GLG

From AI to ROI: Closing the Enterprise Value Gap

Five Expert Diagnoses Across the Implementation Lifecycle

Table of contents

Introduction 3

AI for AI's Sake: Why Most Projects Never Reach Transformation 5

Enterprise AI has a Value Capture Problem 9

What Are You Comparing Progress To? 14

AI's Economics Break the SaaS Playbook 18

Why the Highest-Value AI Takes the Longest to Materialize 23

Conclusion 27

Introduction

Worldwide spending on AI is forecasted to reach **$2.59 trillion in 2026**. A 47% increase from the year prior, the staggering sum looms not only in its magnitude, but for the fact that it will mostly bypass the underwriting rigor that normally applies to capital investments in almost any other category. AI expenditure seems to be skipping steps.

AI expenditure seems to be skipping steps.

Many organizations already feel the impact of that investment, attributing tangible results to their AI initiatives. In McKinsey's 2025 "State of AI" survey, for example, **64 percent of respondents said they already see innovation gains from AI**, with real cost and revenue benefits showing up at the individual use-case level in software engineering, manufacturing, marketing, and sales.

But as money streams into the space, the intense speed of investment, technological development, and pursuant change is widening a gap between AI's material effects inside individual functions and what management can confidently claim about value or revenue at the enterprise level. McKinsey found in the same survey that **only 39 percent of organizations can attribute any enterprise-level earnings impact** to their AI investment, and for most of those who can, the impact remains under 5% of total EBIT. MIT's Project NANDA put a sharper point on the same finding, reporting that **roughly 95 percent of enterprise generative AI pilots show no measurable effect on profit and loss**. While the ability to measure at the initiative level is mixed, the ability to prove ROI at the enterprise-wide level is clearly lagging.

To understand what closing that gap will require, GLG spoke with five preeminent experts in AI implementation, each with a different vantage point on the problem:

- **Matt Baker**
  Former SVP Corporate & AI Strategy, Dell Technologies
- **Mallika Iyer**
  AI Infrastructure Practitioner; Former Google Cloud AI & CoreWeave Engineering Leader; Founder, Arya AI
- **Divya Ashok**
  Founder, Altera Ventures; Former COO, Salesforce AI Research
- **Dr. Ryan Mackey**
  CMIO & AI/RCM/CDI Physician Executive, Lifepoint Health
- **Yiannis Antoniou**
  CTO & Head of AI and Data, Lydatum

Between them, the experts have built AI strategy inside a Fortune 500 enterprise, built the data infrastructure on which financial services firms run their AI, sold and bought enterprise AI as both vendor and customer, deployed AI across a hospital system's clinical and revenue-cycle operations, and built AI products from inside a leading tech company's engineering organization.

Posed with the same direct question in a conversation with GLG's own **AI Moderator** ("Where does AI implementation break down between the technology working and its financial value being provable at the enterprise level?"), all five experts agreed on the gap's significance, but diverged on its causes and implications.

Together, the five perspectives trace the arc of an AI implementation end-to-end, each isolating a different point along that path where the break with provable value most often occurs: some upstream, before a project is scoped or a contract signed, others downstream, once outcomes are already being measured.

This report should help leaders diagnose what's most likely behind an organization's inability to chart and prove AI's financial returns, and to start building the measurement framework that captures that return on the page.

AI for AI's Sake: Why Most Projects Never Reach Transformation

By Matt Baker

Former SVP Corporate & AI Strategy, Dell Technologies

GLG

Today's AI cycle reminds me of the Big Data and Hadoop era. We used to see this notion of doing cloud for cloud's sake, a mindset that led many organizations to pilots that never showed value.

A lot of people continue to approach AI in a similar way, and the value conversation remains similarly immature. The vast majority of AI projects today are AI projects for AI's sake. Leaders match AI to arbitrary acts of innovation, fixating on the power of the technology and throwing it at a part of the business. They may see some benefit at a low level, but the overall process doesn't meaningfully change.

That's where the measurement problem actually starts. It has very little to do with mathematics or metrics and more to do with whether the intent of a project has a strong business focus versus a technology focus. That distinction is crucial to not only making any AI effort successful, but to being able to demonstrate its impact.

In essence, success comes down to intent, focus, and mindset – a business transformation mindset over a technology implementation mindset. Specific business transformations lend themselves to demonstrable goals and outcomes that can be defined upfront and measured.

What hundreds of AI projects taught me

At Dell Technologies, we catalogued more than 800 different projects during the early days of generative AI excitement, all held up effectively as "AI doing something interesting." In the end, all 800 turned out to be small, partial views of what might generate value. But when we decided we were going to change the nature of how we did customer support, or change how we do software development overall, those larger transformation efforts were both more successful and produced far more tangible, holistic, and measurable impacts.

The holistic mindset primes you to make better predictions about what changes might yield. Rather than measuring the implementation of any given AI project in isolation, you're considering a profound transformation to the business. An all-in transformation like that brings expectations for changes that will affect every layer of the business, including its people, tools, and process. You expect a business outcome, and you ultimately plan to measure one.

> "The holistic mindset primes you to make better predictions about what changes might yield."

Some of the best results I've seen came from work that focused on a discrete process managing the overall radiology process and workflow within a hospital, where implementation freed up the time of a radiologist through automation. The approach wasn't merely focused on implementing an AI tool relevant to the field. Instead, the team set out to redesign the entire process across the full patient experience. That yielded a major time benefit in the patient receiving their X-ray results and being consulted by a radiologist. The hospital saved a tremendous amount of costs, and the patient got a result much faster – and was probably healthier for it.

