ISSUE #6 - JUNE 2026

EVALUESERVE

ELEVATE

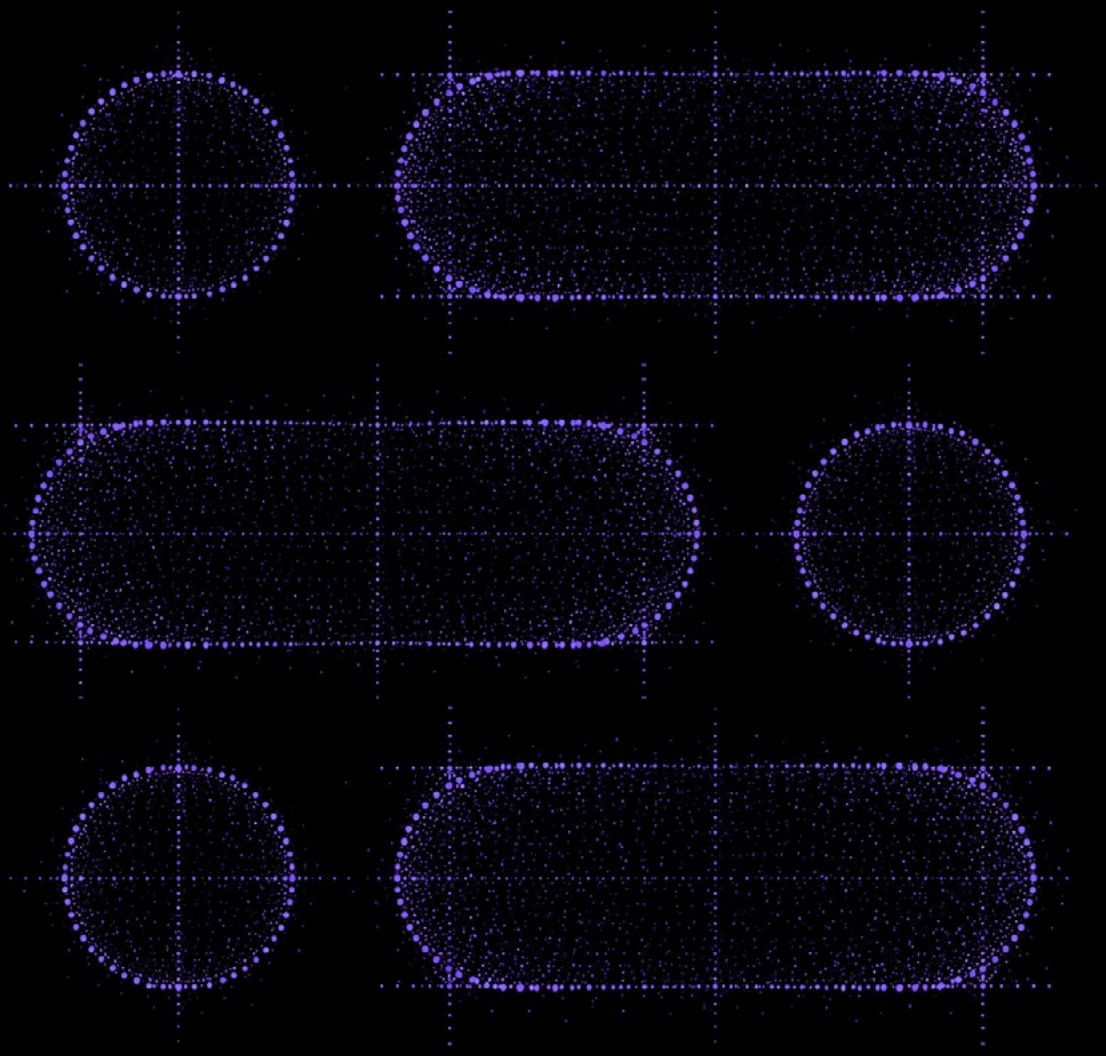
CLOSING THE GAP BETWEEN AI ACTIVITY AND AI IMPACT

MEET OUR CEO: PALLAB DEB

THE REAL RACE BEGINS AFTER THE PILOT

THE TECHNOLOGY IS READY. ARE YOU?

REBUILDING TRUST IN SHOULD-COST MODELING WITH AGENTIC AI



Contents

• Editorial Note

Closing the Gap Between AI Activity and AI Impact

This gap is not a technology problem. It is a matter of orchestration.
And the companies closing the loop are not moving faster; they are moving smarter.

A year ago, AI inside most companies looked impressive from a distance. There were demos, pilots, and a handful of use cases made for great internal presentations. Tools that could answer questions, summarize reports, and promise a transformation that felt just within reach.

Here is where things stand today.

88% of organizations are using AI in at least one business function. But only 23% have scaled AI agents across the enterprise. Within any specific function, that number falls to single digits. Meanwhile, only 6% of organizations qualify as genuine high-performing companies, with AI contributing meaningfully to business outcomes.

As global investment surges toward $2.5 trillion in 2026, most organizations remain stuck in a kind of in-between state: aware of AI's potential, but unsure how to embed it into the messy, everyday reality of how work gets done.

> *Part of the challenge lies in how quickly technology has evolved. When generative AI first entered the enterprise, it lived largely in the hands of engineers. Then came copilots, and soon after, agents, and with that, the number of possible use cases multiplied. But a large implementation did not follow.*

Despite widespread adoption, as 78% of firms are using AI and 75% of SMBs are experimenting, only a fraction have moved beyond isolated use cases. Most remain stuck in a fragmented, "half-seeing" state, where AI exists, but impact doesn't scale.

When every new idea had to pass through a technical team, scale would always lag ambition. So, a different approach began to take shape: place the tools in the hands of the people who understand the work best. Let analysts, researchers, and operators build, test, and refine AI agents themselves, with just enough support from engineering to guide, not gatekeep.

This is the story of that shift.

The Pilot Trap has a Number Attached to It

The average organization scrapped **46% of its AI proofs of concept before reaching production.** In 2025, 42% of companies abandoned most of their AI initiatives — up from 17% the year before.

> *AI doesn't fail at the technology layer. It fails at the operating model layer.*

An MIT study found that 95% of AI pilots delivered no measurable P&L (profit & loss) impact. Over 80% of organizations have piloted tools widely, and nearly 40% report deployment, but these systems mainly boost individual productivity rather than delivering measurable business transformation.

The pattern is consistent: technology gets deployed, individuals get faster, and the organization gets... roughly the same results. Because AI dropped into an existing workflow does not transform that workflow. It just automates the surface of it.

What is actually missing is the layer underneath: the process redesign, the feedback loops, the governance, the human checkpoints that turn an AI output into a decision an organization can stand behind.

The Missing Layer is Orchestration

Think about how a fraud detection system actually gets better over time. It is not due to an algorithm alone.

It improves because investigators confirm or reject its alerts, and those decisions get fed back into the system. The AI accelerates the detection. Humans define what fraud actually means in context. Neither works without the other.

That dynamic holds across every domain where AI is delivering real value.

In financial services, AI models can scan thousands of earnings reports in minutes, surfacing signals at a scale no human team could match. But it is human analysts who validate, interpret, and contextualize those insights before decisions are made (who determine what qualifies as an "investment-grade" insight versus noise). In research and intelligence workflows, AI can process in hours what once took weeks. But it is the people who understand the client's actual decision, the competitive landscape, the nuance behind a dataset, who determine whether that output is useful or misleading.

Remove human judgment too early, and systems lose context. Leave humans buried in manual work, and you never get to scale. Human judgment becomes more important, especially in moments of ambiguity, exception, or risk, when people can provide the layer of context and accountability that machines cannot. This is why organizations that pair AI with human oversight are consistently more likely to outperform their peers.

