looking data: what organizations chose to do with AI-driven productivity gains through early 2025. The pattern of redeployment and acceleration that dominates our sample may not persist as AI capabilities improve and economic pressures intensify.

Research from the Stanford Digital Economy Lab and Anthropic provides early evidence that broader labor market shifts are already underway. Brynjolfsson, Chandar, and Chen (2025), analyzing high-frequency payroll data from ADP covering millions of U.S. workers, found that early-career workers (ages 22–25) in AI-exposed occupations experienced a 16% relative decline in employment since late 2022. A complementary Anthropic study found no systematic increase in unemployment to date, but identified that hiring of younger workers has already slowed in AI-exposed fields, and that actual AI deployment remains a fraction of theoretical capability, suggesting the labor market impact is still in early stages. [5] [23]

This matters because the redeployment and hiring-avoidance strategies documented in our sample are characteristic of an early adoption phase, when organizations are still learning what AI can do. As implementations mature, models improve, and cost pressures mount, the distribution of outcomes is likely to shift. The 45% reduction rate we observed may represent a floor, not a ceiling. Companies that today choose acceleration over cuts may face different calculus when the next generation of models arrives. The canaries are singing.

Chapter 7

Where is AI opening doors that were previously closed?

How enterprises move from efficiency to new revenue, new capabilities, and strategic advantage

Published Findings

Revenue growth is the aspiration, not the reality. Deloitte's 2026 survey of 3,235 leaders found that 74% of organizations hope to grow revenue through AI, but only 20% are doing so today. Only 34% are using AI to deeply transform their business through new products, services, or reinvented business models.[21]

High performers pursue growth, not just efficiency. McKinsey's 2025 State of AI survey found that while 80% of organizations set efficiency as an AI objective, the companies seeing the most value also set growth or innovation as objectives[11]. Yet only 6% of organizations report EBIT impact above 5% from AI, and most revenue gains remain concentrated in marketing and sales, strategy, and product development.[11]

Published research shows that growth via AI is widely aspired to but rarely realized.

What We Found

Most implementations are measured as cost savings. But the highest returns came from companies that pointed AI at revenue: personalizing offers for each customer instead of segments, closing deals in hours instead of weeks, and packaging internal tools as products sold to clients. Others went further and used AI to do work no one had attempted before, like migrating legacy codebases or building sales intelligence in markets where no structured data existed.

As agentic systems changed the workflow of engineers and product managers, they will increasingly be liberated from the drudgery of manual coding and development tasks. This will give them more time to focus on higher value work, experimentation and collaborative innovation. In time, the novel applications and ideas that emerge from this change in working model could yield significant dividends.

Finding 1

New revenue from AI is real, but rare, and follows three patterns

Most implementations in our sample are measured as productivity or cost reduction. But a subset shows direct, quantified revenue impact. What distinguishes these cases is not the technology. It is that someone measured the revenue side, not just the cost side. The revenue mechanisms fall into recognizable patterns.

Personalization that converts.
A retail firm deployed AI to personalize marketing emails at scale, combining a machine learning recommendation engine with generative AI content. In the first month, they measured a 40% increase in purchase intent and a 20% increase in actual purchases. The AI did not change the product. It changed which product each customer saw.

> *"The only thing this did was it gave them better emails to send."*
>
> — Executive, Retail company
>
> *"60% opened the email, 40% went to the site, and probably 20% purchased something."*
>
> — Executive, Retail company

A food delivery company serving millions monthly orders moved from group-based segmentation (500 customers per segment) to individual personalization. The previous approach was not slow. It was structurally incapable of operating at this granularity.

> *"Instead of like taking three weeks to create 50 campaigns, now you can have a campaign for each person."*
>
> — Executive, Food Delivery Company

An enterprise content platform measured a 200% increase in click through rates for AI generated campaigns. Time to launch dropped from seven weeks to six hours. Both are efficiency metrics on the surface, but the volume and precision they unlock translate directly into revenue.

Speed that wins deals.

An insurance services company found that AI powered contract drafting turned speed into a competitive weapon. Contracts that previously took weeks were delivered in four hours. The result was not just efficiency. They won deals that would have been lost.

> *"They were drafting a contract that was really perfect, overseen by a lawyer, in four hours. In the past, it would have taken weeks and they might have lost the contract."*
>
> — Partner Value Creation, Private Equity Firm *(owner of the operating firm)*

In a market where speed of response determines who gets the deal, four hours versus four weeks is not an efficiency gain. It is a different competitive position for small and medium businesses.

> *"SMEs can respond much better to this leverage, and they can actually be the winners of this revolution."*
>
> — Partner Value Creation, Private Equity Firm *(owner of the operating firm)*

In semiconductor manufacturing, the same pattern emerged at a different scale. Reducing testing cycles by 20% and cutting customer issue resolution from 40 hours to under one hour changed how the company competed for enterprise accounts.

> *"When time to market for a product shrinks, it's not 5 million, 10 million in savings. It's hundreds of millions of dollars in savings."*
>
> — Executive, Semiconductor Manufacturer

From insight to product.

Some companies discovered that their internal AI capabilities could become revenue sources. A consulting firm that had built an analytics platform for marketing attribution realized the AI layer could generate predictive recommendations and simulate campaign outcomes. The firm is now launching this as a product offering and expects to "*double*" revenue from the platform.

> *"Then as we go towards productizing the simulator, then yes, probably doubling the revenue is what we would expect."*
>
> — Executive, Consulting Firm

A technology services company went further. After building an AI invoice processing solution for internal use, they packaged it and began selling it externally. A professional services firm's internal AI platform now serves new customers, becoming the foundation for service lines.

> *"We actually packaged it up and took some versions to some of our clients. One of the top three largest consulting companies on the planet is using it."*
>
> — Executive, Technology Services Company

What these cases share is not a common technology stack or industry. It is that someone asked a question beyond "how do we reduce cost?" and measured the answer.

> *"ROI is king. If you can show that in your sales cycle, that is immediately going to get you where you need to go. I've tried to sell efficiency with other things throughout my career and it is really difficult."*
>
> — Founder, Healthcare AI Company

Finding 2

AI is enabling work that was never on the roadmap

Beyond revenue, a separate group of cases shows AI making entirely new work feasible. Not faster versions of existing processes. Work that no one planned or budgeted because it was considered impossible.

Rewriting what was considered technically impossible. A fintech with over 100 million customers needed to migrate millions of lines of legacy code to a modern architecture. The traditional estimate was 18 months with over 1,000 engineers. With AI coding agents, business units began completing migrations in weeks.

> *"Rather than engineers having to work across several files and complete an entire migration task 100%, they could just review the changes, make minor adjustments, then merge their PR."*
>
> — Executive, Fintech

An insurance firm found that AI could rewrite legacy systems from scratch faster than refactoring them. A project originally quoted at 5,000 hours with a team of seven, scheduled for completion in 2027, was finished in 600 hours with a team of three. This opened a strategic question the firm had never considered:

> *"Do you buy a customer base and then try and retool that? Could you start from scratch and go disrupt a company by building their technology?"*
>
> — Executive, Insurance firm

Building market intelligence where none could exist. In healthcare markets with insurance, sales teams buy claims data to know exactly which providers prescribe what, in what volume. That is the standard playbook. But medical aesthetics is entirely cash pay. There are no claims, no centralized registries, no structured datasets. Territory intelligence for this market was not expensive or slow. It was impossible. A healthcare AI company changed that by building a system that scrapes public sources, assembles provider profiles, and scores prospects by estimated procedure volume and growth potential. For the first time, sales reps have a qualified pipeline in a market that never had one. The company is now expanding the platform to serve other manufacturers.

Turning operations into data assets. A robotic inspection company is generating historical datasets from consistent AI powered inspections that enable predictive analytics and incident forensics. Competitors cannot replicate this data without years of similar deployment. The inspections started as an efficiency play. The data they produce is becoming an entirely new asset.

These cases share a common trait: they were not on the roadmap before AI made them possible. No team was asked to do this work faster. AI wasn't just used to improve efficient, emergent solutions made themselves visible. The work itself was new and solved a problem they didn't realize existed.

CASE STUDY

Customer Relations at a Call Center

How AI turned a traditional call center into a growth engine

Call Center Services | Customer Relations | Mid-Market

The Company

A call center as a service (CCaaS) company providing traditional call center services to enterprise customers: answering calls, routing inquiries, and managing ticket queues. In a market increasingly defined by AI native competitors, the company's value proposition was under pressure.

The Problem

The CCaaS market was shifting. Enterprise customers were beginning to expect AI powered capabilities as standard, not as a premium add-on. AI native startups could offer intelligent routing, automated resolution, and real-time analytics from day one. The company's traditional model, built on seat-based pricing and human agents, faced two threats simultaneously: competitors could deliver more value per interaction, and the underlying pricing model was eroding as AI reduced the number of seats customers needed.

> *"One of the issues is that it has an impact on the SaaS model because it is reducing the number of seats. So you need to find a new way to price it."*
>
> — Partner Value Creation, Private Equity Firm

The challenge was not operational efficiency. It was strategic relevance.