Such meaningful, measurable impacts contrast to simply changing the mean time to read a given chest X-ray, and measuring that isolated step outside the totality of the process. The benefit is vague when you only measure the impact on one member of the overall workflow.

Buy-in correlates to pressure

Transformations take time. Organizations with short reporting cycles and high pressure to produce results quickly are liable to bear spurious results. The typical gap between the cost of an AI implementation hitting the books and the financial benefit revealing itself can't be measured in a single quarter.

> "Organizations with short reporting cycles and high pressure to produce results quickly are liable to bear spurious results."

Pressure to rush results usually signals not only a misunderstanding of AI's potential in most enterprises, as well as the challenges to reaching it, but also poor alignment or weak buy-in from the start. When an effort is rightfully viewed as an overall organizational transformation, there's less board pressure or renewal pressure, because at that point everyone's bought in and understands what's happening. Where things get difficult is when a leader promises to spend a fixed amount on a limited, AI-fueled capability, and stakeholders start asking for results. The pressure builds because the intent was never clear, and no one aligned on what results are supposed to look like or when they're meant to reveal themselves. When a board or CEO says, "we must AI all things," that's a dangerous position for an organization to be in.

Leaders should expect a well-scoped process transformation to take at least a year. It starts with a diagnostic phase, e.g. what is the process, how many steps, and how many people? Before beginning the transformation, you grapple with the existing business process to get a deep understanding. Then you move to testing the hypothesis in a constrained implementation.

Major transformations unfold over multiple years. What I worked on at Dell was seen as a multi-year effort from the start, targeting many different processes on different timeframes, with a clear understanding of how they would build on one another.

When everyone buys into a transformational mindset, it means they understand they're going to face a disruptive, costly process together. If everyone in the organization is well informed about the significance of the work and the time commitment, there won't be as much pressure for updates. Regular updates will be built into the overall effort as an inherent component.

> "When everyone buys into a transformational mindset, it means they understand they're going to face a disruptive, costly process together."

Start with the process, not the tool

My advice for leaders seeking real results from AI is straightforward: Make sure you're undertaking a broad business transformation effort. Otherwise, AI will just fuel an out-of-control process that already exists. The overall process reinvention should yield significant upside with or without AI – the AI should accelerate and multiply that upside. Value comes from examining and changing the way you do business with the new tools available, not from taking an AI tool and shoving it into an existing business process and hoping for good things to happen.

Looking back at the cloud example, it took a long time to demonstrate value, and that value didn't end up being cleanly financial – it wasn't always cheaper to implement cloud computing. The same dynamic can play out with AI. The way to close the gap between expectation and reality comes back, again, to a transformational mindset. Winners will focus on business transformation and on how this new class of technology can radically accelerate the upside of any change, rather than on the power of the technology itself.

Matt Baker
Former SVP Corporate & AI Strategy, Dell Technologies

Learn more

Enterprise AI has a Value Capture Problem

By Divya Ashok

Founder, Altera Ventures;
Former COO, Salesforce AI Research

GLG

More than a measurement problem, enterprise AI has a value capture problem. AI can work exactly as intended without creating business value – the models work and adoption improves while the enterprise still struggles to connect that with how the business changed.

> "What economic value are we looking to create, and how will the organization capture it?"

Enterprise leaders need to answer a critical cross-functional question: What economic value are we looking to create, and how will the organization capture it?

That value will differ by usecase. The objective may be to create capacity, reduce churn or risk, generate incremental revenue, reduce hiring, or simply return time to employees for higher value work. Each can be a legitimate outcome, but each will require a different definition and measurement of success.

This is why I believe every large enterprise deployment will need an AI value contract – an agreement to align business, technology, finance, and vendors on the economic value being pursued, the current baselines it will be measured against, what needs to change in the business to realize that value, and which executives are accountable for its delivery.

The vendor/customer disconnect

One place where the value contract breaks is between vendors and customers. Heading into a renewal or a board update, respectively, vendors and customers often bring different ideas of success. The vendor wants to demonstrate adoption and justify renewal, while the customer needs to show business impact. The two stories are related, but they usually sound different – and rarely do stakeholders reconcile them before signing a contract.

The vendor's story tends to focus on adoption rate, agent transactions, and CSAT score. But the customer's board wants to hear a slightly different story, oriented around where the P&L moved. A lot of renewal friction exists between those two framings. A vendor might model the value as 50,000 hours saved at $100 an hour, equaling a $5 million productivity figure. It would be a reasonable capacity estimate, but it doesn't account for the cost of the platform, the integrations required, the human exception handling, or the other layers of the AI stack that most customers initially underestimate. The harder question is what happened to that capacity and how much value the business ultimately captured.

> "The two stories are related, but they usually sound different – and rarely do stakeholders reconcile them before signing a contract."

AI can create capacity. Management needs to determine what that capacity becomes, such as reduced hiring, greater conversions, more service calls handled, or employees getting thinking time back.

The internal version of the problem

Inside the customer organization, a similar misalignment plays out between the teams using AI day to day and the executives evaluating it. Technical and functional teams tend to focus on and measure the day-to-day productivity. For example, a report that used to take an hour to write now takes fifteen minutes, or a BDR has an extra hour because an AI delivered account research in just a few minutes. As material as these improvements may be, they don't always translate into business outcomes, like revenue, margin, or lower cost.