But even that is not enough.

For AI to deliver sustained value, it must be continuously challenged and improved. Over time, AI must adapt to new data, shifting conditions, and changing business needs. This means moving from a static tool to more of a living system.

The Efficiency Ceiling and What's Beyond it

Efficiency is the easiest AI story to tell and the least interesting one to stop at. It is only the first layer of impact. The real opportunity lies in using AI to improve decision quality, uncover new insights, enable faster and more adaptive strategies, and create entirely new revenue streams.

The more consequential opportunity is what becomes possible when AI is embedded deeply enough in decision-making to change the quality of decisions, not just the speed at which they are produced. When intelligence can be synthesized from sources that no human team could realistically monitor continuously. When pattern recognition operates across datasets at a scale that shifts what questions are even worth asking.

Only **15% of organizations** have achieved enterprise-scale AI deployment, despite near-universal experimentation. The gap between those 15% and everyone else is not the sophistication of their models. It is the strength of their operating model: the discipline with which they designed workflows, integrated domain expertise, and built the feedback mechanisms that make AI outputs something a team can rely on.

That gap is widening. And it is closing faster for organizations that treat AI as a system to build, not a tool to deploy.

From Building Tools to Designing Systems

Here is the distinction that separates companies with impressive demos from companies with measurable outcomes: a tool does one thing. A system learns, adapts, and improves the more it is used.

AI does not behave like enterprise software with a fixed spec. It behaves like a living system that depends on continuous input, active human oversight, and alignment with changing business conditions. Remove any of those elements, and performance degrades. Models drift. Trust erodes. Adoption quietly stalls while everyone pretends the pilot is still progressing.

We've spent 25 years inside enterprise workflows. Today, that expertise directs the AI, not the other way around.

The difference becomes clear in practice across five disciplines:

- **Use case design** does not start with what can be automated, but what should be. High-impact use cases are selected based on domain fit, process complexity, and organizational readiness — not theoretical potential. A use case that looks compelling in a pitch but lacks a clear business owner, defined data sources, and a measurable outcome is not ready to scale.

- **Human-AI workflow** embeds AI into existing processes, with explicit roles for both humans and machines at every step. Teams are trained not just to use AI, but to work alongside it — knowing when to trust the output, when to challenge it, and when to override it entirely.

- **Continuous quality assurance** treats deployment as the beginning of the work, not the end. Outputs are monitored against domain-specific benchmarks: in financial research, that means validating insights against the frameworks portfolio managers actually use; in knowledge-intensive industries, it means preserving the standard of rigor clients expect — before the speed AI makes possible.

- **Governance and observability** mean control mechanisms are built into the system from the start, not bolted on after something goes wrong. Leaders need visibility into how decisions are being made, where risk is emerging, and how performance is changing. As AI moves from generating answers to taking actions, firms need strong controls around permissions, auditability, human oversight, and data access.

- **Lifecycle management** recognizes that AI capabilities are not products with a ship date. They are ongoing systems that require updates, retraining, and deliberate expansion as business conditions evolve. Organizations that treat AI as a one-time deployment are the ones watching their pilots fade into irrelevance 18 months later.

What becomes clear across all five is a truth that runs counter to the prevailing assumption: automation does not reduce the need for human involvement. In practice, the opposite is true. Systems perform best when humans are embedded within them — defining decision frameworks, handling exceptions, validating outputs against domain standards, and continuously improving system performance.

Companies will look back on 2025 and 2026 not as the years they launched the most pilots, but as the years they asked the hardest questions: Who owns this workflow end-to-end? What data can the system actually access? Where does human approval happen? How do we audit the output? How do we know whether people are genuinely using it and whether it is changing how they work?

Evalueserve is already making this shift — treating AI not as a series of disconnected projects, but as a system that underpins decision-making. Not a tool to be deployed, but a capability to be built, refined, and scaled. And above all, a partnership between human judgment and machine intelligence, each strengthening the other rather than competing for control.

That requires a different kind of investment: the design of workflows, the integration of domain expertise, the mechanisms that ensure adoption, and the discipline of continuous improvement. Enterprises need to build with an end in mind, designing how AI will be used, not just how it will be built.

The first wave of AI adoption was defined by experimentation. The next will be defined by integration. Progress demands aligning strategy with execution, technology with domain expertise, and automation with human judgment.

That is the shift. And the organizations that make it intentionally will be in a different competitive position entirely.

Author:

Marcela Cortez
Editor & Senior Associate
Evalueserve

Meet our CEO, Pallab Deb

"AI Isn't the Answer. Knowing What to Do With It Is."

You've spent years at the intersection of technology and enterprise adoption. What's the most common mistake you see companies making with AI right now?

Pallab: They're bolting it on. That's the honest answer. They see a capability, they get excited, they run a pilot, and the pilot works. Then nothing happens at scale. The reason is almost never the technology. The technology is extraordinary right now, and it's only getting better. The reason is that they haven't figured out how to get past the friction that's built into every enterprise: data that isn't ready, security concerns, and the hard work of actually transforming the business.

Enterprise environments are brownfield, not greenfield. You've got processes that have existed for 15 years, technology investments that are deeply embedded, and a culture that won't change overnight just because a new tool shows up. If you walk in with AI but you don't understand that context, you'll produce a very impressive demo and very little else. Change management is critical here; without a deliberate plan for how people, processes, and incentives adapt, even the best technology will stall out against the weight of existing habits and systems.

So what does real operationalization look like?

Pallab: It starts with the domain. Before you even talk about technology, you have to understand what the business is actually trying to do. What does an end-to-end workflow look like? Where does it break down? Where is human judgment irreplaceable, and

where is it just friction? You can't answer those questions from a distance. You have to be in it.

The companies that are winning right now are the ones that combine deep domain expertise with the ability to bring the best technology to bear on very specific problems. And then, critically, they stay. They're not delivering a solution and walking out the door. They're sticking around to make sure the value actually materializes for their customers, and then they're building from there.

That domain expertise matters most in the edge cases. The clean, happy path is easy to automate. It's the strange exception, the account that doesn't reconcile, the file stuck in dispute, the situation nobody mapped, that breaks things. If you're not running the process every day, you don't know where it actually breaks. We do, because we've been doing this work across hundreds of companies and thousands of workflows. Those edge cases are the difference between a demo that works and a system that works in production.

The engineering has to match that ambition. It's not enough to build something that works once. You have to build agentic workflows that can scale, adapt as the underlying models evolve, and keep driving the process forward rather than just responding to it. The infrastructure has to be supported over its lifetime. And increasingly, the commercial model has to reflect that, priced on the outcomes the workflow actually delivers, not just on the build. That's a different kind of commitment. It's also the only model that makes sense if you're serious about the value materializing.

But here's where it really takes off: when you can do this in a repeatable way. The breakthrough isn't solving one workflow once. It's solving it faster, better, and more cost-effectively the next time, because you bring assets to the table that already address the domain challenge. Take spreading in commercial lending. Rather than rebuilding the solution for every client, we bring something like Spreadsmart that accelerates how spreading gets done from day one. That's the difference between a one-off project and a capability that compounds.

From right: Pallab Deb, Evalueserve's new CEO, with Evalueserve Co-Founder Marc Vollenweider

You've talked about a three-legged stool. Can you walk through that?

Pallab: The first leg is domain expertise. You need a deep understanding of the customer's industry, their business model, their processes, and what they care about. Without that, you're going to show up with a hammer looking for nails.

The second leg is technological depth. You need to be current, genuinely fluent in what's available and what's coming, and you need to have built that into platforms and solutions so you can move fast. Time to value-realization matters. No one wants a twelve-month engagement before they see results.

