# The Distance Between Delivery and Outcome

*Enterprise leaders on pilots, tokens, capacity, and the role nobody has defined*

**By Omid Razavi, SuccessLab**

*From the Executive Panel at the SuccessLab Executive Forum: From Deployment to Outcomes. SAP Institute for Product and Engineering, San Ramon, September 3, 2026.*

*The Delivering Products, Realizing Outcomes panel at the SuccessLab Executive Forum, SAP Institute for Product and Engineering, San Ramon.*

Cathal McCarthy opened our panel with a date rather than a technology.

ChatGPT launched in November 2022. It reached the mainstream through 2023. The contracts signed in that first wave are coming up for renewal now, and the scrutiny arrives with them.

That framing set the terms for the hour. The panel was billed as delivering products and realizing outcomes. Cathal turned it into a conversation about who pays, who proves it, and what happens when the invoice lands. He named the underlying problem later:

> *"We're solving a distance problem. A distance between delivery and outcome."*

Let me say early what I think that distance is, because it organizes everything below. It is an ownership problem. Somebody has to own the number that says whether an AI program delivered what it promised, and in most enterprises nobody does. Two of the evening's hardest subjects, the forward deployed engineer and outcome-based pricing, are consequences of that unresolved ownership rather than separate topics. Neither settles while the question of who owns the number stays open.

Three people on stage saw that distance from different ends. Anurag Goel is Global Head of Value Consulting at Red Hat, which makes him the only person in the room who runs a function whose entire job is proving return. Ahmed Quadri leads Agentic Observability Field Engineering at Cisco, having arrived through the Galileo acquisition, after chief customer officer roles at ThoughtSpot, Heap, and Reltio. Rama Kolappan is CPO and VP of AI Platforms at HPE, where he runs product and data engineering for product operations end to end, from customer experience and support platforms through code to cash. He buys, and he builds.

I have organized this around Cathal's questions rather than chronologically, because the questions carried the structure and because agreement and difference only become visible when the three answers sit side by side. At the end I explain why they saw things differently, make three arguments of my own, and suggest what to do Monday.

---

## 1. What actually changed

Rama traced the arc in three words. Generate, then assist, then outcome.

He described a few agents running in production rather than in pilot, and a program about a year old. He did not say how many, or where the human sits in the loop, and I did not push. On the panel, he was the one speaking from operating these systems rather than evaluating them.

> *"It's not about pilot to production anymore. It's all about production and operational."*

From that vantage point he described how much the standard has moved. It used to be enough to summarize a deal. Then it was enough to run a competitive analysis and produce the configuration, the pricing, and the quote. Now:

---

> *"It actually also looks at the whole value chain, looks at is the configuration buildable, and do I have the right order in place."*

A buildable configuration and a valid order live in different systems, owned by different functions, on different data. He did not draw the conclusion, but it seems to me the harder half of that standard is organizational rather than technical.

Ahmed separated what changed from what did not.

> *"The outcome is still the outcome. Everybody's still being measured on what they deliver. It's the pace and the expectation of how that outcome is being delivered that's being challenged."*

As a chief customer officer he pushed the value journey for years, and the questions that actually got asked were how long the professional services engagement runs, whether it is profitable, and whether the customer will buy more. The vocabulary of outcomes has been around a long time. The commercial reality sat elsewhere.

What changed is the clock, and one new element. The customer now knows you have a forward deployed engineer. So the product roadmap stopped being an acceptable answer. If the platform does not do the thing, the expectation is that your engineer will build it, inside the customer's timeline, because they are paying.

The largest financial institutions he works with were fine with an 18-month implementation two years ago.

> *"They want things in 18 days now. The problem is, we can deliver it, but their processes get in the way of that as well."*

He means time to first working outcome rather than a full platform implementation. The expectation of when a customer sees something real has compressed by that order of magnitude, and his team lives in the gap between that expectation and a procurement cycle.

**What this means for you.** The bottleneck has moved, which inverts twenty years of assumption about where implementation risk sits. It came up in every session that evening. Prabodh Ambale of ServiceNow took the same question from the floor about shipping hundreds of changes a week into a field that cannot absorb them. If you lead a customer organization, the constraint you manage this year sits inside your customer and inside your own change process rather than inside engineering.