Leaders should read this as a sequencing problem more than an indictment of any productivity gains. After the AI buildout, which can be time-consuming on its own, employees need to use the tools at scale before the organization sees any economic value. They also need time to learn how to use the tools and build meaningful workflows before the freed-up capacity becomes something that appears on a financial statement. The lag should be expected and not necessarily taken as the implementation failing.

> "Leaders should read this as a sequencing problem more than an indictment of any productivity gains."

The gap also exists because we entered the AI era without clean baselines. If you didn't measure the process before, you can't measure the change after implementation. Establishing agreed-upon metrics upfront – hard outcomes, not proxies like number of prompts or tokens processed – distinguishes a deployment that can demonstrate its value.

Start with the outcomes, then establish the baseline

When organizations orient around outcomes rather than usage, the value story gets clearer. Many organizations entered the AI era without clean baselines. If you didn't measure it before, it becomes difficult to determine what changed. The metrics also need to reflect the outcome, not simply the activity. Prompts, tokens processed, agent transactions, and adoption measure usage, not economic value.

> "When organizations orient around outcomes rather than usage, the value story gets clearer."

Salesforce's deployment of Agentforce serves as a simple example. AI handled millions of customer interactions, allowing support staff to focus on the more complex issues. The resolution rates improved over time, and

escalations to humans declined. Management could trace a fairly clear chain from AI activity to financial impact.

Our internal deployment at Salesforce also yielded lessons about the nuanced nature of the net gains that come with implementation. Early on, we saw hesitation among employees, who weren't sure how to use the tools, and the gains were modest. Once adoption took hold, we found we could redeploy people into other parts of the enterprise. This didn't show up as a reduction in headcount or a line item on a spend report. The real net benefit was that we could avoid external hiring by reallocating existing employees toward higher-value work. That not only reduced costs and improved margins, but also improved the customer experience because the people being redeployed already understood the business and its customers.

There are plenty of other ways to conceptualize AI value based on signals that sit even closer to financial outcomes than hours saved. Take for example incremental revenue from increased developer or sales capacity, reduced manual touches, higher throughput, lower customer wait times, churn reduction, etc. The important thing is to define the outcome and measure the baseline prior to deployment.

Build the AI value contract before deployment

The key to closing this gap is to define the value contract before initiating the implementation, not during it. That process should always involve three steps.

1. **Define the value** – an explicit agreement on the metrics that count, whether those are hard P&L figures, capacity measures, customer experience or time returned to employees. Be explicit about the outcome being pursued.

2. **Baseline the current state** for the specific workflows you're targeting as clearly as you can. This is what makes it possible to measure not just whether the AI is working, but how much of an improvement it produced and at what cost. It may not be perfect but it needs to be credible.

3. **Name the executives who will own the outcome**, and give them organization-wide empowerment to follow through. In most organizations, this seems to be the hardest part, but it's crucial. With vague accountability spread out across committees and task forces, the implementation ends up being evaluated on adoption metrics because nobody with enough authority made the call to measure anything harder.

I do see the industry trending in the right direction. Renewal and investment conversations increasingly focus on harder metrics, rather than adoption evidence. Was the hiring requirement altered; did conversion rates improve for sales teams; was churn reduced; was cost per transaction lower? Vendors and customers alike are getting more sophisticated about which models to use and which tools to avoid, thinking about efficiency in the transaction itself, not just its outcomes.

Not every valuable outcome will show up on the P&L. Ultimately, much of the value will continue to live in many organizations' blind spots, a big one being what it really means for employees to get significant time back from routine work. The ability to redirect human attention toward work that requires genuine judgment, creativity, and relationships – the things AI still can't do – is value that won't take shape on a spreadsheet. Value is created when the organization decides the outcome it wants, changes the business in a way that produces that outcome and follows through until the value is realized.