The third leg is trusted relationships. The transformation decisions are being made by lines of business and CXOs. If you don't already have a relationship there, if they don't already trust you, you're starting from behind. That trust takes years to build. It's one of the most durable competitive advantages a company can have.

Most companies have one or two of those legs. Very few have all three in the same place.

There's a lot of pressure on CFOs and finance teams right now to justify AI investment. How should they be thinking about it?

Pallab: First, let's name the conundrum they're sitting in. There's a belief taking hold that if you're not spending heavily on tokens, you're not doing enough. Jensen Huang has made a version of this argument: if an engineer costs $500,000 and you're only spending $250,000 on tokens for them, you're underinvesting. I don't think that's the right way to look at it. Token-maxing is not a proxy for value creation. It's pulling the conversation toward tokenomics and away from business value, which is the only thing a CFO actually cares about.

So reframe the question. Stop asking, "What is this AI tool going to cost?" and start asking, "What is this workflow costing us right now?" The business case for AI transformation almost always lives in that second question.

When you map out what it costs to run a process at the current level of quality, speed, and error rate, the math usually becomes very clear, very fast.

What the CFO really wants to know is simple: when I bring in AI, and knowing what AI costs, is the combined cost below what I'm paying for the human-only approach today? And at scale, is the token-plus-human model producing more throughput than humans alone? The real question is where those cost curves intersect as you scale agents alongside people. That's the analysis worth doing.

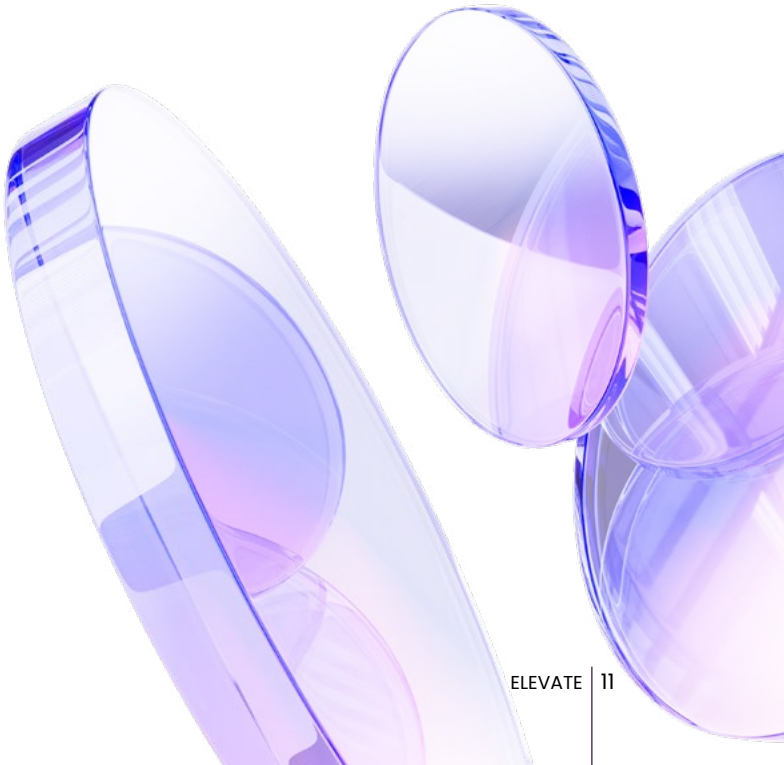
But it requires tech, business, and finance to be in the same room. The finance team needs to understand what the workflow looks like before and after. The business team needs to understand what's actually feasible. And the technology team needs to understand what outcome they're being held accountable for. When those three aren't aligned, you end up with a proof of concept that never becomes a product.

Why should a company come to Evalueserve as their partner for this work?

Pallab: The companies that get this right don't do it alone. They find a partner who can close the gap between what the technology can do and what the business needs it to do. That's a very specific capability, and it's rarer than the market would have you believe.
What makes Evalueserve different is that we come in with all three legs already built. The domain expertise is real and deep, earned across 25 years of work inside specific industries and functions. The client relationships are long-standing, which means we're not spending months getting oriented. We already know the processes, the pressure points, the history. And we're building out the technology muscle to make sure we're bringing the best of what's available to bear on the right problems.

Because we understand the business process, we take a long-term view. We're not just solving for today; we're building something that still holds up five years from now. And critically, we know where the guardrails are. In knowledge-intensive work, the cost of getting it wrong is enormous. For example, a patent filed incorrectly can cost millions. So our people always surround the technology we deploy. We can underwrite the output and account for the edge cases, because we've operated these workflows ourselves. That's the difference between a tool that's confident and a system you can actually trust.

The way I think about it: a great partner doesn't just deliver a solution. They help you figure out what the right problem is, they move fast to show you value, and then they stay to make sure it compounds. That's the kind of partnership we're here to build.



The Real Race Begins After the Pilot

Saikat Choudhury
Director

One of the key observations across pharma organizations, which is observable in other domains as well, is that organizations are getting better at building AI, but not at shipping it.

Across commercial teams, medical affairs, R&D, regulatory, market access, and digital transformation groups, there is still a shared sense that everyone is learning in real time. But there is something important to bear in mind: AI is not a side experiment or a future capability. It is becoming part of how pharma organizations think about insights, operations, decision-making, and growth.

A commercial team can spin up a working market intelligence prototype in a day. A medical affairs team can have a document synthesis tool running by Friday. The demos are impressive, the interfaces are clean, and the outputs sound convincing. And then, a couple of months later, most of it is still sitting in a sandbox waiting for approvals that never come, integrations that were never scoped, and governance questions that were never answered.

"We use AI" no longer impresses anyone.

Clients are asking sharper questions. Internal leaders expect faster results. Commercial models are under pressure. Medical and regulatory teams are being asked to do more with greater precision. The question has moved from "Should we use AI?" to "Where do we start so it actually creates value?"

In pharma, teams can build impressive AI prototypes faster than ever. Tools like Lovable, Cursor, and Microsoft Copilot Studio have made this genuinely easy. A business analyst with no engineering background can have a working prototype of a Scientific Literature Review running before lunch. That is not an exaggeration.

The demo may look polished, the answers may sound convincing, and the interface may make the solution feel almost ready. But a polished prototype does not mean a production-ready solution.

The prototype shows what technology can do. However, it does not prove that the solution is accurate, compliant, scalable, secure, or trusted enough for real-world use.

That is where the real work begins.

The Pilot Trap

Modern AI prototypes create a confidence mismatch, since the experience looks complete, and leaders often assume the remaining work is incremental. If the demo feels 80% done, the assumption is that production is only a few weeks away.

In reality, the visible demo may only represent a fraction of what it takes to operate agentic AI in a regulated enterprise environment.

The invisible work includes access controls, data security, compliance review, validation, workflow integration, observability, auditability, change management, and user adoption. In pharma, these are not administrative extras, but the conditions for trust.

Here are some examples of originally "invisible" issues with AI:

- A commercial intelligence agent that works in a sandbox but cannot connect to approved data sources will not change how decisions are made.
- A medical content agent that drafts responses but cannot support compliance review will not scale.
- A regulatory assistant that summarizes documents but cannot show what it used, why it made a recommendation, or how a human reviewed it will not pass the trust threshold.

This is why so many AI pilots stall. The technology works well enough to impress people, but the operating model was never designed to support daily use.