## 2. Who owns the number

Cathal refused to let outcome stay a word. Outcomes is a lovely term, he said, but let us talk about economics, because that is what shows up in the P&L.

Anurag's answer was a diagnosis. The alignment gap is vertical.

Board expectations on AI return have been set, often in language the working level never sees. The people executing are not aligned to those expectations. The C-suite sits between them. Then the part that makes it structural:

> *"When we are going out and talking to them, the highest we usually get to is the C-suite, but the board carries a different expectation on ROI. So that connection point is missing right now."*

The expectation is often set a level above the highest person a supplier meets, and the people executing sit below where it was set. In many companies nobody in between is measured on reconciling the two. Anurag believes closing that gap is the differentiator, and that it happens before the pilot rather than during it. He put it against the news cycle directly: a new model release is fine, the harness around it is impressive, and without a clear definition of ROI nobody can say whether outcomes are being delivered at all. Harness is this year's term for the scaffolding around a model, the orchestration, tools, memory and evaluation that turn it into something that does work.

He has 100-plus business cases behind that view, and he built similar functions at Salesforce, Adobe, and SAP before Red Hat. He noted, with a nod to our host, that it began at SAP with Ferose.

For buyers, the practical version is uncomfortable. You are the only person who can carry the board's definition down to the project team. That translation is real work, and in most companies nobody has been assigned it.

## 3. No naked pilots

Cathal put the pre-pilot question directly:

> *"You can define an outcome pre-pilot, or you can define an outcome post-pilot to justify the pilot. What happens when you do it upfront?"*

Anurag and Rama both answered it as a methodology question rather than deflecting.

Two years ago, Red Hat measured ROI at production and tried to work backward, which was hard because no baseline existed by then. So they stopped.

> *"We introduced this concept of no naked pilots anymore. Every pilot that we sign up for has to have somebody from my group, from a business value side, and an industry expert attached."*

*Anurag Goel, Red Hat, on defining value before the pilot.*

The second half of the method is what most ROI conversations skip. Attaching a value person and a domain expert does more than create a baseline. It improves which use cases get picked, because those two can also score how hard the change will be. When Red Hat advises a large bank they bring what the twenty largest banks have done, where the return landed, and where the change management proved hardest. Anurag's framing of that second variable is the one I keep using: yes, you can have impact, but what is the effort required to get there? Impact and effort scored together, at selection time, before anyone builds. By his account the pilot-to-production rate is three times what it was, and he credits selection rather than measurement.

He added one more discipline. Quantify in the language of the vertical. Cost to serve for a hospital and cost to serve for a telecom are different numbers with different owners. Generic value models do not survive contact with a specific industry's P&L.

Rama arrived at the same place from the buying side, and pushed the line even earlier. Delivery used to begin at production, which is when you started measuring. Now the experimentation starts far upstream to justify the spend, and before the pilot HPE runs a workshop to pressure-test that the claimed outcome is close to the one they will actually see. In his framing, delivery has shifted back from the moment of deployment to the decision to consume the product at all.

Two organizations, one selling and one buying, moved outcome definition to before the pilot independently of each other. Neither framed it as agreement.

## 4. Cost per outcome

Cathal asked what is actually moving the economics. Token cost, model choice, or which components are reversible at scale.

Anurag rejected all three.

> *"It's not the cost per token, but rather the cost per outcome. That is what CFOs are caring about more and more."*

Worth pausing on what that metric requires. Cost per outcome is a ratio, and most organizations asking for it cannot yet name the denominator. A resolved support case is countable. A buildable configuration is not, at least not in a way procurement would sign against. The CFOs asking this question are ahead of the systems that would answer it, a point I return to at the end.

He described a large European bank running a thousand AI use cases, sponsored throughout by the CIO. Tokens were never the constraint. What changed was who walked into the room. The CFO arrived in the last six months, opened by worrying about token cost because of what he had read, and moved quickly to asking about outcomes. The measurement existed to answer him, because the work had been baselined at the start. The customer has since spoken about the results publicly.

> *"That is the CFO's love language."*

The alternative, in his framing, is a race to the bottom. Argue about token cost and you will lose, and that is the argument you are forced into when you cannot produce a value number.