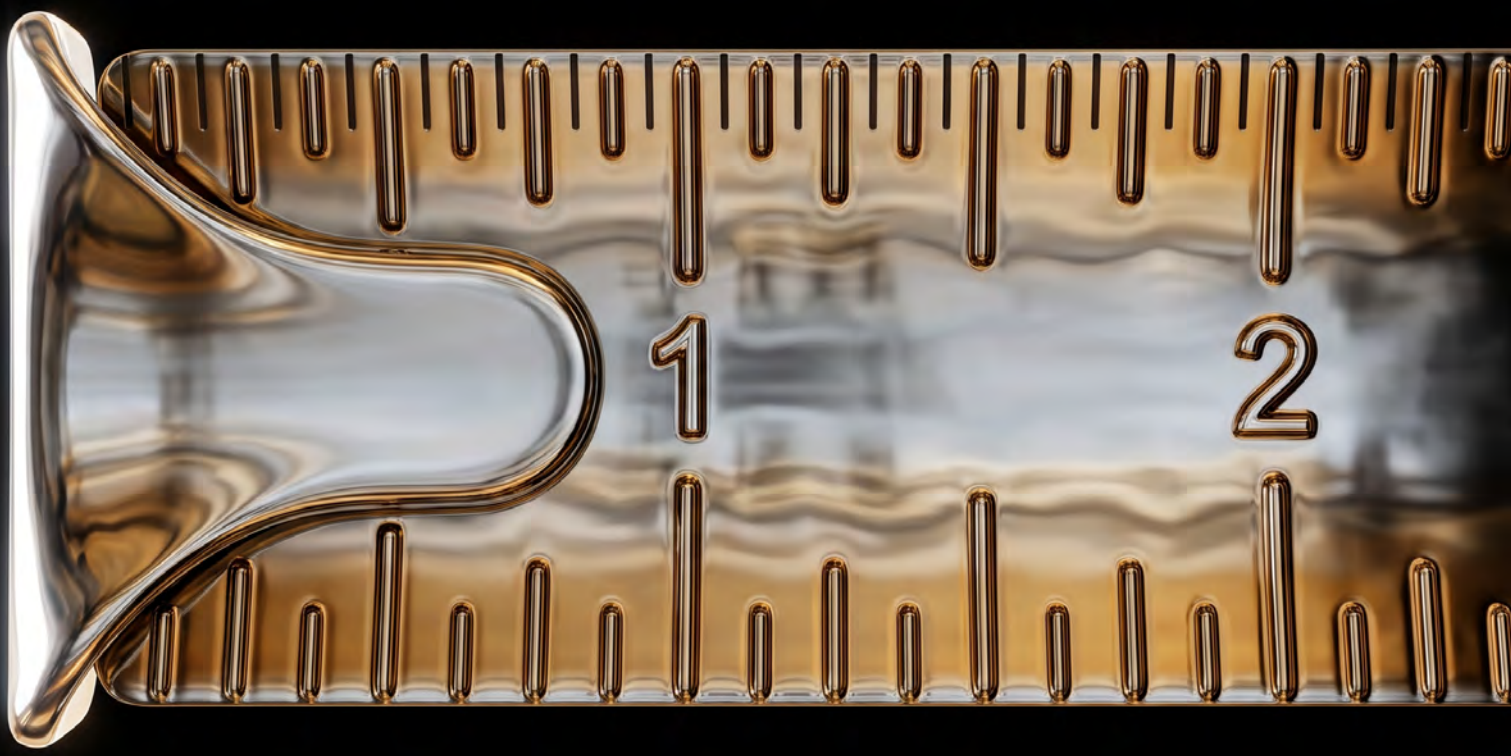
Divya Ashok

Founder, Altera Ventures;
Former COO, Salesforce AI Research

Learn more

What Are You Comparing Progress To?

By Yiannis Antoniou
CTO & Head of AI and Data, Lydatum



GLG

The key breakdown I see most often for organizations implementing AI has little to do with the extent to which they can prove the technology "works." That's the easy part. The team builds the technology, and sooner or later it effects some kind of change and "works" to some end. Where most organizations struggle to prove value is in their difficulty answering another basic question: What are you comparing progress to?

I've seen so many projects stall or fail simply because the implementation team never even tried to capture a snapshot of the before era. In such cases, the motivation behind implementation usually boils down to vague interest in the novelty and raw power of the technology, and an urgent sense of pressure to "modernize." Rarely do leaders in those environments seriously ask themselves, "What business outcome am I trying to improve?" If you can't say so before the build, then you can engineer the most advanced product, and it still won't prove anything by way of value.

> "The motivation behind implementation usually boils down to vague interest in the novelty and raw power of the technology, and an urgent sense of pressure to 'modernize'."

Take an investment research example, where a firm has a team of analysts covering a particular sector but only has the capacity to review 5% of the companies in the space. "We implemented an AI research tool" will never be a business outcome for that team. It would have to sound more like, "With the same number of analysts, in the same amount of time, we reviewed 10x as many companies." That's a clear, measurable before-and-after that you can agree on upfront, track against, and point to when someone asks about the investment's payoff. Most organizations still aren't thinking in those terms, though, which is exactly why we're talking about this gap today.

Why the baseline doesn't exist

In most situations I encounter with clients, the lack of a baseline appears when technology teams and business teams don't work hand in hand. If it's an engineering-driven project with its own timeline and origin story, there's often no connection to a business outcome whatsoever. And then even in well-run organizations where the business and technology teams are aligned, the business side often still hasn't thought about measuring a baseline. It can be as simple as "This task takes us X amount of time, and we want to reduce it by Y."

But I've found the root of the problem runs deeper than organizational misalignment. In almost every client environment I walk into, the underlying data that would be needed to establish a baseline just doesn't exist

> "The underlying data that would be needed to establish a baseline just doesn't exist or isn't standardized enough to trust."

or isn't standardized enough to trust. It's common for companies to overestimate the quality and governance of their data estate. You can't have a baseline when you don't have common definitions of what a "customer" or a "product" is – or when you have five different definitions depending on where the data originated and how it's stored.

AI has especially brought this vulnerability to the fore, because the investments aren't paying off in organizations where the data estate is disconnected, ungoverned, owned by different people, and doesn't conform to shared definitions. AI actually gives organizations a pivotal opportunity here – albeit one that may be forced upon them – to modernize and govern their data estate, both to use it as contextual input for AI and agents, and to build a foundation that a real baseline can actually measure against. More and more vendors are catching on to this issue, investing in the semantic or contextual layer, to help AI understand data not just in technical terms but in a real business context.

Organizational misalignment adds another layer to the data problem. Technology teams and business teams don't always see eye to eye, nor do they work with the same incentives. An engineering-driven project can be technically excellent while also being organizationally invisible.