There is also an organizational problem underneath the technical one. In most pharma teams, there is no clearly defined owner for the path from prototype to production. The data science team builds it; IT is responsible for infrastructure; the business is the end user; and compliance and the GenAI governance board review it. But nobody owns the end-to-end journey from working demo to something that runs reliably in a production environment and is actually used by the people it was built for. That gap is the absence of a role or structure that bridges all of those functions. Complex workflows like Omni-Channel or HTA Filing require deep domain understanding as well as the capability to translate them into Agents that can augment existing processes rather than adding complexity or broken workflows.

Part of why this persists is how AI progress is measured internally. Most organizations track the number of pilots launched, while some track the pipeline of Agentic AI. Very few track the number of AI tools currently running in production and actively used. If the scorecard rewards starting things, the organization will get very good at starting things.

Agentic AI Raises the Stakes

Traditional software follows defined rules and works in environments where requirements are written, workflows are built, outputs are tested against expected behavior, and the system ships.
Agentic systems, in comparison, are more adaptive. They reason, retrieve information, call tools, interpret context, and sometimes trigger downstream actions. Their behavior emerges from the interaction between models, tools, prompts, data, memory, and human feedback. That creates new questions pharma teams must ask much earlier:

- What data can the agent access?
- What actions can it take?
- Where does human approval happen?
- How are outputs validated?
- What happens when an agent makes the wrong recommendation?
- How do we audit the reasoning path?
- How do we know whether people are actually using it?

That creates questions pharma teams are not used to asking before they build. What happens when the agent makes the wrong recommendation? In a traditional system, a wrong output is a bug with a traceable cause. In an agentic system, the wrong output may be perfectly reasonable given the inputs the agent had access to at that moment, and reproducing the exact conditions to debug it may not be straightforward. For a commercial team, that is a concern when accurate numbers need to be reported on sales, engagement, and various leading indicators, bifurcated by indication, MoA, and country, among others. Similarly, for a medical or regulatory team, such issues become a compliance event. Questions also arise around how the reasoning path can be audited. A human analyst can explain their logic. An agent that synthesized 200 documents, called three tools, and produced a recommendation needs to be able to show its work, which is a regulatory requirement for certain use cases (e.g., HTA/Clinical Protocol/SLR, etc.). Finally, one of the most important questions is: how do you validate a system whose outputs are non-deterministic? Standard Computer System Validation approaches in pharma assume you can define expected outputs for a given input. Agentic systems do not work that way, and many use cases are still being worked out on what validation looks like for these systems. Organizations that are not thinking about this now will face it as a blocker when they try to move to production.

The mistake is treating the agent as the product. In pharma, the product is the workflow. The agent only creates value when it is embedded into the real work of evidence synthesis, launch planning, medical review, market monitoring, regulatory intelligence, KOL engagement, or field force enablement.

So, a stronger approach is to place the agent that is connected to the right data, governed by the right controls, reviewed by the right experts, and measured against the right business outcomes, to later become an operating infrastructure.

When we talk about architecture, Agentic AI is not a new feature sitting on top of an existing technology stack. It is a new stack.

A production-grade agentic system has several layers that need to be designed deliberately:

- A data layer that defines what the agent can see, how that data is governed, and how its lineage and quality are tracked; a model layer that handles the underlying LLMs, fine-tuned models, and embedding systems

- An agent and tool layer where single agents, multi-agent networks, and reasoning loops operate and call external tools, APIs, and data sources

- An orchestration layer that manages task decomposition, agent coordination, and memory, including short-term memory within a session, long-term memory that persists across sessions, and shared memory that multiple agents can access simultaneously

- An observation layer that logs every decision, every tool call, every data source accessed, and every output produced

- A UI and engagement layer that defines how humans interact with, review, and override the system.

Cutting across all of these layers are two non-negotiable rails: security and compliance, and governance and audit. The challenge I see most often is that organizations treat these as separate workstreams managed by separate teams, but in an agentic system, these layers are deeply interdependent. A gap in the data layer creates a problem in the observation layer, and a gap in the observation layer makes the governance layer impossible to operate.

One emerging standard worth understanding is MCP (Model Context Protocol), which is basically an agent-to-tool communication protocol that does what HTTP did for the web. MCP establishes a common interface so that agents built on different platforms can connect to tools and data sources without bespoke integration for every combination.

Focus is Key

Trying to make AI touch everything at once is exciting. Unfortunately, it usually translates into dozens of pilots, overlapping initiatives, impressive slides, and limited production impact.

This keeps teams busy, with the feeling of riding the AI wave, but in reality, value remains scattered. Pharma leaders should stop launching disconnected experiments and choose a small number of high-impact workflows tied directly to business value. These could include:

- AI copilots for competitive intelligence and market monitoring: Commercial intelligence agents that surface competitor label changes, trial readouts, and pricing moves in real time, reducing the lag between signal and strategy update from weeks to hours, and replacing the quarterly slide deck that is already outdated by the time it is presented.

- Automated literature reviews and evidence synthesis: Evidence synthesis tools that screen and extract data from scientific literature in hours rather than weeks, freeing medical affairs teams from the manual work of literature reviews that currently consume analyst time that

should be going into interpretation and recommendation.

- AI-assisted HTA filing and dossier generation: Extracting PICO-structured data, flagging evidence gaps against each body's specific threshold criteria and surfacing the indirect comparison analyses that are almost always the thing that delays a submission because the bottleneck in HTA is rarely the evidence itself.

- Medical content generation with compliance review workflows: Medical content drafting agents connected to approved sources, with built-in compliance review workflows, so that the time between response and approval is measured in hours rather than weeks.

- Predictive analytics for launch readiness and patient targeting: Launch readiness analytics that synthesize patient identification data, HCP engagement signals, and market access indicators into a single view, replacing the manual process of pulling five reports and reconciling them in a spreadsheet the night before a review meeting.

- Omnichannel personalization for HCP engagement: Analysing which clinical topics and sub-topics drive engagement versus unsubscribes by specialty and feeding those signals directly back into content generation.

Do not rank Agentic AI use cases by how innovative they sound. Rank them by whether they can survive production.

A pharma workflow is a strong Agentic AI candidate only if five things are clear from the start: the business owner, the approved data sources, the human approval point, the system it must integrate with, and the KPI it is expected to move. If a team cannot name those five things, the use case is not ready to scale.

That means moving away from broad ideas like "an AI copilot for medical affairs" and toward specific workflow bets, such as:

- Reducing the time required to screen and extract evidence from scientific literature.

- Accelerating competitive intelligence updates for a priority brand.

- Shortening launch-readiness reporting cycles.

- Improving KOL identification beyond existing relationship networks.

- Supporting medical content drafts that move through approved compliance review.

A simple test for leaders is this: Can this agent improve one measurable workflow within 90 days without creating unacceptable compliance, quality, or adoption risks? If the answer is no, keep the idea in discovery. If the answer is yes, fund it like a production initiative — with integration, governance, validation, change management, and adoption metrics built in from day one.