Ahmed supplied the counterweight. CFOs are only now discovering what they are actually paying for. They bought software that would deliver outcomes, and the bills arriving are for what that software consumes underneath, in inference.

Two reasons this stayed hidden. Inference is metered at the model provider and folded into a vendor's subscription, so there is no line item for a CFO to find. And per-task consumption is genuinely variable, because a reasoning model working an agentic loop may spend ten times the tokens on one request that it spends on the next. Cost per seat was predictable. This is not, which is why finance cannot forecast it even once it becomes visible.

> *"Over the last 18 months, the velocity has kind of eclipsed the value conversation."*

His customers are past pilot and production-oriented, and because they are moving fast, the value story often went unbuilt. He expects that to break within about six months, when organizations can finally see what a unit of agent work costs and start asking what they are getting in return. Today, in his words, it is still a black box.

Rama showed what the cost stack looks like inside a company that also makes hardware. It used to be the LLM cost. Now HPE accounts for infrastructure, models, the harness, and the surrounding ecosystem. He also drew a distinction worth holding onto. **There are two ROI audiences inside one company, and they carry different proof burdens.** Operational teams justify on turnaround time, experience, CSAT, and cost reduction. Product teams justify on time to market and on headcount: I had 300 people building this, can I do it with half, and put the other half on something new. Same word, two audiences, two sets of evidence. Most ROI frameworks I see assume there is one.

HPE built its case about a year ago, when evals were immature and open weight models were less common. Since then the commercial model underneath has shifted from per user to consumption, so the economics they modeled are being rebuilt.

---

## 5. Where the hours went

Cathal asked the question almost nobody asks about efficiency. The promise is capacity creation. How do we make sure the created hours do not disappear?

He offered his own evidence and then undercut it. He has found an extra hour a day through simple automations and multiplied across a company the size of Cisco that is hundreds of thousands of hours. He has also noticed the number of 40-page documents circulating at Kore.ai looking for reaction has gone up.

The three answers did not agree.

Ahmed says the hours never appear as free time. Expectations rose to match the new speed, so the hours refill. His concern is what fills them, because the why question gets lost inside the velocity. Then he named the reason the value conversation stalls:

> *"When you, as a vendor, ask your customer the why, the value question, they kind of fall back to say, no, I'm not going to tell you, because then you're going to charge me more for it."*

Value transparency is blocked by pricing fear. The buyer knows an honest answer weakens their position at renewal, so they decline, the supplier cannot build the case, and both sides arrive at the renewal without a number. No one else on the panel raised it, and I suspect it explains a good deal about why outcome measurement so often stays theoretical.

*Ahmed Quadri, Cisco, on connecting delivery velocity with measurable value.*

He closed the point by extending something Prabodh Ambale had said earlier in the evening about AI slop in code. Ahmed asked the same question one level up: there is no shortage of ways to fill the time being saved, but is any of it valuable, and is it being filled with slop?

Rama answered it the way most leaders will not on a stage.

> *"If I were to say that absolutely 100 percent of the capacity that's being created is being reused somewhere else, that would be a lie."*

Some of it is efficiency, and he said so plainly. Then he described where people actually go. Product support moved to solution support, and solution support into the account roles they call ASMs. In deal management, AI took the lower-level work and managers now cover more deals. In customer success, digital scales the long tail so CSMs concentrate on larger accounts. In engineering, reclaimed capacity goes to other products.

The jobs move up the ladder, and some of the bottom rung goes away. That is a more specific answer than most of us have been giving to the question our teams are asking.

Anurag reported the effect on his own team: 30 to 40 percent more accounts covered, with real top line impact alongside the bottom line.

Be careful about promising your organization that all reclaimed capacity gets redeployed. Rama's version is the one your teams will recognize as true, and saying it early costs less than saying it late.

## 6. The role nobody has defined

Ahmed raised forward deployed engineering, then apologized for going off script. Cathal set his own agenda aside and went there. FDE had already surfaced in the Rocketlane workshop readout, in Ferose's keynote discussion, and in Prabodh's session, where he mentioned he is hiring a principal forward deployed engineer reporting directly to him.