It gets worse when organizations, under pressure to prove something is working, try to shoehorn some measure of progress in the absence of a real business baseline. They tend to lean on adoption metrics to fill the vacuum (e.g., 80% of potential users engaged with this tool last month), which then get mistaken for evidence of financial impact. Adoption is surely important, and so is engineering excellence, but neither directly demonstrates value. Fundamentally, organizations need a commonly understood business goal that AI – or any other initiative, for that matter – will help them reach, broken down into clear components, with metrics attached and agreed upon before the build begins. Internal adoption rates and engineering milestones just aren't enough.

> "Adoption is surely important, and so is engineering excellence, but neither directly demonstrates value."

I'd also push back on the notion that adoption itself primarily represents an organizational problem. If people get real value out of something, adoption should follow organically. If it's forced top-down – "We have to use AI, so everybody will now use AI" – that's a sign there was never a business case to begin with.

The early red flags

In my experience, certain signs appear very early in a build, even in the first week, that tell you the value case is going to be difficult to prove later, regardless of the technology's performance.

First, if you run a quick demo or discovery session with stakeholders and at the end of it they still can't say what success would look like, that's a serious red flag. Uncertainty about timelines or technical approach is normal, but an inability to articulate what "this worked" means in business terms represents a grave problem.

The second one is subtler: If improvements depend on behavioral changes on the part of teams that aren't involved in the project, that is a clear red flag. If your metric requires a particular team to change their process in order to use an AI capability, but that team hasn't been part of the discussions and doesn't have skin in the game, you're building in a closed bubble and taking on dependencies you can't control. You can create the technically "perfect" system only to discover that you're relying on other teams to adopt it or change their processes, even though they were never involved and have no reason to prioritize it. I've seen this pattern end many implementations that were otherwise well-executed.

What to ask before you greenlight anything

Any leadership team should be able to answer two questions before they spend a single dollar on a major AI implementation, whether using internal or external resources.

First off, what am I trying to achieve as a business outcome, and what are my measurable success criteria? In crisp business terms, why are you spending this capital relative to everything else you could be spending it on, and what would the metrics actually look like at the end? If you can answer these questions clearly, you either get a clear mandate or you realize you don't have one and can stop and regroup before wasting any resources.

Secondly, are the right people at the table? Organizational alignment and real sponsorship require genuine buy-in from the teams whose behavior and processes will need to change. If you don't have that before you start, you'll have to struggle for it later, at a point when it's much more expensive to resolve.

Those two questions sound simple. In my experience, the organizations that can answer them clearly before a build starts will reap value, and be able to prove it.

Yiannis Antoniou
CTO & Head of AI and Data, Lydatum

Learn more

AI's Economics Break the SaaS Playbook

By Mallika Iyer

AI Infrastructure Practitioner;
Former Google Cloud AI & CoreWeave
Engineering Leader; Founder, Arya AI

GLG

Many enterprise leaders still walk into an AI implementation as though they're in the market for SaaS. With software, once you buy the license and pay for the build, the marginal cost of each subsequent use is negligible, while value reliably compounds over time. The economics of AI aren't so simple. Every inference request draws on new context, consumes GPU time, and incurs a cost, which means the technology could be simultaneously working well and costing more money. Organizations that overlook this phenomenon tend to read their cost curve as a failure signal when it's more a usage signal, a confusion that can cascade into important measurement decisions down the line.

Herein lies the most underappreciated reason for enterprise AI's ROI gap today. Organizations are measuring against baseline assumptions that don't apply to AI. It's not surprising many leaders still cling to the SaaS precedent given how crucial it has been in shaping how they evaluated successful technology investments for over a decade. But leaders need to develop new models to find and prove value in the AI era.

> "Organizations are measuring against baseline assumptions that don't apply to AI."

Why KPI mapping breaks down

Traditional software deployments lend themselves to relatively clean attribution, with legible causal chains. For example, if you replace a manual sales process with a CRM, you can trace inbound call volume, response times, and ARR prediction accuracy, and chart improvements in those metrics to specific software features.

AI outputs, on the other hand, aren't deterministic (i.e. the same input doesn't reliably produce the same output every time), so you can't directly tie a KPI improvement to an AI-enabled feature. While this reality presents measurement challenges, it's also what makes AI valuable. The model draws on contextual information to generate something new rather than executing straightforward, fixed instructions. But as a result, any cost-benefit analysis that assumes linear KPI attribution will produce a skewed picture. An ROI case built on the assumption that AI improvements map to clean attributions won't hold up to close scrutiny for long.

> "AI outputs, on the other hand, aren't deterministic so you can't directly tie a KPI improvement to an AI-enabled feature."

Furthermore, companies tend to measure AI against OKRs and KPIs that were designed for a different kind of technology deployment. The metrics aren't wrong per se, they're just not calibrated for the latency and variability that characterizes AI implementations. What you

should measure instead is whether the outcomes associated with the AI are shifting, over what time horizon, and at what total cost, which then brings in the cost of ownership question.

The total cost of ownership problem

When organizations plan for AI deployment, they typically apply a cost model revolving around how much it costs to host the technology in the first year, and then maintenance projections for years two, three, and four. While that model may suit traditional cloud-based applications, it systematically undercounts the cost of running AI in production.