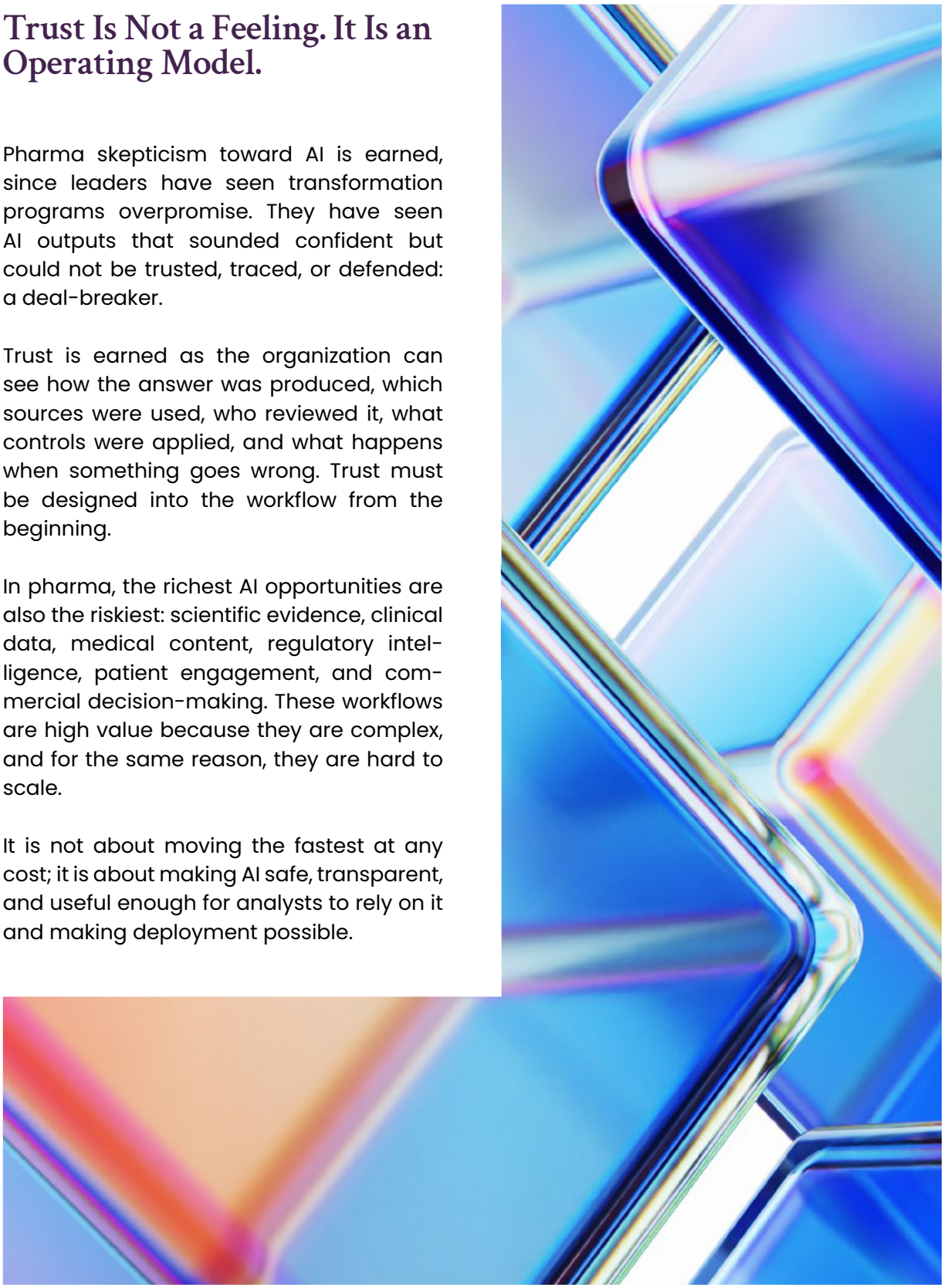
Trust Is Not a Feeling. It Is an Operating Model.

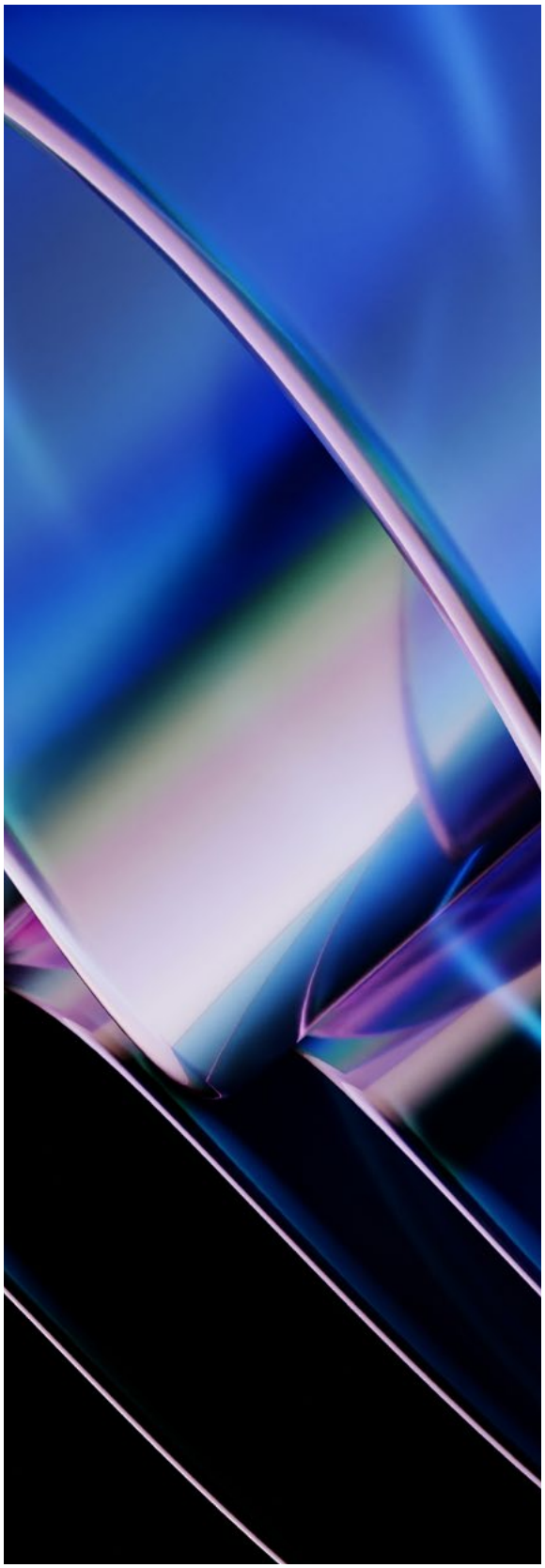
Pharma skepticism toward AI is earned, since leaders have seen transformation programs overpromise. They have seen AI outputs that sounded confident but could not be trusted, traced, or defended: a deal-breaker.

Trust is earned as the organization can see how the answer was produced, which sources were used, who reviewed it, what controls were applied, and what happens when something goes wrong. Trust must be designed into the workflow from the beginning.

In pharma, the richest AI opportunities are also the riskiest: scientific evidence, clinical data, medical content, regulatory intelligence, patient engagement, and commercial decision-making. These workflows are high value because they are complex, and for the same reason, they are hard to scale.

It is not about moving the fastest at any cost; it is about making AI safe, transparent, and useful enough for analysts to rely on it and making deployment possible.





Where Evalueserve Is Different

The mission is to help pharma use AI in ways that are useful, governed, measurable, and ready for real work. Pharma needs a practical way to turn promising AI ideas into reliable operating models.
That requires connecting three agendas that often move separately:

1. Business teams need faster insight generation, better content workflows, stronger research support, sharper analytics, and more effective decision-making.

2. CIO, CISO, and CAIO teams need governance, compliance, integration, architecture, and control. They cannot let every function build its own disconnected AI workflow.

3. CFOs need measurable impact. They need to see whether AI reduces cost per process, shortens cycle times, increases capacity, or improves business outcomes.

Evalueserve brings together domain expertise, AI-enabled workflows, human quality review, and managed delivery models to move AI beyond experimentation. The goal is not to build an agent and declare success, but to design the operating model around the agent — where it fits, who uses it, how teams govern it, how experts check quality, and how leaders measure impact.

What's Next for Pharma Leaders

The path from pilot to production is a discipline:

First, prioritize a small number of high-impact use cases. Choose workflows tied to revenue growth, speed-to-insight, operational efficiency, launch performance, or risk reduction.