The Solution

The management team embedded agentic AI directly into the company's product offering. Rather than using AI to make human agents faster, they redesigned the service so that AI could resolve tickets end to end, not just take calls or answer questions, but actually close issues.

The technical approach used an agentic AI framework that could orchestrate the full resolution process. This went beyond copilot-style assistance into autonomous task completion.

The Results

New project wins

20+ attributed to AI

Market position

Top 4 in AI for CX

Customer acquisition

Winning new logos

AI in every deal

No project without AI

Competitive repositioning

Traditional CCaaS → Benchmarked against AI natives

What AI Unlocked Beyond Cost Reduction

The same technology stack that could have been used to reduce headcount or lower cost per ticket instead repositioned the company in its market. Three things changed that had nothing to do with efficiency.

First, the company started winning deals it could not have competed for before. AI capabilities in the product became a differentiator. Thirty new projects were won not because the company was cheaper, but because it was more capable than competitors still operating on traditional models.

Second, the company's competitive set changed. An independent technology assessment ranked the company among the top four for AI capabilities in customer relations. The other three were AI native companies. A traditional call center was now benchmarked against startups, not against other incumbents.

Third, the pricing model began to shift. Instead of selling seats, the company could sell outcomes. AI capabilities became the product, not a cost reduction tool applied to the old product.

> *"Basically, what we see is that this [AI] is an approach that helps us win new deals."*
>
> — Executive, Call Center Services Company

Key Lesson

AI deployed for efficiency saves money. AI deployed into the product changes the competitive position. The difference is not the technology. The question is not "how do we reduce cost?" but "how do we win deals we could not win before?" This company asked the second question, and thirty new projects later, the answer was clear.

> *"I strongly believe that mid-sized companies and small size companies are very well positioned to win the AI revolution if you provide them the right capabilities. Decision making is taken much faster. They don't have that much legacy systems. They didn't know what to do with unstructured data, and now they can use it. And they lack resources, and the resources can get augmented with AI."*
>
> — Partner Value Creation, Private Equity Firm *(owner of the operating firm)*

Chapter 8

Is agentic AI generating real value?

Where autonomous AI works and where simpler approaches win

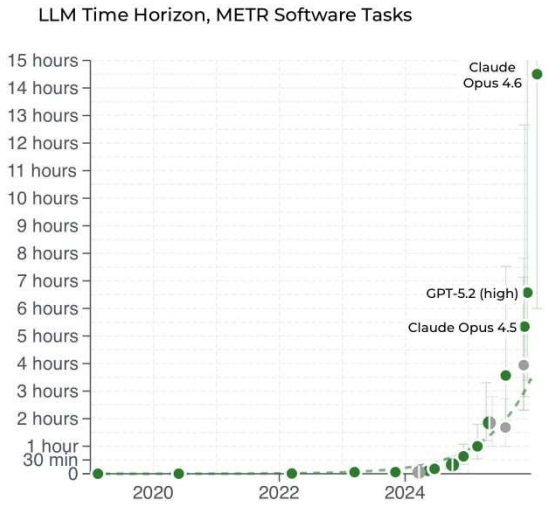
Published Findings

High hype, low scale. McKinsey reports that 62% of organizations are experimenting with AI agents, but only 23% are scaling them. Scaling is limited to one or two functions, most commonly IT and knowledge management.[11]

Emerging value in niches. OpenAI reports that enterprise AI adoption is accelerating unevenly by sector, with technology (11×), healthcare (8×), and manufacturing (7×) showing the fastest year-over-year growth, while finance and professional services operate at the largest absolute scale.[12]

Reliability limits. Anthropic warns that success rates decline as task complexity increases. Their data suggests API success rates drop below 50% for tasks requiring approximately 3.5 hours of human effort.[10]

Agentic Capability is growing exponentially. METR, an independent AI evaluation organization, measures the length of software tasks that frontier models can reliably complete autonomously. Their research shows this metric had been doubling approximately every seven months since 2019, but in recent months it has accelerated. As of early 2026, the most capable models can now reliably complete tasks without human intervention that would take a human expert approximately 15 hours.[22]. (*see Figure*) This trajectory suggests that the set of enterprise tasks suitable for agentic AI will expand massively in the near term. These are benchmark capabilities; real-world deployment still depends on integration, permissions, and exception handling.



Task completion time horizon of frontier AI models on software engineering tasks, 2019 to 2026

What We Found

Agentic implementations are currently a minority at 20% of cases. Most likely due to the reality that enterprise AI agent frameworks only emerged into the popular zeitgeist in 2025. But even with such immature scaffolding, agentic AI is delivering higher median productivity gains (71% vs 40% for high automation) in functions with high volume, clear success criteria, and recoverable errors. As these systems mature and use cases broaden, we expect the advantages of agentic AI to accelerate. It's important to note that agentic AI isn't a just new way to access AI, it's a redefinition of the role of

humans and machines in the workflow. Companies started treating AI as an extension of the team, not just a tool, guided and supervised by humans but increasingly capable of acting on their behalf and amplifying human capabilities beyond what one would have expected purely from looking at the team headcount.

Finding 1

Agentic AI is in production, but most implementations use simpler approaches

Level	Definition	%
Agentic	AI takes autonomous actions, completes multi-step tasks end-to-end without human intervention	**20%**
High Automation	AI handles >80% of work autonomously, humans review only exceptions or final outputs	34%
Human-in-Loop	AI and human work together, human reviews or approves each output before action	46%

Agentic implementations are the minority. The majority of successful enterprise AI uses simpler approaches: high automation with exception handling or human-in-loop collaboration. This does not mean agentic AI fails. It suggests many use cases do not yet require full autonomy to deliver value, and likely the bigger blockers to agentic AI adoption are the lack of technology maturity and limited deployment experience of the workforce.

Finding 2

Agentic AI delivers higher productivity but with wider variance

Level	Median Gain	Range	Top Use Case
Agentic	71%	20% – 80%	Field service: 80%
High Automation	40%	10% – 90%	IT Ops: 90%
Human-in-Loop	22%	10% – 85%	Invoice: 85%

Agentic implementations show the highest median productivity gain at 71%. The highest gains came from field service where a multi-agent framework gathered data across repositories automatically. Human-in-loop shows 22% median, appropriate for document review and clinical documentation where human judgment is essential.

As agentic frameworks mature over time, we expect an increasing percentage of use cases to fall into the full autonomy category. In the coding space, that trend is already increasingly clear. Examples of coding agents (*Claude Code, OpenAI Codex, etc.*) running for days autonomously delivering tens or hundreds of thousands of lines of working code is not uncommon in recent months.

This level of autonomous capability will not only increase productivity, but it will also redefine roles in the organization. Team members with limited or no technical experience will soon be able to build and deploy complex projects just by having a natural language conversation with the toolset, as they had in the past with the development team lead. This level of capability won't be restricted to the software development realm alone. It's foreseeable that such workflows will extend into financial, accounting, consulting services and other data focused sectors rapidly. The macro labor implications of this on the wider economy can be dramatic when adoption widens.

Finding 3

Successful agentic implementations share common characteristics

The ten agentic cases clustered in specific functions (procurement, field service, security operations, coding, and customer support triage), but what matters more than the function is what these implementations have in common: **High volume, repetitive tasks.** Security operations processing thousands of alerts. Procurement handling hundreds of decisions. Customer support triaging tickets. The volume justifies the investment in building autonomous systems.

Clear success criteria. Alert is valid or not. Procurement decision is correct or not. Ticket is resolved or not. The AI can evaluate its own outputs against objective criteria.

Recoverable errors. A missed alert can be caught later. A wrong procurement recommendation can be overridden. A failed ticket resolution escalates to a human. Errors are costly but not catastrophic.

Data access across systems. Agentic AI requires the ability to query multiple systems, gather information, and take actions. Implementations that succeeded had invested in data infrastructure and API access.

> *"Don't just apply AI to your existing processes. That's a mistake. We're **redesigning our workflow** and that's what makes us successful."*
>
> - Head of Operations, Technology Company

CASE STUDY

Procurement at Supermarket Chain

How they built agentic AI that delivers real value

Retail | Procurement | Mid-Market

The Company

A regional supermarket chain with approximately two dozen stores. Unlike market leaders with substantial margins and massive procurement power, this company operated at roughly half the industry benchmark with minimal negotiating leverage against suppliers.

The Problem

Supermarket economics are unforgiving. Margins are thin, waste is constant, and stockouts lose customers permanently. The company faced three interconnected challenges:

First, waste. Perishable goods expiring on shelves, seasonal products ordered in wrong quantities, and promotional items that did not sell.

Second, stockouts. Empty shelves do not just lose one sale. They lose the customer who drives to a competitor and may not come back.

Third, procurement timing. A human buyer made decisions based on gut feeling, supplier relationships, and whatever data they could manually compile. They could not possibly optimize thousands of SKUs across 25 stores.