Cathal set it up. Years ago, there was exactly one person you never wanted talking to a customer, and it was an engineer. Now they are out front.

This is the ownership gap in human form. When nobody owns the outcome definition, somebody still has to absorb the difference between what a product does and what a customer expected. That somebody is now an engineer.

A definition first, because the term gets used more than it gets explained. A forward deployed engineer works inside the customer's environment, on the customer's data and problems, building whatever closes the gap between what the product does and what that customer needs. The pattern started in the AI-native companies and spread quickly over the past two years. That is where the agreement ends, which is why this section exists.

*Rama Kolappan, HPE, describing an internal forward deployed model for operational transformation.*

The three definitions that followed diverged on the essentials.

**Ahmed took ownership of getting it wrong, in public.** He built Galileo's FDE team as an extension of product, on a software engineering skill set, with a very technical interview process. He calls that a mistake, and he was precise about why:

> *"It's a mistake not because they're software engineers, but because we overlooked the complexity of understanding problems from an outcome perspective, and understanding what it takes to deliver that outcome in a complex customer environment."*

He tied it back to the keynote, where Ferose told the room that a good engineer falls in love with the problem before the solution. Ahmed's application: if you hire people who only love to solve, they will not speak to value.

Then a warning about imitation. Every peer he talks to wants to build it the way the AI-native pioneers did. That profile was entrepreneurial by design, and nearly all of those people have since left to start their own companies.

> *"I can tell you that's not what I hired. And I don't think that model works for everybody. You have to look at your product first. Does your product even support that FDE model?"*

There was a human cost to the hype. A year ago the FDE was positioned as the answer to everything, and the people who landed in those roles are, in his word, shell shocked.

**Anurag's answer was structural, and it is the direct alternative.** Do not look for that person at all. Thousands of FDE job posts share one thread, which is that everybody is hunting a unicorn. He splits it into three flavors: technical depth, engagement management in the classic consulting sense of running a project profitably, and outcome definition, which his own team supplies alongside the FDEs. Three roles, kept separate, played as a team rather than an individual.

He added a commercial note almost nobody is discussing. Red Hat does not charge for FDEs yet, and nobody is asking whether FDEs are profitable.

**Rama described a version I have not seen named elsewhere.** His organization has no FDE facing customers. It has the same pattern facing inward.

> *"I should probably name it something before it gets named."*

A team goes into the operational functions, CPQ, invoicing, order management, and treats those functions as its customers. They find the problem, decide whether the answer is an internal build or a vendor product, identify the workflow transformation and the process changes, work out where AI fits, and then deliver it with those teams. He says the returns have shown up in turnaround time, experience, and efficiency.

That is forward deployed engineering pointed at your own company. Many organizations are building this right now under names like transformation, internal products, or business technology, without recognizing it as the same role with the same hiring problem.

Then an audience question cut to the commercial core of it. Are your FDEs checking code into core platform repos, or into customer-specific repos? Because otherwise, the questioner argued, they are higher-intellect service engineers. Ahmed's answer: definitely the core product, they are extending core product, and they try hard to avoid customer-specific repos though it happens in certain situations.

---

## 7. Outcome pricing, and the blurred lines

Pricing is the same gap expressed commercially. You cannot price an outcome nobody has been made accountable for defining.

An attendee asked it directly. If I want to charge on outcomes, what does that look like, given the outcome could be delivery, adoption, experience, or hard dollars? And where does FDE end and customer success begin, when each demands a different skill set?

---

Anurag and Ahmed both answered the same way. Not yet.

Anurag pointed to the companies that have moved. Intercom and Fin, Agentforce, Sierra, Decagon. Each has one specific outcome. Buy Decagon and you pay based on calls resolved. Then the caveat most pricing discussions miss:

> *"They have built an entire system, including rev ops, sales incentives. It's not just that you publish a new pricing page out there."*

Ahmed added the precision requirement. There can be no ambiguity in the outcome your revenue model rests on. On the second half he was blunt: the lines between FDE and customer success are blurred and will not stabilize until outcomes start being delivered and renewal pressure forces a realignment. He had noted the structural version earlier, that pre-sales and post-sales are converging into single technical organizations.