Second, build cross-functional AI squads. AI adoption fails when innovation, IT, data science, commercial, medical, and regulatory teams work in silos. The right squad combines business owners, domain experts, analytics, technology, compliance, and workflow execution.

Third, design for production from day one. Before funding a pilot, ask what production would require: integration, access control, validation, monitoring, user training, change management, and measurement.

Fourth, embed AI into existing workflows. The goal should not be to create another portal people have to remember to use. The goal should be to bring intelligence into the processes where work already happens.

Fifth, measure adoption, not just deployment. A launched tool is not the same as a changed workflow. Track usage, decision acceleration, quality improvement, cycle-time reduction, workflow integration, and business outcomes.

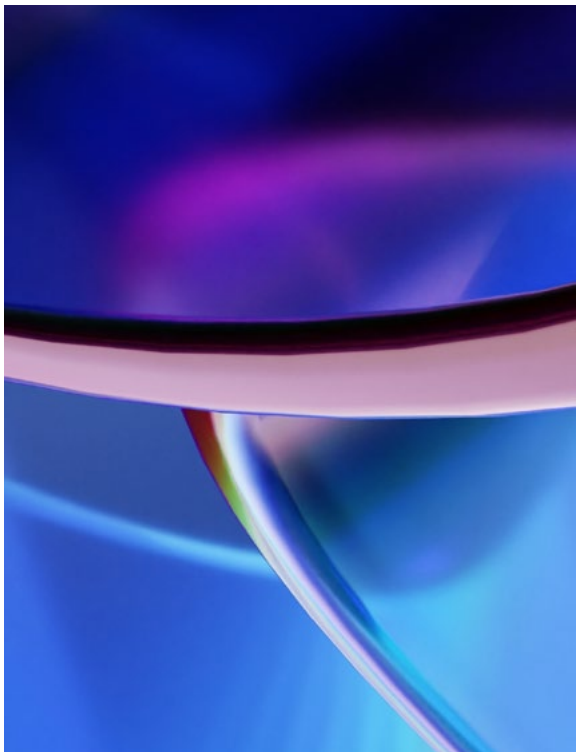
Finally, start before everything feels ready.

No organization has complete clarity; you must stay willing to learn through execution while building the governance and quality systems that pharma requires.

Not more demos, not more sandboxes, not more "AI initiatives" that never touch the operating model.

- Can a medical team cut evidence review cycles from weeks to days?
- Can a commercial team detect competitor moves before the next quarterly planning meeting?
- Can regulatory teams reduce repetitive document work without weakening control?

That's the real race.



The Technology Is Ready. Are You?

Susan Xie
Director

The numbers tell a familiar story. Billions of dollars invested in artificial intelligence, and yet productivity statistics remain largely unmoved. It is enough to invoke **Robert Solow**, who observed in the 1980s that you could see the impact of computers everywhere except in the productivity data. The gains eventually came, but only after organizations did the slow, unglamorous work of reorganizing around the technology, not just adopting it, but genuinely rebuilding processes, infrastructure, and ways of working to make it useful.

> "You can see the computer age everywhere but in the productivity statistics."
>
> - Robert Solow, 1987

We are living through a similar lag with AI. Earlier this year, there was a push to maximize usage, marked by token consumption leaderboards and incentive structures designed to get people using AI everywhere, for everything. The costs have proven significant. Uber reportedly burned through its entire annual AI budget in four months. And where that spend is going is telling. Leaked audio from Accenture revealed that one of the top enterprise use cases is converting PDFs to PowerPoint slides. LLMs are powerful, but powerful is not the same as efficient, and there are far more effective ways to convert a file.

That gap between investment and return is especially visible in financial services. What it actually takes to make AI work inside a large financial institution was the central question at a recent Executive Exchange dinner hosted by Evalueserve, where senior leaders from across banking, lending, and private capital gathered for a discussion led by **Columbia Business School Professor Stephan Meier** alongside a panel featuring Evalueserve Chief Growth Officer **Mangesh Patnaik** and Co-founder and Chief Strategy Officer **Marc Vollenweider**. Two obstacles came up repeatedly: the context gap and the perception gap. Together, they go a long way toward explaining why the productivity gains have been so hard to capture.

Unpacking the AI ROI Gap

1. The Context Gap

The first obstacle is data. In software engineering, AI tools exploded almost overnight for a straightforward reason: everything a model needs to be useful already lives inside the codebase. The context is self-contained, structured, and accessible. In financial services, the picture looks quite different. The context that powers good decisions is scattered across a CRM, an email thread, a decade-old Excel model, and a relationship manager's accumulated judgment that has never made it into any system at all.

Before AI can deliver on its promise, institutions need to give models the context they need. Sounds straightforward, but financial institutions are full of silos. Some have sprung up unintentionally, but others are deliberately constructed to protect sensitive information, and the work of navigating the complex web of structured and unstructured institutional knowledge is no small undertaking.

2. The Perception Gap

The second obstacle is organizational. Large institutions are historically resistant to change, and that resistance is not always irrational. With large-scale tech layoffs making headlines, concerns about job security are widespread, and senior leaders tend to underestimate just how prevalent that fear is among their people.

Professor Meier presented some striking numbers illustrating this gap in perception. When executives were asked how informed and enthusiastic their employees were about AI strategy, 80% believed their people were well informed and 76% believed they were enthusiastic. When employees were asked directly, both figures landed at around 30%. That is a gap of roughly 50 percentage points, and it is where a great many AI transformations quietly stall before they ever get started.

Pic: Professor Stephan Meier, Columbia Business School

Building a Human-Centered AI Strategy

It is tempting to treat the context gap as a data challenge with a technical solution. But surfacing institutional knowledge, aligning on architecture, and rethinking long-standing processes all require people working collaboratively toward a shared goal, and that only happens when people feel safe enough to do it. Beyond the structural silos, institutions routinely contend with people who gatekeep information because it gives them standing, and that tendency only intensifies when job security feels uncertain. The result is a workforce pulled in competing directions: under pressure to demonstrate AI engagement, wary of appearing too enthusiastic about a technology that might replace them, and quietly using unsanctioned personal tools in the meantime. None of these behaviors move the organization forward.

Ultimately, both the context gap and the perception gap are people problems. Their solutions will involve better data infrastructure, new workflows that improve access, and stronger leadership. But running through all of them is the same underlying challenge: change management.

Getting people genuinely on board is arguably more important right now than investing in the best models. New models come out regularly, each with improved capabilities, and there is a legitimate case for using lighter, more cost-effective models for less complex work. But model selection is a moving target. What remains constant is the need to align teams around AI transformation, and that requires positioning AI not as a threat to existing roles but as something that augments what people do best. The question to lead with is not what can AI automate, but what can AI make more room for.

In wealth management, for example, what humans do best is also what clients value most: building trust, exercising judgment, maintaining relationships over years. When Morgan Stanley began rolling out AI to its financial advisors, a case Professor Meier studied closely, the technology was not positioned as a cost-cutting measure. It was framed as a gift of time, with hours saved on research and administrative work freed up so an advisor could make one more call to a client, have one more substantive conversation. That reframe proved decisive. Adoption started slowly, then spiked, and usage rates among financial advisors now sit above 90%.

From the left: Mangesh Patnaik, Marc Vollenweider, Professor Stephan Meier.

The lesson is transferable across financial services, and the question is how to replicate it deliberately. Drawing on his research, Professor Meier presented a framework for driving AI adoption inside large organizations: five principles targeting the underlying barriers of will and skill that determine whether a transformation takes hold or stalls.

The Five I's: A Framework for AI Change Management

Professor Meier presented five principles that have emerged as the engine of effective AI adoption, each targeting the underlying barriers of will and skill that determine whether a transformation takes hold or stalls.