A sound plan must account for model drift, for example, in which a model's performance degrades over time as the distribution of inputs it encounters migrates from its original training. Staying competitive and relevant for your specific use case means continuously evaluating, fine-tuning, and periodically upgrading the model. Human-in-the-loop costs can therefore shift, and in some cases expand as the model matures. Building evaluation harnesses and test harnesses for the edge cases a model handles poorly introduces more ongoing work and costs. Leaders regularly overlook these factors in total cost of ownership, assuming AI automates everything and therefore presents a SaaS-style cost structure after the initial deployment.

As they think about what they're getting, organizations also often make the mistake of treating an array of outcomes as simultaneously achievable. There's a tendency to bucket all expected returns into a single ROI figure, like cost savings, revenue uplift, freed capacity, or risk reduction. In practice, gains on one aspect usually come at the expense of another. A revenue uplift play tends to consume additional capacity and require more human oversight. Freed up capacity doesn't automatically reduce cost unless you've made explicit decisions about what to do with it. These nuances warrant specificity upfront about which outcomes you're actually optimizing for, rather than assuming the model will deliver across all four at once.

What productivity gains actually mean

The assumption that time saved equals money saved or new revenue probably accounts for the single most common measurement error in enterprise AI. It manifests most awkwardly in the board room when someone asks where an investment is showing up on the P&L.

Consider a code generation example. Augmenting software development with an AI coding agent will almost always increase headcount initially, because you need more "humans in the loop" to evaluate the code the model generates before you have a clear picture of its edge cases. The time the AI saves on generation then gets offset by the time humans spend reviewing output they didn't write. Gradually, as you develop a clearer picture of the model's weak spots and build

"The assumption that time saved equals money saved or new revenue probably accounts for the single most common measurement error in enterprise AI."

evaluation tooling around them, the implementation velocity accelerates, but on a curve that resembles an increased cost before it looks like savings. Organizations that measure only the early portion may draw the wrong conclusion.

A similar dynamic plays out in customer support, albeit in a very different way. A well-deployed agentic AI with sufficient context and clearly defined guardrails can reduce headcount and cost while simultaneously reducing customer churn. But that outcome assumes the implementation gets the guardrails right early, which, while possible, shouldn't be taken for granted. More typically you see some cost increase before it comes down as the system is tuned and the edge cases come to the surface.

When it comes to margin improvement, leaders should view high usage as a prerequisite, not evidence of success. If a user is interacting with an AI model for something that generates no economic benefit on either side (like a conversational loop that doesn't solve a problem the user couldn't otherwise solve on their own) then higher usage translates to higher hosting cost, not higher margin. Margin improves when two things happen together: the model's productivity is high and that productivity is captured economically, meaning the model is helping solve problems that would otherwise require a human being or wouldn't get solved at all.

How to actually measure, and what to ask before you deploy

The baseline is the starting point, and it needs to be more specific than organizations usually make it. Once you have a clear before, you can measure a real after. If you're replacing a human customer support agent with an agentic AI, how many cases were resolved, in what timeframe, at what cost, with what churn rate? Did resolution time change? Did churn increase or decrease? What was the mean time to resolve escalated cases that the AI couldn't handle? The metrics need to track the entire workflow, not just the step the AI touches.

> "I recommend a best practice that I hardly see any organization do: Play devil's advocate against the deployment decision itself."

I recommend a best practice that I hardly see any organization do: Play devil's advocate against the deployment decision itself. If a team comes to leadership arguing that AI needs to be deployed to counter a competitor gaining ground in a particular domain, the right counter-question should be, "what do we actually lose if we deploy this 30 days later?" What is the realistic cost of waiting, versus the cost of deploying in an area that may not generate any meaningful economic benefit because the ROI case wasn't properly scoped? If the implementation idea can clear that bar, it has a good chance at proving value.

Few leaders want to go down that road, though, because the honest answer is often that nobody knows for sure what would be lost. That uncertainty feeds the FOMO that drives a lot of AI deployment into use cases that are low-stakes enough to feel safe but too low-value to ever show up on a P&L. The companies that avoid that trap are willing to get extremely granular before

deployment about exactly how they expect an AI-enabled aspect of a particular application to impact productivity, revenue, or operating cost – and what the cost of that impact actually is, including the human cost of working alongside the AI and recovering from its mistakes.

It can be uncomfortable to demand that level of granularity, and it's always hard to produce. At the same time, it's also the only thing that makes the ROI question answerable.