The Solution

The company deployed an AI system that replaced the human procurement function entirely. The system does not assist humans or generate recommendations for review. It makes purchasing decisions autonomously.

The architecture has three components: a data platform that pulls inventory, sales, and supplier data from multiple systems; demand forecasting models that predict sales at the store and stock keeping unit (SKU) level; and an autonomous procurement agent that decides what to buy, when to buy it, and from which supplier.

> *"They replaced the human procurement guy with an AI tool that is buying. Telling them what to buy. And so again they have the supermarket full of stuff and the stock is optimized."*
>
> - Project Lead, Retail company

What Makes It Agentic

This implementation crosses the line from automation to agentic AI in three ways:

First, it replaced a human function, not just a task. The AI took over the procurement role entirely, determining what to buy, when to buy it, and how much, across all stores simultaneously. A human buyer making the same decisions could not optimize at this scale.

Second, it connects multiple decision steps that were previously handled by intuition. A single procurement decision requires predicting demand, checking current inventory, factoring in supplier lead times, and balancing waste against stockouts. The system handles this chain continuously across every product in every store.

Third, it operates across multiple systems without human orchestration. The AI pulls inventory data, supplier catalogs, and sales history from different sources, processes them together, and outputs purchasing decisions. Before, a human buyer was the integration layer.

The Results

Waste reduction

40%

Stockout reduction

80%

EBITDA margin

Doubled

> *"The market leader has much higher margins. These guys are super small. But they do almost as well, and their procurement power is zero compared to the big players. But what they have is that they don't have waste."*
>
> - Project Lead, Retail Company

Key Lesson

This case illustrates that even for more traditional sectors, when agentic AI is applied properly, it creates real value: tasks too complex for rules but too repetitive for humans, with clear success criteria and recoverable errors. Thousands of SKUs across dozens of stores required continuous optimization no human could perform. The AI could evaluate its own performance because outcomes were measurable: did the product sell, expire, or run out? For a small retailer competing against giants, agentic AI turned intelligence into a substitute for scale.

Sample Limitations and Future Outlook

Our findings on agentic AI should be interpreted with an important caveat: agentic technology was still emerging during our data collection period (August 2024 to January 2025). Only 20% of the implementations in our sample involved agentic workflows, and most organizations were experimenting rather than scaling. The limited sample of agentic cases reflects the state of the technology at the time, not its long-term trajectory.

That trajectory is likely to be transformative. Foundation models and agentic frameworks are improving rapidly in their ability to reason, plan multi-step workflows, and recover from errors.

These are precisely the capabilities that define agentic AI. As these models advance, the share of enterprise use cases suitable for agentic approaches will grow substantially. Tasks that today require structured automation with human oversight may increasingly be handled by autonomous agents capable of navigating ambiguity and making context-dependent decisions.

The 71% median productivity gains we observed in agentic implementations, compared to 40% for high automation, suggest that when agentic AI is applied to the right use cases, the impact is significantly larger. As technology matures and the conditions for successful deployment become better understood, we expect agentic implementations to represent an increasingly dominant share of enterprise AI value creation. The patterns documented in this chapter capture the early innings of a trend that will likely reshape how organizations think about the boundary between human and machine work.

Chapter 9

How clean does enterprise data actually need to be?

Why the real data challenge is access and storage, not cleanliness

Published Findings

Clean data is a scaler's advantage. Among Strategic Scalers, 61% possess a large, accurate data set, compared to just 38% of companies stuck in Proof of Concept. Strategic Scalers are adept at "tuning out data noise" to focus on priority domains like financial, marketing, and customer data.[9]

Data products are key. McKinsey notes that high performers are more likely to have created "reusable, business-specific data products."[11]

Unstructured data tolerance. OpenAI reports that enterprise use of structured workflows (Custom GPTs and Projects) grew 19× year-to-date, now handling approximately 20% of all enterprise messages. This suggests organizations are succeeding by building access layers to existing data rather than requiring perfect data structures before deploying AI.[12]

The above research establishes that clean data correlates with AI success. It does not quantify how messy data can be while still yielding results.

What We Found

Only 6% of implementations had data that was fully ready for AI. But in the majority of cases where data challenges existed, LLMs were part of the solution, not just the consumer of clean data, but the tool that made messy data usable. Models unlocked previously inaccessible data in 88% of cases, processing voice transcripts, scanned documents, legacy code, and scattered knowledge bases that no prior technology could handle. A caveat: because this sample focuses on successful implementations, it likely underrepresents cases where data quality proved insurmountable. The finding reflects what is possible with deliberate design, not a universal guarantee.

Finding 1

LLMs are not just consuming data. They are fixing it!

The conventional narrative assumes AI needs clean data to work. Our data tells a different story. Only 6% of implementations had data that was fully ready for AI deployment. **The vast majority faced data challenges ranging from moderate to severe.** Yet in most of those cases, LLMs were part of the solution to the very data problems they were expected to struggle with.

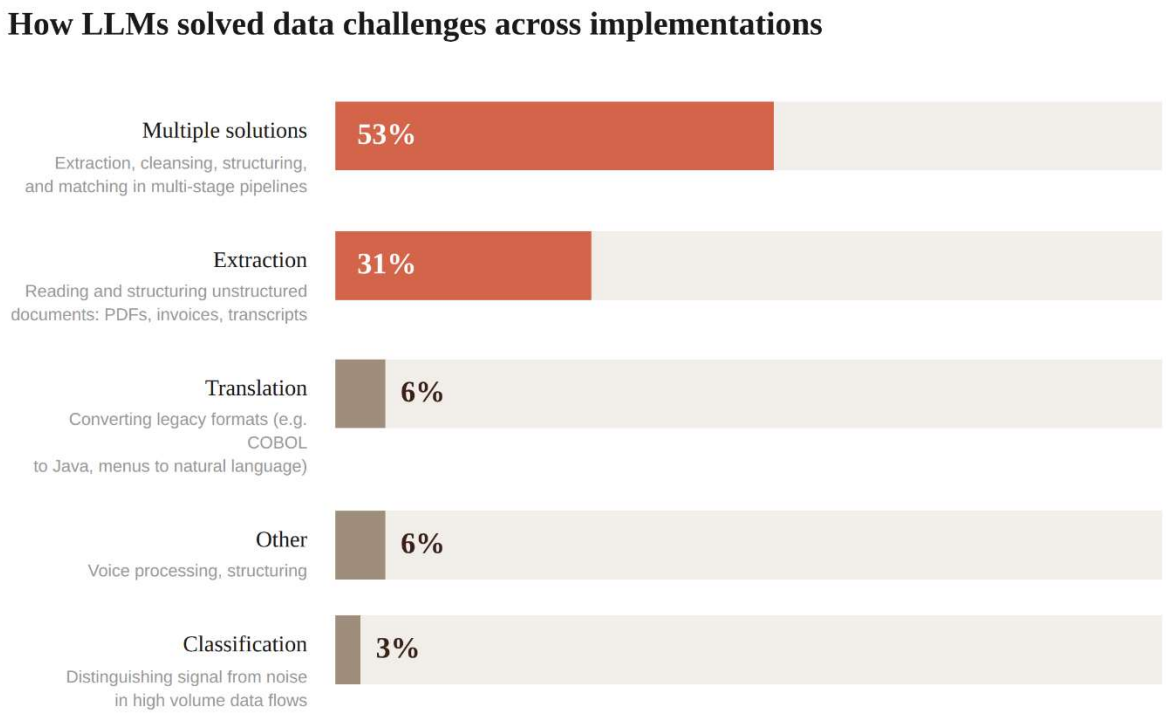


Figure 9. Data quality challenges across deployments

This is a fundamental shift. Previously, unstructured data required human analysts to impose structure before any analysis could happen. Now, 91% of our implementations successfully processed unstructured data, including voice transcripts, scanned documents, images, chat logs, and legacy code, that would have been unusable two years ago. In 88% of cases, LLMs unlocked data that was previously inaccessible, not because it did not exist, but because earlier approaches (OCR, rules engines, manual tagging) couldn't process it at the accuracy and scale required.

> *" We've had partners tell us, hey, it would have taken us two months to clean this up, and you guys flagged all the data issues within a day."*
>
> - VP of AI, Professional Services Firm

The types of data made newly accessible span the full range of enterprise information.

Voice and conversation data. Ambient transcription in healthcare made doctor-patient conversations accessible to coding teams for the first time. Call center transcripts became sources of real-time coaching and quality assessment. Previously, coding teams had no window into clinical decisions. Now they have the full conversation.

> *"With an ambient transcription technology, now that person does have access to everything that was discussed as part of that person's medical care. The auditability and traceability of medical care is now much more possible with this technology*
>
> - Executive, Healthcare Company

Scattered documents and knowledge bases. A semiconductor manufacturer reduced data gathering more than 10 times by deploying a multi agent framework that pulls information from five or six different repositories automatically. The data was technically available but practically inaccessible due to organizational silos and the sheer time required to assemble it manually.