1. Inform. A common mistake is communicating AI strategy at a level that is too abstract. Statements like "we are becoming AI-first" may be directionally true but they do not answer the question employees actually care about: what does this mean for my work? Effective communication is practical, repeated, and role-specific. A financial services analyst, a data engineer, and a client-facing advisor will not experience AI in the same way, and the message needs to reflect that.

The Morgan Stanley story, described earlier, is a useful illustration of this principle in action. Rather than mandate adoption, leadership held hundreds of meetings with financial advisors, brought in peer voices, and kept the message consistent: this is about freeing you to do more of what only you can do. Adoption rates exceeded 90%.

2. Incentivize. Employees need to see concretely what is in it for them, whether that means reducing repetitive work, improving quality, or creating more space for higher-value client work.

In banking, where incentive structures are deeply embedded in how people work and what they prioritize, it is worth asking early whether adoption can be tied to the metrics people already care about. But the choice of metric matters as much as the decision to use one. As the token maxing phenomenon illustrated, poorly designed incentives can drive costly behavior that looks like progress without delivering any.

3. Involve. People are far less resistant to change they helped create. Employees who sit close to client workflows, sector-specific problems, and data challenges know where AI can help and where it can introduce risk. Involving them through cross-functional teams that bring together domain experts, data architects, technologists, and process specialists transforms employees from subjects of change into active agents of it.

BBVA, a European bank operating across 25 countries with over 25,000 employees, put Involve and Incentivize into practice when faced with widespread shadow AI spreading through its workforce. Rather than restrict it, they channeled it, creating scarcity around 3,000 licenses awarded by enthusiasm rather than seniority, then giving employees the autonomy to build their own tools within a governed framework. Around 5,000 applications were created, and learning spread peer to peer rather than top down. Today, 83% of employees use AI tools weekly, saving an estimated two to five hours each per week.

4. Inspire. Too much of the AI conversation is framed around efficiency, and when leaders talk only about productivity and cost reduction, employees hear a different message: fewer people, higher expectations, less security. History suggests that is not how transformative technology tends to play out. ATMs reduced the number of tellers needed per branch, but cheaper branches meant more branches, and the teller role evolved into something more valuable: relationship banking, sales, client service. The technology changed the job rather than eliminating it, and demand for those evolved roles grew. AI will follow a similar pattern, though the full shape of those new opportunities is still emerging.

What is already becoming clear is that the most significant gains from AI will not come from cost reduction alone. They will come from entirely new capabilities. Marc Vollenweider, Evalueserve's co-founder and Chief Strategy Officer, offered a compelling example drawn from his years in pharma consulting. The probability of a drug reaching market is one of the most consequential data points an investor can have. The earlier and more accurately that intelligence can be captured, the better the models built around it. With LLMs, it is now possible to transcribe and analyze conference proceedings in real time, surfacing the most relevant signals and delivering structured insights directly into the analytical process. Information that once required weeks of synthesis can now be captured from the ground as it happens. For banks, better and more accurate models mean better pricing, more precise risk assessment, and improved underwriting decisions. That translates directly to revenue, and it is the kind of opportunity that should fundamentally change how organizations think about what they are building toward.

5. Instruct. Even willing employees cannot adopt tools they do not know how to use. Training needs to operate at three levels: basic AI fluency, so people can use tools responsibly and effectively; workflow fluency, so they understand how AI fits into specific processes; and judgment fluency, so they can evaluate outputs, challenge assumptions, and apply domain expertise.
Financial services has traditionally devel-

oped that judgment through apprenticeship, with junior analysts building intuition by working through foundational tasks alongside more senior colleagues. As AI automates more of that entry-level work, the risk is not just a skills gap today but a leadership gap further down the line. The goal of Instruct is not compliance but confidence, and for financial institutions specifically, it means being intentional about redesigning the learning pathways that the industry has long taken for granted.

Every organization will weigh these principles differently depending on its culture, its leadership style, and where its specific friction points lie. The five I's are not a sequence to be followed in order but a set of levers, and the leaders who use them most effectively are the ones who understand their own organization well enough to know which ones will move people, and how hard to push each one.

Getting Off the Starting Line

Every institution is in this race, whether they are ready or not. The question is how to move beyond the starting line with purpose and in the right direction. With AI, nobody knows exactly how industries and institutions will look on the other side. What is clear is that the organizations best positioned for whatever comes next are the ones that start moving now, deliberately and with a clear sense of what they are building toward.
That is easier said than done for large financial institutions, which move carefully by design. Coordinating thousands of people across functions, geographies, and legacy systems toward a shared new way of working unfolds over years. In some areas, there are not just perception gaps and context gaps, there are chasms, and the productivity paradox is glaring.

But AI is advancing in weeks, and the gap between institutions actively building toward AI maturity and those still deliberating is already widening. The challenge is not whether to move, but how to move faster without losing control. One of our major banking clients describes Evalueserve as the speedboats to their battleships, partners who can move quickly, test in different directions, carve out well-defined use cases, and start assembling the context layer that makes agentic AI adoption possible, while the larger organization catches up. The right partner brings specialized expertise and implementation experience, but perhaps more valuably, an outside perspective on where the real blockers are, whether those are data silos, misaligned incentives, or a perception gap that has gone unaddressed. That combination of near-term wins and longer-term structural change is what creates the organizational momentum needed to carry a transformation forward.

Speakers:

Professor Stephan Meier
Columbia Business School

Marc Vollenweider
Co-Founder and Chief Strategy Officer, Evalueserve

Mangesh Patnaik
Chief Growth Officer
Evalueserve

• Client Story

Rebuilding Trust in Should-Cost Modeling with Agentic AI.

Overview

A Fortune 500 home appliance manufacturer partnered with Evalueserve to modernize and accelerate its procurement cost modeling and supplier quote evaluation process. The company needed a faster, more scalable way to compare supplier quotations against internal should-cost models while improving confidence in procurement decision-making.

By combining Evalueserve's procurement and manufacturing domain expertise with agentic AI capabilities built on the Hybrid-X platform and Google Vertex AI Agent Engine, the client significantly reduced the time required to analyze supplier responses, identify cost discrepancies, and prepare negotiation strategies.

What previously required weeks of manual review and reconciliation could now be completed in minutes, enabling sourcing teams to move faster during supplier negotiations and new product introduction (NPI) programs while restoring trust in should-cost modeling across the organization.

Challenge

The manufacturer managed substantial direct materials to spend across steel parts, plastic components, electrical assemblies, and sensors. Over time, the company had invested heavily in a commercial cost modeling platform, and thousands of legacy should-cost models had been developed for mechanical parts.

However, the underlying cost of libraries and assumptions within the platform had become outdated. As supplier pricing evolved, the gap between model predictions and actual supplier quotes widened significantly, reducing organizational confidence in the system.

To improve sourcing accuracy, the procurement organization began requesting highly detailed supplier submissions, including:

- Detailed line-item cost breakdowns
- Technical drawings and specifications
- Manufacturing assumptions
- Supporting commercial documentation

The goal was to compare supplier quotations line by line against internal should-cost outputs and identify pricing discrepancies.

In practice, however, the process became highly manual and time-intensive.

Suppliers submitted information in inconsistent formats, often combining spreadsheets, PDFs, engineering drawings, annotations, and unstructured commentary. Procurement teams had to manually navigate large volumes of information to reconcile supplier responses against internal models.