Rama put the question back on the business model. For a digital team serving support the outcome is case resolution, deflection, and mean time to resolve, and you price against those and against adoption. His summary: everyone is exploring, any claimed perfect solution is perfect for exactly one product, and the answer depends on whether you run capex, SaaS, or SaaS-led capex.

---

**What all three agreed on**

They converged on the mechanics, and I have flagged those convergences where they occurred. The deepest agreement came out differently from each of them, which is what makes it convincing.

The scarce skill is understanding the problem. Anurag wants people in the room who care about the business problem. Ahmed wants engineers who fall in love with the problem. Rama starts inside the function, understanding what is broken, before deciding what to build. Three roles, three companies, one conclusion, and the same one Ferose had put to the room in his keynote earlier that evening.

Hold onto it, because everything they saw differently sits downstream of it.

---

## Their advice, and what keeps them up

Cathal's closing question was what one thing would shorten the distance between delivery and outcome.

---

**Anurag:** "Technology, right from the days of SAP to AI, is about solving business problems. So define the business problem and get people in the room who care about that, who will be able to match your definition of ROI."

**Ahmed:** "The outcome is in the eye of the beholder. Whoever gives you an outcome, break it down into smaller pieces. The outcome might be a bigger outcome than you can deliver."

**Rama:** "Be flexible. The initiative we built, we started a year ago and finished recently. We would probably be doing it very differently now. It previously used to be prompt engineering, then context engineering, now it's harness engineering. That changes the dynamics of what you build. Break it small, show some wins, get the outcome successful. And get the blessing of your CFO, because your next round of funding has to be approved."

Atul Nanda then asked from the floor what keeps them up at night a year from now, with outcomes as a key measure.

Ahmed said the organization around his team, and whether it can adapt fast enough. He closed with a callback to Ferose's keynote about the best salespeople earning a week in Hawaii while the best engineers earn more work. He is in the sales organization now, so he qualifies. What stands between him and Hawaii is how quickly everyone around him learns this.

Rama said adoption. You can have the right outcome, break it down, write the KPIs, execute the build. Then it has to be adopted, and only then does the return appear.

> *"And you don't have a lot of control over that, because that's something that you're building for someone."*

Anurag turned it inward. The change lands on the people delivering, under multi-year pressure to become AI fluent while still hitting outcomes. In his words, it is a human business at the end of the day.

Three leaders asked what they fear, all answered with people.

### Three views, three settings

Where they saw things differently, the differences seemed to me less about opinion than about vantage point.

Anurag works across a broad account base, where three specialists can be staffed centrally and spread across engagements, so splitting the FDE role is affordable. Ahmed's engineers extend a core product, so the person has to be inside the codebase, which collapses those three roles into one and makes the hiring bar brutal. Rama's team has no external customers, so his version points inward and looks like internal consulting with engineering attached.

The same pattern held on capacity and on where proof of ROI belongs. Each answer was true from that seat. Which suggests the industry may be hunting for a single right answer to questions that have a right answer per setting, and the useful work is knowing which setting you are in. That is also why copying a definition from a company whose economics do not resemble yours goes wrong so reliably.

## Three arguments I would make

The panel gave me three arguments I did not walk in with. The first is about the ownership problem itself. The second and third are what happens while it stays unresolved. This is where reasonable people should push back.

1) **The value conversation is blocked by incentives, not by instrumentation.**

A buyer who declines to disclose value is behaving rationally, and no amount of better observability changes that calculation. The industry is investing heavily in measurement tooling on the assumption that the constraint is technical. I think it is commercial. Until a buyer can disclose realized value without it being used against them at renewal, we will keep producing numbers neither side will validate. I suspect the fix is contractual before it is technical: pre-agree that the pilot's value baseline is not an input to pricing for the current term. Someone has to go first, and it should be the vendor, who has more to gain from a measured account than from one point of margin. I would like to see an enterprise vendor try this publicly in the next year.

2) **FDE is the industry hiring around a product gap.**

When a customer says your engineer will build it because your product cannot, that is a product statement. Every FDE hire is a bet that the gap is temporary and the work flows back into the core product, which is why the repo question from the floor was the sharpest commercial question of the night. If the work lives in customer-specific code, you have a services business carrying software margins, and few are auditing whether it clears. So the question for any leader standing up the function is not how to hire an FDE. It is how long you expect to need one, what has to become true in the product for the answer to be "not much longer," and who owns making that true. Without an answer to the second question, you are not building a bridge. You are funding a permanent department.