> *"Documents, different data sheets, different test libraries - it all is not centralized. Each of it is owned by different teams."*
>
> - VP of Engineering, Semiconductor Manufacturer

Visual and multimodal data. Field technicians can now photograph equipment and receive instant AI generated repair instructions. Retail procurement systems process scanned paper forms, emails, and Excel spreadsheets that previously required armies of manual data entry clerks.

> *" You can take a photo of it and AI will instantly give him a detailed description of that device and how to fix it.*
>
> - VP of AI, Telecom Company

Finding 2

Process documentation and access matters more than data perfection

Of the implementations where we could assess the data architecture, 59% had data scattered across multiple systems owned by different teams. Only 16% had fully centralized data. Yet success did not require centralization. It required access.

Organizations that built integration layers—whether APIs, RAG architectures, or multi agent frameworks—to connect scattered data performed as well as those with centralized data stores. In the pre-LLM world, enterprises had to structure and centralize data before extracting value. Today, RAG architectures and knowledge base connectors work with messy data if the retrieval layer is well designed.

> *"All the hard work is in process documentation and data architecture. If you can do those two things, everything else is quite simple."*
>
> — VP of AI, Professional Services Firm

A telecom company built different knowledge bases for different equipment types, indexed them, and gave AI agents access through model context protocol (MCP), without ever centralizing the underlying data.

> *"We've basically built different knowledge bases for different objects. The MCP can go out to these various tools that we have built for different situations."*
>
> — VP of AI, Telecom Company

Finding 3

Proprietary data is the durable competitive advantage

Every frontier lab is training on every piece of public data it can access. Organizations cannot compete on that axis. But every company has proprietary data that no frontier lab has ever seen or is allowed to see. That data is their edge.

Across our sample, 75% of implementations mentioned proprietary data as a key factor in their AI strategy, and 47% explicitly described their accumulated data as a *competitive moat*. The pattern was consistent across industries: the organizations generating the most value from AI were those that had been storing data, even imperfect data, long before they knew how they would use it.

> *"Our differentiator, the reason why people buy from us, is because in our last 13 years in business, we created this knowledge graph. We have over 20 billion data points."*
>
> — Executive, HR Technology Company

> *"The power comes from leveraging unique assets: data, SMEs, customer base, relationships. Differentiation requires what others can't replicate quickly."*
>
> — CTO, Education Technology Company

The implication is straightforward. ***Save everything.*** The cost of storing data is negligible compared to the cost of not having it when the right use case arrives. Organizations that preserve their data, however imperfect, are building a competitive advantage that compounds over time. As open-source models close the performance gap with proprietary ones, the differentiator shifts from which model you use to what data you feed it.

CASE STUDY

Procurement at a Construction Services Company

How they succeeded with bad data on both sides

Construction Services | Procurement | Enterprise

The Company

A large construction services company with field technicians who need parts delivered to job sites.

The Problem

Technicians submitted requests via paper forms, emails, and Excel spreadsheets. A team manually entered these into the procurement system and matched items to the parts catalog. Slow, error-prone, expensive.

The Solution

AI extracts requests from unstructured sources, matches to catalog, and creates requisitions automatically. Also identifies when items are not in catalog and suggests alternatives.

The Data Challenge

Data quality issues on both sides:

> *"The quality of extraction from unstructured sources was not good - OCR was not giving us good results. And the quality of the structured data we were matching to is also not consistent. The core data itself isn't of the best quality."*
>
> - AI Practice Lead, Professional Services Firm

How They Overcame It

A four-stage pipeline that improved data progressively:

1. Extract with Python when OCR failed.

2. Cleanse with generative AI using vectorization and embedding to eliminate stray characters.

3. Fuzzy match to catalog despite imperfect reference data.

4. Human-in-loop for exceptions rather than requiring 100% accuracy.

> *"We shifted from 'this is your requirement; this is what it will do' to 'what does good enough look like?' AI will improve if you monitor it and give it better data over time."*
>
> - AI Practice Lead, Professional Services Firm

The Results

Investment	Productivity gain (projected over 3–5 years)
$500K - $1M	**30%**

Expected ROI

10x over 3 years

Key Lesson

Design for **'good enough**,' not perfection. Each pipeline stage added value even with imperfect input from the previous stage.

Chapter 10

Does rigorous security protect the project or kill it?

How security requirements affect AI project outcomes

Published Findings

Security is a top priority. McKinsey reports that 51% of organizations are working to mitigate cybersecurity risks, the second most common mitigation after inaccuracy.[11]

High performers take more risk, not less. Counterintuitively, AI high performers are more likely to report negative consequences (like IP infringement) and more likely to mitigate risks than peers. This suggests they are using AI in mission-critical contexts rather than avoiding risk entirely.

Regulated sectors are not blocked. OpenAI reports that healthcare, one of the most heavily regulated sectors, is among the three fastest-growing for enterprise AI adoption at 8× year-over-year, while financial services operate at the largest absolute scale. This suggests that stringent security and compliance requirements are not preventing adoption in high-stakes fields.[12]

What We Found

In our sample, security was not a pure project killer. In every case where security created barriers, those same requirements eventually enabled the project to handle sensitive data that would otherwise be off-limits. Shadow AI (where employees use unauthorized AI tools) emerges when formal channels fail to keep pace. The security tax is real but front-loaded.

Finding 1

Security requirements that initially block eventually enable

Of 12 cases with complete data, we found that security was never a pure project killer. In every case where security created barriers, those same requirements eventually enabled the project to handle sensitive data that would otherwise be off-limits.

The pattern was consistent: teams forced to build robust data protection infrastructure unlocked use cases that competitors without such infrastructure cannot touch.

> *"At most banks there was a mindset that it has to be completely within house. We're not going to use any software or hardware that is not within the firewall. Now when you introduce AI, everything is cloud based and so we had to update that kind of policy."*
>
> - Executive, Large Financial Institution

The same institution, after years of security work, now runs customer-facing AI that handles sensitive financial data, scrubs personally identifiable information (PII) before sending to external models, and reassembles it on return. Security investment became the foundation for capabilities competitors cannot replicate quickly.

Finding 2

Shadow AI emerges when formal channels fail to keep pace

Shadow AI refers to the use of AI tools and platforms by employees without formal authorization from IT or security teams. It is the AI-era equivalent of shadow IT, but carries amplified risks: employees routinely upload proprietary data, customer records, and internal documents to consumer AI platforms that lack enterprise security controls.[13]

The problem is pervasive. Industry surveys find that **70% to 80%** of employees who use AI at work rely on tools not approved by their employer.[14] An IBM study found that while 80% of workers use AI, only 22% use exclusively company-provided tools.[15] Among those using unauthorized platforms, **57% admit to entering sensitive company information**.[16] The financial consequences are real: AI-associated data breaches cost organizations an average of more than **$4mm per incident**.[13]

Our case studies confirm these findings and reveal the organizational dynamics behind the numbers.

Shadow AI was explicitly mentioned in 15% of cases. Two distinct patterns emerged:

Pattern A: Enthusiasm outpaces governance. A semiconductor manufacturer discovered a massive collection of different AI tools in use across the company. Employees were not being malicious. Leadership had signaled "use AI" before any platform existed.

> *"When I did the security analysis, we found the company staff are using 1,500 or 1,600 different AI tools. So our objective was building working internal platforms before we go and say you cannot use non-approved tools."*
>
> - Executive, Semiconductor Manufacturer

Pattern B: Desperation beats bureaucracy. In healthcare, physicians adopted ambient transcription tools without formal approval because hospital systems were too slow to evaluate and procure them. The doctors were burned out, the technology existed, and the formal process took too long.

> *"A lot of these doctors have been adopting these technologies without approval or a formal vendor selection process."*
>
> - Executive, Healthcare AI Company

The insight is not that shadow AI is good or bad. It is that shadow AI is a symptom that policy moves slower than technology, and it needs to be expected but accounted for to some level. When formal security processes cannot keep pace with demand, users find workarounds. In industries like healthcare, finance and government, legal and regulatory liabilities could be massive.

Finding 3

The security tax is real, but the investment pays forward

Quantifying security delays proved difficult, but qualitative evidence from regulated industries suggests the tax is substantial.

> *"I come from tech startups where you would perish if you didn't just try stuff. Going to a large, regulated institution within financial services... just a night and day difference."*
>
> - Executive, Large Financial Institution

Projects in this environment take multiple years to set up. That is the security tax in its most extreme form.

But the tax is front-loaded. Once the infrastructure exists, subsequent projects leverage it. The company built data scrubbing pipelines, established contracts with cloud providers, and created compliant archival systems. Each new AI use case now builds on that foundation rather than starting from scratch.

When the tax makes sense: The security investment is justified when it enables use cases that would otherwise be impossible. Handling customer financial data, processing healthcare records, managing confidential M&A documents: none of these are possible without robust security.

When the tax is wasteful: The security tax is wasteful when it blocks work without enabling solutions to employee/customer issues. When formal processes are too slow, shadow AI fills the gap, heightening many of the security risks the process was designed to prevent.