Mallika Iyer

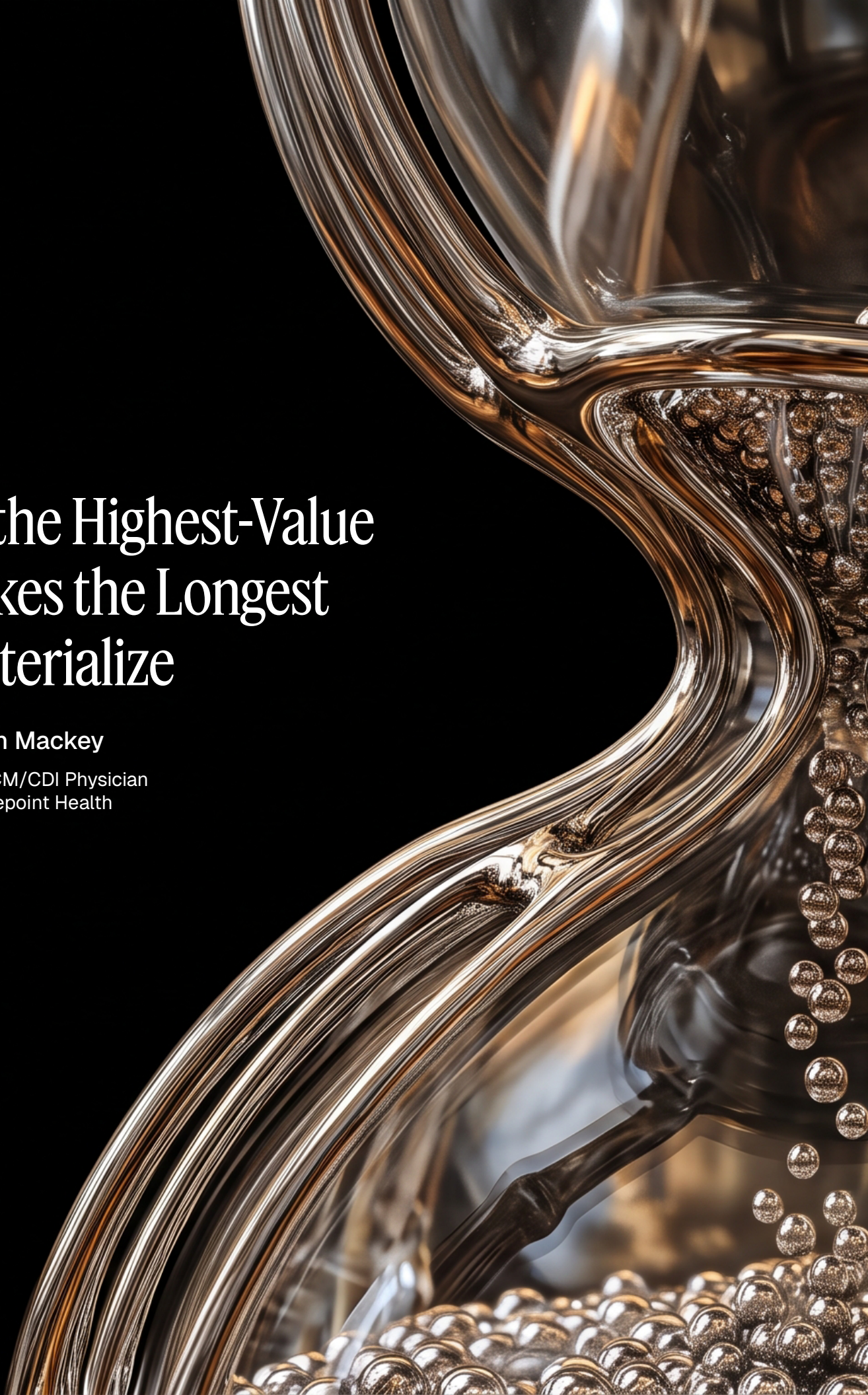
AI Infrastructure Practitioner; Former Google Cloud AI & CoreWeave Engineering Leader; Founder, Arya AI

Learn more

Why the Highest-Value AI Takes the Longest to Materialize

By Dr. Ryan Mackey

CMIO & AI/RCM/CDI Physician Executive, Lifepoint Health



GLG

Any realistic conversation about AI ROI in healthcare begins with an understanding of three layers within a healthcare system: clinical, financial, and operational (or administrative). The AI success story sounds slightly different for each, and takes shape through specific metrics in different timeframes.

In the financial and operational layers, the story is relatively straightforward. AI applied to revenue cycle management (RCM) should streamline billing processes and reduce denial rates. On the operational side, patient engagement improves, compliance goes up, and day-to-day tasks run more efficiently. The proof points may arrive quickly and present clearly on a dashboard. Healthcare systems are implementing AI aggressively in these areas right now, where they can make an ROI case with some confidence.

It's a different story on the clinical side, where outcomes are more subjective and may bear out over a longer timeline. But it's where I see AI ultimately having the most dramatic positive impact. Understanding why it's harder to measure is essential for any leader in the space who hopes to set realistic expectations around AI investments.

What makes clinical different

While clinical AI outcomes may be as tangible as healthier patients, they're harder to isolate and quantify. Quality of patient care, clinician burnout, and earlier detection rates don't map neatly onto a billing code or a throughput metric. Seeing more patients or coding at a higher level are not the sole ROI for achieving true clinical efficacy and impact from AI.

The causal chains are long, the end users more varied, and the care settings ever changing. Clinical AI touches outpatient settings, emergency and inpatient care, and alternative care environments, each of which moves at a different pace, where lower complexity tends to yield faster proof points. A straightforward outpatient documentation tool will likely demonstrate value faster than an AI system designed to support diagnostic reasoning in an emergency department. Any sound implementation plan should account for that variation inherent to the paradigm of clinical work.

While it's harder to measure, clinical AI presents the biggest opportunities for positive transformation in healthcare. The direction of the evidence is already encouraging, demonstrating strides on important, high-potential fronts like earlier detection, improved care methodology, and more meaningful patient engagement. But the models still have a lot of maturing ahead of them. Machine learning systems improve through feedback loops, which take time to accumulate enough signal to meaningfully shift outcomes. Until there's broader consensus and awareness around AI utilization in clinical settings, some areas will feel much further along than others. The highest-stakes applications, namely acute care complexity and direct patient-impact, will take the longest to develop.