The process created several operational challenges:

- Buyers spent weeks comparing supplier responses and validating assumptions
- Missing information was often identified late, resulting in multiple clarification cycles with suppliers
- Sourcing events and NPI timelines experienced delays
- Negotiations remained heavily dependent on buyer experience rather than standardized cost intelligence
- Trust in the existing should-cost platform continued to decline

The core challenge was the inability to operationalize it efficiently and consistently at scale.

Our Approach and Solution

Evalueserve addressed the challenge by combining deep procurement domain expertise with agentic AI designed specifically for manufacturing and sourcing workflows.

The engagement began with a proof of concept (PoC) focused on three key objectives:

- High-accuracy extraction of unstructured supplier data
- Rapid comparison of supplier quotes against cost models and peer suppliers
- AI-driven recommendations for supplier follow-ups and negotiation preparation

Intelligent Parsing of Supplier Responses

The first challenge involved extracting information from text-based documents, technical drawings, specifications, and embedded graphical data.

Using Evalueserve's proprietary Hybrid-X platform, a multi-agent architecture leveraging Vision Language Models (VLMs) and Large Language Models (LLMs) was deployed to process complex procurement documentation at scale.

The solution:

- Extracted structured data from PDFs, engineering drawings, tables, and annotations
- Converted outputs into standardized JSON templates for consistent comparison
- Applied AI confidence scoring to indicate extraction reliability
- Enabled domain experts and sourcing analysts to validate and refine outputs during the PoC phase

This created a standardized foundation for evaluating supplier responses regardless of how the information had originally been submitted.

Automated Cost Model Correlation

Once supplier responses were structured, the AI agents evaluated them against expected cost-model inputs.

The platform automatically:

- Identified missing or incomplete supplier information
- Generated targeted follow-up questions for suppliers
- Correlated supplier line items against should-cost model outputs
- Flagged discrepancies across materials, labor, tooling, overheads, and manufacturing assumptions

This significantly reduced the manual effort required to reconcile supplier quotes with internal cost models.

AI-Assisted Negotiation Recommendations

Beyond discrepancy detection, the solution also generated negotiation-oriented insights.

The agentic framework analyzed likely supplier assumptions relative to internal cost computations and highlighted:

- Potential drivers behind supplier pricing variances
- Areas where internal models may require recalibration
- Key negotiation levers available to sourcing teams
- Recommended discussion points and clarification questions for suppliers

The result was a shift from static cost comparison toward actionable sourcing intelligence that directly supported procurement negotiations.

Business Impact

By embedding agentic AI into the procurement cost modeling workflow, the manufacturer transformed how sourcing teams evaluated supplier quotations and prepared for negotiations.

Supplier responses that previously required weeks of manual review could now be processed in minutes.

Buyers and sourcing analysts spent less time reconciling documents and more time evaluating commercial strategy and supplier positioning.

Category managers received clearer visibility into supplier assumptions, pricing discrepancies, and potential negotiation levers.

Faster quote evaluation reduced delays in sourcing events and supported quicker decision-making for new product introduction programs.

Restored trust in should-cost modeling

One of the most significant outcomes was rebuilding organizational confidence in the company's should-cost platform. Procurement stakeholders who had relied primarily on instinct and experience were directly involved in validating AI-generated outputs and refining model assumptions. This collaborative validation process helped position the solution as a trusted decision-support capability rather than a black-box analytics tool.

The PoC was successfully executed across 15 supplier responses covering steel and plastic components, delivering overall turnaround-time improvements of 50–80%.

Following the success of the PoC, the solution is now progressing toward production deployment on Google Cloud Platform (GCP), with integrations planned across SAP, Apriori, and Google Vertex AI Agent Engine to create a scalable, connected procurement intelligence ecosystem.

• Client Story

PharmaHTAInsight: Accelerating NICE & G-BA Submissions.

A global pharma client was preparing simultaneous HTA submissions for a rare disease asset across three European markets. Their market access team faced a familiar but acute challenge: evidence packages were being built manually by separate country teams, comparator alignment was inconsistent across dossiers and critical submission risks on OS maturity gaps, ICER threshold breaches, subgroup fragmentation were only surfacing during internal review cycles, weeks before deadline. The cost of a delayed or failed submission in a rare disease indication with limited patient population was significant, both commercially and reputationally.

Evalueserve deployed **PharmaHTAInsight** as the client's evidence and risk intelligence backbone across all three submissions. The platform's Evidence Intake agents aggregated SLR outputs, CSR data, and cross-jurisdiction HTA precedents into a unified repository indexed by PICO framework. A dedicated Risk Assessment agent ran pre-submission stress-testing for each country scoring comparator alignment against NICE's single technology appraisal methodology, flagging OS maturity gaps for G-BA's AMNOG process, and modeling ICER sensitivity curves against each country's reimbursement threshold. Country-adapted dossiers were generated in GxP-ready format with full audit trails, enabling the client's medical writers to focus on narrative quality rather than evidence assembly.

The engagement reduced evidence synthesis time by 38% across all three submissions,

with pre-submission risk identification enabling the team to address two comparator alignment issues that would have triggered formal objections from NICE assessors. Dossier localization that previously required parallel workstreams across three country teams was consolidated into a single agentic workflow, and the client has since extended the platform to cover four additional pipeline assets entering market access planning.

Driving AI Conversations Globally.

Evalueserve is actively shaping the dialogue on how AI is moving from experimentation to real business impact. Across industries and regions, these conversations reflect a shared focus on scaling intelligent solutions, strengthening decision-making, and turning innovation into measurable outcomes.

ISPOR 2026

Philadelphia welcomed Coleen Hobaugh and Dr. Jagriti Prasad at ISPOR 2026, one of the world's leading health economics forums. Coleen and Jagriti had four posters, one private dinner, and countless conversations at the intersection of evidence and impact.

Google Cloud Next 26

Where enterprise AI meets a serious scale. Evalueserve's team hit the floor at Booth 4707 at Google Cloud Next 2026 to show how domain-specific AI and the Mind+Machine™ approach help organizations move from cloud ambition to real business impact, faster than they thought possible.

4th S&S Innovation Summit

Corina Peiu and Abhay Singh joined ministers, investors, and infrastructure leaders in Riyadh for the 4th Stadiums & Sports Innovation Summit. There, they had conversations around AI and public-private partnerships turning ambition into lasting impact.

Reuters Pharma Exec Dinner

Cristina Dracea and Swaraj Sahay hosted a dinner in Barcelona for pharma executives and industry leaders. At the dinner, they unpacked what's working, what's fragmented, and what it takes to turn scattered progress into something that performs at scale.

Adobe Summit 2026

Arjun Vishwanath (pictured) and Coleen Hobaugh attended the Adobe Summit 2026 in Las Vegas. They learned how specialized AI agents can eliminate campaign inefficiencies, surface customer journey friction, and keep tag compliance from becoming an afterthought.

HTAi 2026

Ankit Dhaundiyal attended the HTAi 2026 conference in Istanbul, exploring emerging trends, innovations, and collaborations shaping the future of health technology assessment.

What’s next?

Every organization has a pilot running somewhere. Almost nobody can tell you what shipped. This issue is about the gap between the two — and the surprisingly human reasons it exists. AI is reshaping where value sits inside organizations and the distance between the firms that lead, and those that follow is opening faster than most leadership teams realize.

The real advantage has moved beyond speed. Routine research, synthesis, and drafting are being absorbed by intelligent systems. What remains, what cannot be automated, is judgment. Context. The hard-won understanding of how a particular industry actually operates.

But intelligence alone is not enough. AI does not fix broken processes. Without clarity on how decisions are made, automation simply scales inefficiency. This issue is about what it demands, what it makes possible, and what it looks like when organizations get it right.

Elevate Mag
Issue #6- June 2026

marketing@evalueserve.com
www.evalueserve.com