CASE STUDY

Customer-Facing at a Large Retail Bank

How they went from "everything within firewall" to cloud AI

Financial Services | Customer Support | Enterprise

The Company

A large US-based retail bank serving millions of customers through mobile apps, branches, and call centers. The bank operates under federal banking regulations and, following past compliance issues, faced consent orders that created a deeply risk-averse culture.

The Problem

The bank wanted to deploy AI-powered virtual assistants in their mobile app. But their technology policy prohibited using any software or hardware outside the corporate firewall. Modern AI is cloud-based. The policy made cloud-based AI impossible.

The Solution

The team developed a data protection architecture with four components:

PII scrubbing on exit. Customer utterances are stripped of names, account numbers, and dollar amounts before leaving the firewall.

Synthetic data substitution. Fake values replace real ones during external processing.

Intent processing externally. The cloud model determines customer intent and selects the appropriate workflow.

Reassembly on return. Real values are reinserted internally before presenting the response.

> *"What we send to the Google Cloud platform is a minimum scrubbed set of what's needed. We swap in fake names, a fake dollar amount. It can still discern what the intent is. Then on the return we remarry all of that into the response."*
>
> - Executive, Large Financial Institution

The Results

Channel cost	Call containment
Lowest to serve	**48-72hr reduction**

Next phase

Agentic AI for scheduling

Key Lesson

Security is infrastructure, not overhead. The years spent on security built the foundation for handling sensitive financial data at scale. Without that investment, the bank could not offer AI services that touch customer accounts.

The tax is front-loaded. Most security cost came before the first deployment. Subsequent use cases leverage the same pipelines, contracts, and archival systems. ***Caveat:*** *With the many highly capable open-source solutions available now, there may be viable options to deploy the AI function in-house that would deliver similar results without altering existing policy.*

Risk-averse culture is the hardest barrier. The technical solutions weren't completely straightforward, but enabling cultural flexibility did give it a faster time to market advantage over competition. Changing a culture shaped by past consent orders was the real challenge.

Chapter 11

When is foundation model choice not a commodity?

When model selection matters, when it does not, and what drives the open vs. closed decision

Published Findings

Model quality matters more as complexity increases. Anthropic shows that success rates decline sharply as task horizon grows, implying that for complex, long-horizon, or agentic workflows, model capability becomes a binding constraint.[10]

Reasoning workloads are growing. OpenAI reports a 320x year-over-year increase in reasoning token consumption, suggesting enterprises are pushing models into more complex domains where performance differences may matter.[12]

Open models reach 90% of closed model performance. MIT research analyzing five months of OpenRouter inference data found that open models routinely achieve 90% or more of proprietary model quality at release and rapidly converge. Yet closed models still account for approximately 80% of token usage.[17]

Over half of enterprises already use open-source AI, often alongside proprietary tools. A McKinsey, Mozilla, and McGovern Foundation survey of 700 technology leaders found that more than 50% of organizations use open-source AI somewhere in their stack, rising to 72% among technology companies[18]

The cost gap is widening faster than the performance gap is closing. Open models achieve roughly 90% of proprietary model performance while costing on average six times less per token in observed API pricing, based on Artificial Analysis benchmark and pricing data. [19][20]

Published research captures both the commodity question and the cost and performance tradeoff between open and closed models. What it does not address is how enterprises in production actually make these decisions, where factors beyond benchmarks shape outcomes.

What We Found

For 42% of implementations, model choice is fully commodity. The commodity boundary exists at task complexity: routine tasks are 4x more likely to be commodities than advanced tasks. Multi-model strategies are the emerging norm, and abstraction layers separate leaders from laggards. Despite growing open-source availability, enterprises overwhelmingly default to proprietary models, and the decision is driven by capability and speed rather than cost. **Caveat:** *Going forward, as the capability gap between open and closed models continues to narrow [which is happening rapidly], it's likely the percentage of tasks in the commodity category will grow significantly.*

Finding 1

For most enterprise use cases, model choice is commodity

Verdict	%
Commodity (interchangeable)	**42%**
Moderate importance	39%
Critical differentiator	19%

The pattern was consistent: success came from everything around the model - data quality, process documentation, integration architecture, change management - not from the model itself.

> *"The most important thing that we've ever done was spending a tremendous amount of time with our RAG and really nailing down our chunking strategy."*
>
> - Director, Professional Services Firm

Finding 2

The commodity boundary is defined by task complexity

We grouped implementations into two categories based on cognitive demands:

Routine tasks: Repetitive, rules-based work with clear success criteria. Customer support triage, document search, marketing content, recruiting screening.

Advanced tasks: Work requiring multi-step reasoning, domain expertise, or consequential decisions. Complex coding, compliance analysis, clinical documentation, agentic workflows.

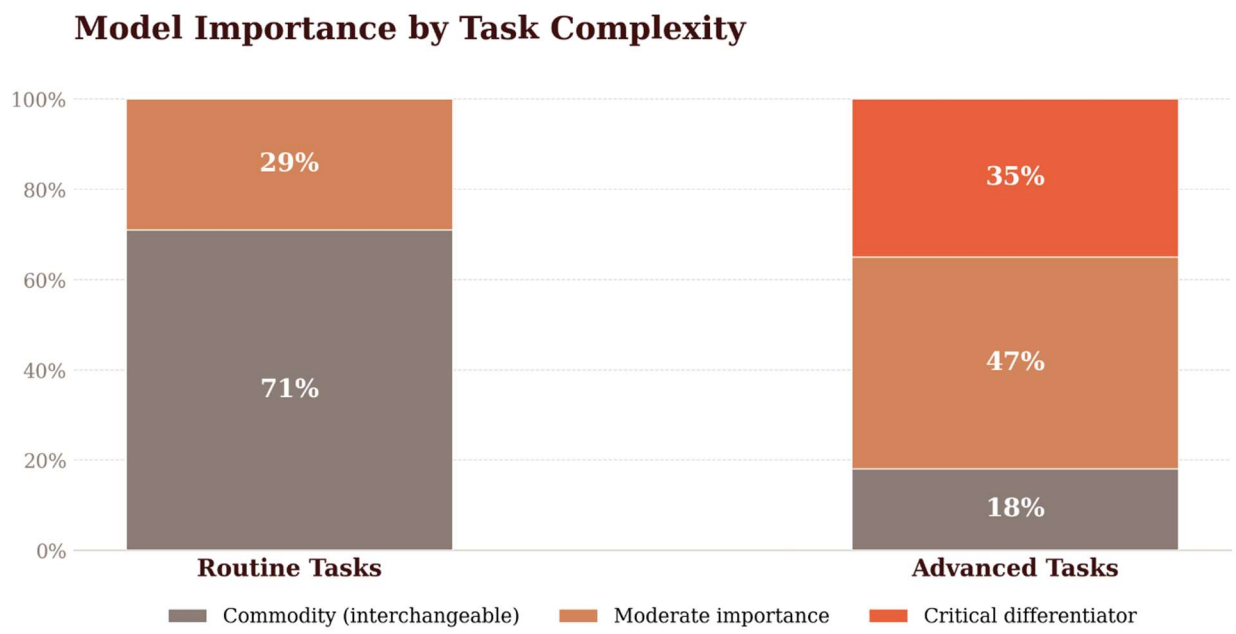


Figure 10. Model interchangeability by task complexity

Among routine tasks, 71% treated the model as fully interchangeable and none considered it a critical differentiator. Among advanced tasks, only 18% treated it as commodity while 35% saw it as critical.

> *"Barriers to entry here are much lower than they have traditionally been as LLMs become more commoditized."*
>
> - Executive, Healthcare AI Company

Finding 3

Multi-model strategies are the emerging norm

The majority of implementations used multiple models rather than committing to a single provider. The multi-model approach took several forms:

Task-specific routing. Different models for different tasks: using a fast, cheap model for classification and a more capable model for generation or reasoning. The cost difference can be 10x or more. Often small models can be run locally meaning it's essentially just the cost of electricity rather than using costly model APIs which can result in big savings, not to mention the privacy benefits.

Validation through redundancy. Running the same query through two different models and only accepting matching answers.

> *"I think we ran it against two different RAG models. And if we got the same answer, then it was a good answer. And if the two RAG models weren't matching, then it was a bad answer."*
>
> - Project Lead, Professional Services Company

Query-based optimization. By adding an intelligent evaluation and routing layer for each model, query systems can derive significant benefits based on the use-case requirements.

> *"We built a multi LLM gateway. The ability to really solve for cost, accuracy, relevance, latency based on the query. At each query, the goal is to say okay, does this result require a deep search? Or is a mini-model good enough?"*
>
> - Head of Operations, Technology Company

Finding 4

Model abstraction layers are becoming a competitive advantage

The most sophisticated implementations included abstraction layers that allow model switching without rearchitecting the system. These organizations treat models as interchangeable components within a larger platform.