The staff capacity reality

One of the more persistent misconceptions about healthcare AI is that it creates an efficiency vacuum, where the same work gets done with fewer people. A closer look at how healthcare systems are actually staffed reveals flaws in that assumption.

> "What AI can do, importantly, is pull people back toward the work they're actually trained and equipped to do, and away from the administrative and operational burden they've absorbed out of necessity."

Healthcare isn't sitting on surplus staff that AI can displace. The average healthcare staffing model already stretches individuals across three or four roles when they used to cover one or two. There's not much reason to expect the labor crunch in healthcare to go away anytime soon. What AI can do, importantly, is pull people back toward the work they're actually trained and equipped to do, and away from the administrative and operational burden they've absorbed out of necessity.

When AI takes on the tasks that staff were never meant to own, it creates an opportunity for redeployment, as opposed to redundancy. Clinicians get back time for the patient-facing, judgment-intensive work that demands their expertise. Administrative staff can focus on exception handling and escalation rather than routine processing. AI should fit into the well-defined gaps, and more importantly, offer staff members oversight, control, and feedback impact to ensure models keep evolving.

How implementation actually works

The timeline between adopting an AI tool and seeing it move a metric that leadership cares about varies significantly by the layers I mentioned above. On the clinical side, a tool that's quick to provision and adopt can impact end users demonstrably in three to four weeks. RCM implementations take closer to 90 days, because the AI needs to ingest historical data before it can prove itself, and then has to be integrated into existing processes before the improvements compound. Once it's running, though, the RCM lane tends to operate with more consistency, with fewer end users touching it, less subjective feedback to incorporate, and cleaner iteration cycles.

Pilots are succeeding more quickly, which suggests the value case is becoming easier to make in every layer. Before AI, technology pilots in healthcare often stalled in extended evaluation phases – organizations were cautious about integration risk and unsure about outcome validation. With AI, there's more confidence going in, partly because the ROI mission is clearer and because the problems being solved are ones that healthcare systems know are essential. A pilot that succeeds in one market often promptly moves to five or six more.

> "Pilots are succeeding more quickly, which suggests the value case is becoming easier to make in every layer."

What enables that acceleration is a combination of local champions, clear milestones, and vendor partnership. Markets need engaged clinical or operational leaders

who can manage expectations, surface feedback, and advocate for the implementation internally. Milestones tied to real data keep the effort honest and give leadership meaningful check-ins rather than vague progress updates. Meanwhile, vendors need to show up not just to sell products but as scalability partners, helping manage the tool as it grows, evolving their features in response to what the data reveals, and staying integrated into the journey beyond the initial launch.

Setting expectations that hold

For a leadership team about to greenlight a major AI implementation, the mindset should be less about technology selection and more about organizational readiness. You need an engaged team that spans corporate leadership and market-level champions – clinical, operational, or both – who are part of the mission from the outset, not brought in for counsel after the decisions are made. Without markets in the decision-making process, implementations tend to arrive lacking genuine buy-in, and fall flat.

You need well-defined milestones and a clear sense of what you're measuring in each layer (financial, operational, and clinical) because conflating the varying proof timelines sets everyone up for disappointment. And you need the patience to let the clinical models mature, which means conveying to internal staff that AI isn't meant to take away from what they do, but make them more effective at what they do best.

The transformative impact of AI in healthcare will happen in clinical care delivery. The financial and operational wins are already arriving, but they represent just a foundation for what's coming over the next 12-18 months. Clinical AI will take longer because the end user landscape is so varied, as are the care settings. But as the feedback loops compound and the models mature, the opportunity to touch every care setting, every clinician type, every dimension of patient experience and outcome becomes more plausible.

Healthcare systems that maximize their AI journey will be the ones that tempered expectations early enough to keep their implementations intentional rather than reactive, and who stayed the course long enough to see the value materialize.

Dr. Ryan Mackey
CMIO & AI/RCM/CDI Physician Executive, Lifepoint Health

Learn more

Conclusion

Individually, each of these diagnoses isolates a single point where an AI implementation can lose its ability to prove value. Together, they explain why doing so at the enterprise level has become a compounding challenge rather than a single hurdle to clear: the definition of success, the baseline it's measured against, the cost model applied to it, and the time horizon allowed to prove it out each introduce an independent way for a technically sound deployment to still fall short of demonstrating value. Clearing one is no guarantee of clearing the rest, which is what makes enterprise-level ROI a materially harder question to answer than whether the technology itself is working.

There is no silver bullet for enterprise AI implementation. Unlocking its transformative potential will require time, resources, and the right mindset – a mix of ambition, humility, and creativity, from the C-suite down to the end user. The process of proving the investment paid off will reflect the complexity and magnitude of the undertaking. Every organization, team, and leader faces unique hurdles, and stands to win on different terms.

GLG is the world's leading expert network, and has a deep bench of senior-level AI experts on a range of topics including infrastructure and compute, foundation models and agentic systems, enterprise adoption and transformation, industry-specific applications in financial services and healthcare, and AI risk, policy, and investment strategy.

Closing the AI ROI gap looks different at every organization. GLG can help clients bring in the expertise they need to navigate these complex challenges at every stage of a company from **Deal Advisors** to embedding experts within a portfolio company or corporation in a fraction role, to targeted **Board Placements** that extend an advisory engagement into a long-term role. AI experts are also available to produce customized workshops and trainings, or speak at internal or external event panels and keynote addresses.

Reach out to GLG to learn more.