3) **Outcome-based pricing will not spread beyond narrowly countable service outcomes this cycle.**

The companies that have moved all price a single countable outcome, and it is resolution. A case either closed without a human or it did not. That property is rare. A buildable configuration, a faster deal cycle, a CSM covering more accounts: none produce a unit procurement will sign against, and rebuilding rev ops and incentives makes the switching cost multi-year even where the unit exists. So this renewal wave will not be fought over outcome pricing. It will be fought over consumption spend, with cost per outcome as the justification rather than the model. Stop preparing for outcome-based pricing. Start preparing to defend a consumption line to a CFO who has never seen one.

One caveat on all three. Panels are a biased sample by construction. Companies where none of this is working do not put their leaders forward, so the skeptical case, that many enterprise AI programs still show no measurable P&L impact, had no one in the room to make it. You are reading the experience of teams who got this working, mine included. Calibrate accordingly.

## What to do Monday

**Baseline one number this month, even if your pilot is six months out.** Pick a named workflow and measure end-to-end turnaround time before AI touches it. Turnaround time survives every ROI conversation, because the CFO and the operator both recognize it. Add cost to serve for one segment, in your industry's language.

**Put two names on every pilot.** One accountable for value definition, one who knows the function deeply. Not a committee. Anurag's method, reduced to what a mid-size company can afford.

**Score effort alongside impact when you pick use cases.** Ask where the change management will be hardest before you ask where the return will be largest.

**Ask your vendor four questions at renewal.** What baseline did you record before the pilot, and who recorded it? Show me the consumption line separated from the license line, trended monthly. Does your forward deployed work land in your core product or in code specific to us? And will you agree to the value principle below?

**Negotiate the principle before you negotiate the paper.** The principle is simple: measurement is a joint exercise, and what it produces is not ammunition for the next negotiation. Raise it as a working agreement first, because a supplier who will not accept it in conversation will not accept it in a contract, and that tells you what the measurement exercise is for. If you want language to hand counsel, something in this spirit: *value baselines and realized-value measurements established jointly during the evaluation and initial term are for joint reporting and case development, and are not an input to pricing, list adjustment, or uplift at renewal of the current term.* Your counsel will shape it to your paper.

**Decide where reclaimed capacity goes before the pilot starts.** If part of it is efficiency, say so to your team now. Saying it late costs more.

**If you are hiring an FDE, settle three things first.** Whether the work lands in the core product. Whether you are hiring one person or building three capabilities. And what has to be true for the role to stop being necessary. Then ask your product organization to own that third answer.

**If you run internal operations, look for Rama's unnamed function.** A team that goes into CPQ or billing or order management, treats them as customers, and rebuilds the workflow. If it exists in your company, it deserves the same design thinking as a customer-facing FDE team, and a name.

# The number nobody owns

Many major AI renewal conversations in the next few quarters will ask a version of the same question. What did we get for this?

In most companies, nobody owns the answer. The board set an expectation the project team never saw. The vendor measured what it could reach. The CFO arrived late and is working backward from an invoice. And the function that sits closest to the customer, the one that knows what actually changed in the work, is usually not in the room when the number gets decided.

That gap is the opening.

Customer leaders have spent years asking for real influence over product, pricing, and investment decisions. This is the argument that earns it, and it comes with a deadline. Produce the number that survives the renewal and you will not have to ask again.

Baseline something this month.


*SuccessLab Executive Forum participants at the SAP Institute for Product and Engineering, San Ramon.*

*My thanks to Ferose V R for hosting us at the SAP Institute, to Prabodh Ambale of ServiceNow, to Srikrishnan Ganesan and the Rocketlane team for the Agentic Delivery Workshop, and to Cathal McCarthy, Ahmed Quadri, Anurag Goel, and Rama Kolappan for a specific and candid conversation. Each*

*spoke in a personal capacity, from their own experience, rather than on behalf of their employer. Quotes are drawn from the event transcript and lightly edited for readability.*