> *"My focus is not so much about the tools. My focus is to build a platform and once the platform is there, then they will use the platform. You have flexibility to pivot between models if and when one gets better or cheaper than the other."*
>
> - Head of Operations, Technology Company

A food delivery company built their own AI Chatbot on top of multiple foundation models: OpenAI, Gemini, and Claude. This abstraction layer allowed them to achieve 90-95% automation in customer service without dependency on any single provider.

The organizations with abstraction layers share a common philosophy: models are improving rapidly and unpredictably. Rather than betting on a single provider, they built infrastructure that allows them to adopt improvements from any source. With the rapid releases from all the frontier and open-source labs, the ability for the system to pick the right model for the job is becoming a competitive advantage in itself.

The highest-performing implementations treat models as interchangeable components within platforms they control. **The durable advantage is in the orchestration layer, not the foundation model.**

Finding 5

Open-source models are entering production, but in specialized roles

Open-source is not absent from enterprise deployments. It is showing up in specific functions where customization and control outweigh the need for frontier capability: specialized tasks like named entity recognition, security functions requiring full model visibility, and startup products where fine-tuning on domain data drove the initial architecture.

A major financial services institution illustrated the emerging pattern. Its core customer-facing capabilities run on proprietary models, but its information security functions use open-source models where the team needs to customize and control exactly how the model behaves.

> *"Some of the models that we use for information security, I would say the supporting or helper type models... we might use open-source models for some, like specifically NER, named entity recognition. Those are open-source."*
>
> - Senior Executive, Major Financial Services Institution

A cybersecurity vendor took a different path, building its entire product on Llama as a base model and fine-tuning extensively. The choice was not about performance leadership but about the ability to adapt the model to a narrow domain at manageable cost.

> *"We just took Llama as a base model and then modified it."*
>
> - Executive, Technology Services Company

At the other end of the spectrum, a cloud-native software company rejected open-source models entirely, prioritizing the security and privacy guarantees that come with proprietary enterprise-tier licensing.

> *"We avoided most open or free or unsanctioned tools due to privacy and security terms and conditions. We're cloud-native; no local self-hosting."*
>
> - Engineering Leader, Software Company

Open-source adoption in enterprise AI is entering through specialized, lower-risk functions. The question is not whether enterprises will use open-source models, but how quickly the supporting infrastructure will catch up to make them viable for core production workloads. As the gap between closed and open-source offerings narrow, open will increasingly gain mindshare given the sizable cost and technology sovereignty advantages.

The rise of capable Chinese open-source models in early 2026 (Qwen, Kimi, Minimax, GLM, etc.) has narrowed the capability gap with proprietary models while maintaining a significant cost advantage. On OpenRouter, a platform that routes API requests across 400+ models for over 4 million users, 4 of the top 5 models by token volume are now Chinese open-source, driven largely by agentic workloads that consume exponentially more tokens than traditional chatbot use. That said, most U.S. enterprise deployments still rely heavily on American providers such as OpenAI, Anthropic, Meta's Llama, and Google's Gemini, where compliance, support, and vendor qualifying remain key factors. As agent-driven architectures scale, managing model selection and inference costs (sometimes called tokenomics) with real-time model selection will become an increasingly important capability for technical teams.

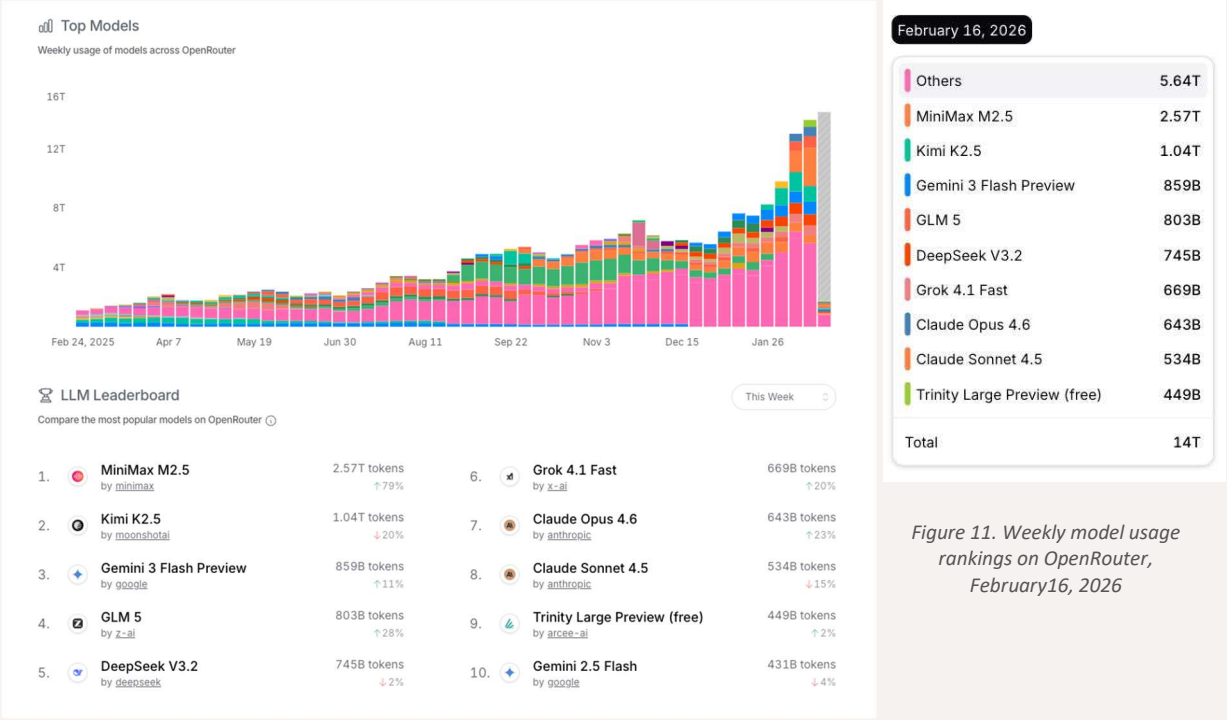


Figure 11. Weekly model usage rankings on OpenRouter, February16, 2026

Finding 6

The current focus is capability and speed, not cost

More than two thirds of the enterprises that discussed model selection criteria cited capability as the primary reason for their choice, not cost. Enterprises chose the model that could deliver results fastest, not the one that cost least.

This pattern makes sense when viewed through total cost of ownership rather than inference price alone. One retail company initially built a custom solution on top of a specialized vendor, then discovered it could replicate the same functionality using a general-purpose proprietary model at lower total effort.

> *"We ended up redeveloping the same code in Claude. And we cancelled the contract and we have our own proprietary solution."*
>
> - Executive, Retail Company

Several organizations built their competitive moat not around model-choice but around proprietary data accumulated over years as was discussed in Chapter 9.

Data sovereignty, often cited as a driver toward open-source and self-hosting, was resolved differently than expected. Rather than deploying local models, enterprises negotiated contractual protections with cloud providers. Sovereignty was addressed through contracts and data minimization, not through model architecture.

The cost advantage of open-source at the inference level is real, but enterprises we spoke with are not optimizing for inference cost yet today. They are optimizing for time to value, operational simplicity, and risk reduction. The pattern may look different for startups, where inference cost is a more binding constraint from the outset and open-source adoption tends to be higher. **This will likely change in the near future. With token-hungry agent implementations on the rise, inference cost will likely be the *primary* factor in model choice**. As the capability of small and open-source models closes the gap with frontier models, cost and technology independence factors will see them play an increasingly dominant role in the model selection process.

CASE STUDY

Customer Support at a Technology Company

How they built model-agnostic infrastructure

Communications Technology | Customer Support | Enterprise

The Challenge

Customer support volume was growing faster than their ability to hire agents. The technical challenge was not building a chatbot. It was building infrastructure that could route queries to the optimal model, optimize for cost/accuracy/latency, avoid vendor dependency, and improve as new models became available.

The Solution

Rather than selecting a single model provider, the company built a multi-LLM gateway that abstracts away model choice from the application layer. The gateway routes each query based on four optimization dimensions: cost, latency, relevance, and accuracy.

> *"We don't use just one. We use Claude and we use OpenAI and we use some Llama. So we use different models. We also sometimes use Bedrock."*
>
> - Head of Operations, Technology Company

The Results

Ticket deflection

82%

Resolution rate

71%

Agent productivity

40%+ improvement

Support headcount

32% reduction

Key Lesson

This case illustrates why model abstraction matters more than model selection. The business outcomes above were driven by automating customer support at scale. The multi-LLM gateway primarily optimized cost and latency within that system, and helped the company avoid three architectural traps:

Vendor dependency. They can adopt improvements from any provider without rearchitecting.

Cost-optimization. They optimize per query rather than making a single global model choice.

Future-proofing. As models improve, the infrastructure can absorb those technology and cost improvements automatically.

For routine customer support queries, any frontier model works. For complex queries requiring deep search or high accuracy, more capable models are routed in. The system optimizes continuously without requiring anyone to make a strategic "model choice" decision.

Conclusion

This research began with a deceptively simple question: what actually happens when enterprises deploy AI in production? After studying 51 successful implementations across 41 organizations, 9 industries, and 7 countries, the answer is more nuanced and more actionable than prevailing narratives suggest. The technology works. The challenge is everything else.

Perhaps the most counterintuitive finding is what the work is actually about. AI is new and exotic so, many executives enter AI projects expecting the technology to be the hard part. In practice, the majority of the hardest challenges had nothing to do with the technology. They were about understanding the opportunities, redesigning processes, earning trust from skeptical teams, and building the data infrastructure that allows a model to operate in a real business environment and leaders to measure the results. For 42% of cases, the model itself was fully interchangeable. The organizations that succeeded did not necessarily have better AI. They had better process and execution.

That execution followed a recognizable pattern. Executive sponsors who stayed through failures, not just successes — in *every case* we could track, the same executive who oversaw a failed attempt led the one that worked. Iterative development that delivered working software within weeks. And deliberate strategies for managing resistance, not only from frontline workers fearing replacement, but from Legal, HR, Risk, and Compliance departments, which accounted for the biggest segment of project resistance. Overcoming these functions required mandates tied to corporate OKRs, not persuasion.

The human oversight question produced one of the research's most practically relevant insights. Escalation-based operational models, where AI handles 80% or more of work autonomously and humans review only exceptions, was associated with 71% median productivity gains compared to 30% for more traditional operating models, where human approval was required on every output. This does *not* mean less oversight is universally better. Regulated industries and high-stakes decisions require human review by design. But for high-volume, recoverable tasks, organizations that gave AI more autonomy achieved dramatically better results.

The employment picture is more complex than either optimists or pessimists suggest. Reduction was the most common headcount outcome at 45%, but not the majority. Alternatives, including hiring avoidance, redeployment, and explicit decisions to maintain headcount all played prominent roles. In

many cases, new types of value creation, not simply cost avoidance, were the key to sustainable business value. These patterns, however, reflect responses in an early adoption phase. As model capabilities and agentic frameworks mature, economic forces will likely push the market towards increasing labor substitution but it will inevitably also produce some new opportunities for augmentation.

The playbook that emerges from data

Start with the invisible and intangible work. Process documentation, data access layers, and change management are not overhead tacked on to the real work. They often *are* the real work. Organizations that treated these as prerequisites rather than afterthoughts reached production faster and achieved higher returns.

Invest in measurement. Clear KPIs should be identified before deployment. Organizations with strong metrics are significantly more likely to demonstrate value and scale up their projects. Key indicators include metrics of quality, customer value, and revenue growth that go beyond headcount reduction or costs savings.

Save everything. Even messy, incomplete, or seemingly useless data has value now that LLMs can clean, structure, and extract meaning from unstructured sources. Organizations that hoarded data, even imperfect data, found themselves with a compounding advantage once the models caught up. The cost of storing data is usually negligible compared to the cost of not having it when the right use case arrives.

Build a multi-model architecture from day one. The most successful implementations treated models as interchangeable components within an orchestration layer they controlled. Route each task to the optimal model based on cost, accuracy, privacy and latency. Use small models for classification, large models for reasoning and planning, open-source for specialized functions or regulated industries, and proprietary for industry specific capabilities. The organizations that built this flexibility early avoided vendor lock-in and captured the rapid improvements from providers automatically.

Plan for agentic AI. The productivity gap between agentic and non-agentic implementations, 71% versus 40% median gains, will only widen as models improve. Organizations that build the infrastructure for autonomous workflows now, including clear decision boundaries, structured escalation, and multi-system data access, will be positioned to capture the next wave of value. Open-

source models will grow in importance in agentic implementations as controlling inference costs is significantly easier and more predictable.

The broader picture

The productivity J-curve is precisely what our data captures: heavy early investment in integration, process redesign, and change management before the returns materialize. At the company level, the competitive dynamics are visible. While everyone has access to the same models, that gap between leaders and laggards is widening. That's because for every company that has redesigned its workflows around AI and begun capturing its benefits, there is a competitor still debating which model to use or struggling with organizational issues. The agentic implementations that represent 20% of our sample today will likely represent the majority within three years. Models will become cheaper and more capable. The organizations that built multi-model architecture, invested in data infrastructure, and developed the muscle for continuous process redesign are not just ahead. They are compounding their advantage with every iteration.

We are at a "productivity fork" in which the macroeconomic outcome depends on whether organizations use AI to create new tasks and augment workers, or primarily to cut costs and reduce headcount[2] Our data shows both paths being pursued simultaneously with success, and it's still unclear with way the balance is shifting as the technology matures. How organizations and governments navigate this fork in the coming years will shape whether this transformative technology delivers broadly shared prosperity, or concentrated gains and societal instability.

It is plausible that government programs and policies to retrain or support displaced workers will be increasingly necessary in many developed economies given the proportion of exposed jobs in those countries. Even though we are highly optimistic about the potential for AI to create massive value in the economy, the transition will not be smooth, as was the case with previous general-purpose technologies.

The window for experimentation is closing. The question is no longer whether AI will deliver value. It is whether organizations can evolve fast enough to capture it and what's the social responsibility of organizations to help soften the transition for workers and communities as efficiencies are realized. The stability of the economy and our social fabric may depend on how today's leaders answer this question.

Appendix

The preceding chapters analyze how enterprises are creating value with AI. This section translates those findings into two practical references: the indicators organizations are measuring and the failure patterns they are learning to avoid.

These are not theoretical frameworks. Every KPI and every failure mode below was reported by at least one of the 51 implementations in our sample.

Key Performance Indicators by Function

Organizations that define clear KPIs before deployment are significantly more likely to demonstrate value and secure continued investment. Yet many teams default to a narrow set of *efficiency-focused* metrics — often measured by *headcount reduction* — while overlooking indicators of quality, customer value, and revenue growth that often prove more sustainable and impactful over time. We all know the saying, you get what you measure, and with AI enabled projects, this is especially true. We hope these options for KPI measurements may help your organization find new ways to realize value beyond raw productive efficiency.

KPI	What It Measures
Customer Support	
Ticket / Call Deflection Rate	Share of inquiries resolved by AI without human agent involvement
Average Handle Time (AHT)	Total time to resolve a customer interaction, including talk time and after-call work
Self-Service Resolution Rate	Share of AI-deflected interactions that are actually resolved, not just deflected
Customer Satisfaction Score (CSAT)	Customer-reported satisfaction with support interactions
Support Headcount Reduction	Reduction in support staff enabled by AI automation
Sales	
Sales Rep Time Saved	Daily time freed for sales reps to focus on high-value selling activities
Lead Discovery Speed	Time to identify and research potential leads compared to manual process
Conversion Rate	Rate of converting prospects into paying customers

Tool Adoption Rate	Share of sales team actively using the AI tool on a recurring basis
Engineering	
Development Time Reduction	Reduction in engineering hours to complete development tasks
Task Completion Speed	Time to complete individual coding or migration sub-tasks
Team Size vs. Output	Change in engineering team size relative to delivered output
Code Quality Score	Success rate and accuracy of AI-generated code outputs
New Product Offerings	New or unplanned products that emerge from AI implementations
Marketing	
Campaign Time-to-Market	Time from briefing to campaign launch
Content Production Cost	Cost of creating marketing content and campaign materials
Click-Through / Conversion Rate	Customer engagement and purchase behavior driven by AI-generated content
Personalization Scale	Ability to create individualized campaigns vs. batch segments
Legal & Compliance	
Document Review / Drafting Time	Time to review, draft, or process legal documents
Document Processing Volume	Number of documents processed and searched within a given period
Information Retrieval Accuracy	Accuracy of extracted information from legal documents with source validation
Procurement	
Waste / Stock-Out Reduction	Reduction in inventory waste and out-of-stock incidents through demand optimization
Cost of Goods Sold Reduction	Savings through better negotiation, supplier matching, and timing
Processing Time Reduction	Time to process procurement requests from intake to purchase order
Finance Operations	
Processing Accuracy	Percentage of invoices or transactions processed correctly by the AI system
Staff Reduction / Cost Savings	Headcount and cost savings from automated financial processing
Backlog Elimination Speed	Time to process transactions and eliminate processing backlogs

HR & Recruiting	
Screening Time Per Role	Time to screen the full candidate pool for a given role
End-to-End Recruiting Efficiency	Overall productivity improvement across the entire recruiting pipeline
Candidate Conversion Rate	Rate of screened candidates who convert to successful hires
IT Operations	
Operating Cost Reduction	Cost savings from automating IT support and internal operations
Staff-to-System Ratio	Number of humans required to manage AI-automated systems or robots
Technician Self-Sufficiency	Ability of field staff to resolve issues without escalating to support
Field Service	
Data Gathering Time	Time to collect all technical data needed for customer issue triage
SLA Achievement Rate	Percentage of customer issues resolved within the agreed SLA timeframe
Healthcare	
Clinical Documentation Time	Time clinicians spend on documentation vs. patient care
Revenue Cycle Time	Time from service delivery to payment receipt
Coding Accuracy Rate	Agreement between AI-suggested billing codes and doctor-approved codes
Insurance Operations	
Claims Processing Efficiency	Reduction in repetitive task time for claim handlers

Common failure modes and how to overcome them

Across our interviews, 61% of AI implementations experienced at least one significant failure before reaching production value. Common symptoms like “projects stuck in pilot” or “inability to prove ROI” appeared frequently, but these are consequences, not causes. The table below consolidates failures into six root causes, shows how they manifest, and captures how companies overcame them.

Root Cause	% Cases	Manifests as	How companies overcame It
The organization wasn't ready to adopt	**35%**	• Pilots stall and never scale • Low usage despite deployment • No internal champions	• Secure visible CEO mandate tied to OKRs • Frame AI as removing repetitive tasks, not replacing people • Empower junior ambassadors to bypass resistant middle management • Deliver structured training on specific use cases — not just tool access
Critical knowledge was never captured or stored	**27%**	• Model gives generic or incorrect answers • Output quality below what an experienced employee would deliver • Users lose trust and stop using it	• Build accessible data architecture before starting any AI project • Make knowledge documentation a prerequisite, not an afterthought • Use AI itself to extract and structure tacit knowledge from employees
Legal or compliance teams blocked the project	**18%**	• Project delayed months waiting for approvals • Use cases restricted to low value safe zones	• Engage legal early as partners, not last minute gatekeepers • Implement PII scrubbing, redaction, and audit trails from day one • Build risk and controls processes before they are demanded
Technology broke or wasn't mature enough	**16%**	• System fails at production scale • Costly rework cycles • Users lose trust after seeing errors	• Build modular frameworks that absorb rapid tech evolution • Use hybrid approaches: 80% technology, 20% human refinement • Start with dual model validation before trusting single outputs
Wrong problem chosen or unrealistic expectations set	**14%**	• Solution looking for a problem • Leadership kills project prematurely	• Map processes end to end and find real bottlenecks first • Validate use cases with end users, not just executive sponsors • Set expectations that most AI projects fail on the first attempt • Frame success as iterative improvement, not perfection on day one
Talent or sponsorship gap	**12%**	• Slow iteration, vendor dependency • Project loses priority when champion leaves	• Create dedicated data science roles — do not just retrain existing staff • Secure sponsorship across multiple leadership levels • Build internal capability so progress doesn't depend on one person • Document wins continuously to maintain organizational commitment

Research sample profile

Our 51 case studies draw from 41 organizations across multiple geographies. The table below presents each organization in anonymized form.

41 organizations · **7** countries · **5** regions · **1M+** combined workforce

Organization	Sector	Functions	Region
Manufacturing company	Manufacturing	Recruiting, Compensation & Benefits	Asia
Consumer goods company	Other	HR shared services	Asia
Logistics company	Other	Training	Asia
Technology company	Software & Technology	HR support	Asia
IT security services firm	Software & Technology	IT Operations	Asia
Insurance company	Financial Services	Claims processing, document review	Europe
ERP consulting firm	Financial Services	Finance operations	Europe
Manufacturing company	Manufacturing	Supply chain & ops	Europe
Call center	Other	Customer support	Europe
IT services company	Other	Finance operations	Europe
Supermarket chain	Retail	Procurement	Europe
Talent management platform	Software & Technology	Recruiting	Europe
Semiconductor company	Manufacturing	Sales	Global
Global consulting firm	Other	Construction procurement	Global
Digital bank	Financial Services	Software development	Latin America
Stock exchange	Financial Services	Software development, compliance, customer support	Latin America

Organization	Sector	Functions	Region
Energy trading firm	Other	Customer support, marketing	Latin America
Food delivery & retail platform	Retail	Customer support, marketing	Latin America
Insurance software company	Financial Services	Software development	North America
Retail bank	Financial Services	Customer support	North America
Medical device manufacturer	Healthcare	Sales	North America
Senior care provider	Healthcare	Patient monitoring	North America
Hospital	Healthcare	Clinical documentation, revenue cycle	North America
Law firm	Legal Services	Document review	North America
Global automobile manufacturer	Manufacturing	Customer support, IT operations	North America
Industrial robotics company	Manufacturing	IT operations	North America
Semiconductor storage manufacturer	Manufacturing	Field service, product development	North America
Industrial manufacturer	Manufacturing	Quality assurance	North America
Energy utility company	Other	Marketing content generation	North America
Packaging logistics company	Other	Cold chain monitoring	North America
Online learning platform	Other	Software development, content drafting	North America
Professional services firm	Professional Services	Due diligence, recruiting	North America
Consulting firm	Professional Services	Document review	North America
Retail company	Retail	Marketing content generation	North America
AI-powered BPO	Retail	Sales automation	North America
Retail company	Retail	Procurement	North America

Organization	Sector	Functions	Region
Contact center	Software & Technology	Customer support	North America
Mobile ad attribution company	Software & Technology	Marketing	North America
Technology company	Software & Technology	Customer support, sales	North America
Telecom company	Telecom	IT operations	North America
Telecom operator	Telecom	Marketing content generation	North America

Endnotes

[1] Erik Brynjolfsson, Daniel Rock, and Chad Syverson, "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies," American Economic Journal: Macroeconomics 13, no. 1 (2021): 333–372.

[2] Erik Brynjolfsson and Gabriel Unger, "The Macroeconomics of Artificial Intelligence," IMF Finance & Development, December 2023.

[3] Daron Acemoglu and Pascual Restrepo, "Automation and New Tasks: How Technology Displaces and Reinstates Labor," Journal of Economic Perspectives 33, no. 2 (2019): 3–30.

[4] Daron Acemoglu, "The Simple Macroeconomics of AI," NBER Working Paper 32487, April 2024.

[5] Erik Brynjolfsson, Bharat Chandar, and Ruyu Chen, "Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of AI," Stanford Digital Economy Lab Working Paper, 2025.

[6] Erik Brynjolfsson, Avinash Collis, W. Erwin Diewert, Felix Eggers, and Kevin J. Fox, "GDP-B: Accounting for the Value of New and Free Goods," American Economic Journal: Macroeconomics 17, no. 4 (2025): 312–344.

[7] MIT NANDA Initiative, "The GenAI Divide: State of AI in Business 2025," July 2025.

[8] McKinsey & Company, "The State of AI in Early 2024: Gen AI Adoption Spikes and Starts to Generate Value," McKinsey Global Survey, May 2024. Survey of 1,363 participants across regions, industries, and seniority levels.

[9] Accenture, "AI: Built to Scale – From Experimental to Exponential," Accenture Applied Intelligence, 2019. Survey of 1,500 C-suite executives across 16 industries and 12 countries.

[10] Anthropic, "The Anthropic Economic Index," multiple reports, 2025–2026. Available at anthropic.com/research/economic-index.

[11] McKinsey & Company, "The State of AI in 2025: Agents, Innovation, and Transformation," McKinsey Global Survey, November 2025.

[12] OpenAI, "The State of Enterprise AI," December 2025. Based on de-identified data from 1M+ business customers and survey of 9,000 workers across ~100 enterprises.

[13] ISACA, "The Rise of Shadow AI," September 2025. IBM, "2025 Cost of Data Breach Report." AI-associated breaches cost organizations an average of $4.88M, highest of any breach category.

[14] WalkMe / SAP, "AI in the Workplace Survey," August 2025. 78% of AI users bring their own tools to work.

[15] IBM / Censuswide, "Is Rising AI Adoption Creating Shadow AI Risks?," November 2025. 80% of workers use AI; only 22% use employer-provided tools.

[16] TELUS Digital / Fuel iX, "Shadow AI in the Enterprise: 2025 AI at Work Survey," February 2025. Survey of 1,000 U.S. enterprise employees conducted by Pollfish, January 2025.

[17] Frank Nagle and Daniel Yue, "The Latent Role of Open Models in the AI Economy," MIT Initiative on the Digital Economy / Georgia Institute of Technology, November 2025. Based on OpenRouter data (May–September 2025).

[18] McKinsey & Company, Mozilla Foundation, and Patrick J. McGovern Foundation, "Open-source in the Age of AI," February 2025. Survey of 700+ technology leaders across 41 countries.

[19] OpenRouter and Andreessen Horowitz, "State of AI: An Empirical 100 Trillion Token Study," December 2025.

[20] Artificial Analysis, "Intelligence Index v4.0," January 2026. Hugging Face model repository data from AI World, December 2025.

[21] Deloitte AI Institute, "The State of AI in the Enterprise: The Untapped Edge," January 2026. Survey of 3,235 director to C-suite-level leaders.

[22] METR, "Measuring AI Ability to Complete Long Tasks", updated February 2026

[23] Labor Market Impacts of AI: A New Measure and Early Evidence." Anthropic, March 2026

Acknowledgements

Special thanks go to all the interviewer subjects who contributed their time, insights, and data to this paper. We would also like to express our appreciation to the reviewers who took helped us fine tune the content to make it significantly better: *Sandy Pentland, Andy Haupt, Jing Wang, Christie Ko, Tomas Castagnino, and Matty Smith*.