Would You Recognize AI ROI If You Saw It?

How engineering leaders measure AI's value — and what their metrics miss — **as the bottlenecks keep changing.**



Uplevel

Table of Contents

Leaders are confident AI is working 3

Methodology 4

How we define maturity levels 5

Nearly every leader says AI is creating value 6

Leaders are sure AI is the cause 7

“Sure” is far more common than “Checked” 8

Defining AI’s value through productivity 9

Leaders define AI’s value as efficiency 10

AI’s value extends beyond code generation 11

Faster development exposes new bottlenecks 13

AI moves work into review 14

No single obstacle dominates 16

Reported value dips in the middle of the maturity curve 18

Most leaders diagnose bottlenecks without a tracked metric 20

Proving business value is another matter 21

Leaders extended pre-AI ROI approaches 22

Separating AI impact from other changes 24

VPs are AI's most confident advocates 26

Three findings to take to your next board meeting 28

Would you recognize AI ROI if you saw it? 29

The AI Reality Check 30

What the Data Reveals 31

Leaders are confident AI is working

96% of engineering leaders are confident AI is responsible for the value they're seeing. But how do they know?

A few years ago, engineering leaders were still trying to establish whether AI would deliver meaningful value. Today, that debate has largely been settled. Organizations have widely deployed AI tools, engineers are using those tools every day, and leaders are seeing results they attribute to AI.

But recognizing improvement and knowing what's responsible for it are two different things. Engineering performance is influenced by everything from changes in team composition to the way work moves through the delivery process. AI is being introduced into systems that are already changing, making it difficult to isolate its contribution.

That distinction matters because leaders are making decisions based on what they believe AI is accomplishing. If they attribute an improvement to faster coding when the benefit actually came from a process change, they may invest in the wrong thing next. And if AI makes coding faster while moving the bottleneck into review, the metrics they use may capture the speedup without accounting for what happens downstream.

Our 2026 research examines how engineering leaders are making those judgments and whether their measurement practices have kept pace with their confidence.

Methodology

Uplevel surveyed 428 engineering and technology leaders between August 31 and September 3, 2026, to understand how organizations evaluate AI's impact on software engineering performance.

Respondents represented organizations of varying sizes and AI maturity levels, with seniority ranging from engineering managers to C-level executives. The survey examined how leaders define and measure AI value, attribute performance improvements, and identify obstacles to getting greater value from AI.

All respondents completed the survey and passed an attention-check question. Results are self-reported and reflect leaders' perceptions and measurement practices rather than independently verified engineering outcomes.

Percentages may not total 100% due to rounding. Findings for smaller respondent groups, particularly C-level executives (n=19), should be interpreted as directional.



How we define maturity levels

We asked organizations to self-report their maturity on a five-point scale. Here's how we define each level:

LEVEL 1

Organizations at this level have rolled out company-wide AI tools, and engineers consistently use them for tasks, including code completion, test case generation, and code review.

LEVEL 2

Teams run agents on bounded tasks, often using lightly modified, tool-native agents. Agents in use tend to be fairly generic, with limited org-specific context.

LEVEL 3

Teams configure agents with organizational context, including codebases, conventions, and runbooks. Engineers rarely prompt AI to generate code and tests; their job is to review, correct, and direct agent output.

LEVEL 4

Teams build and maintain agents at the org level and share them across teams. Engineers pull from a common capability layer, with ownership, versioning, and intentional design behind agents and skills.

LEVEL 5

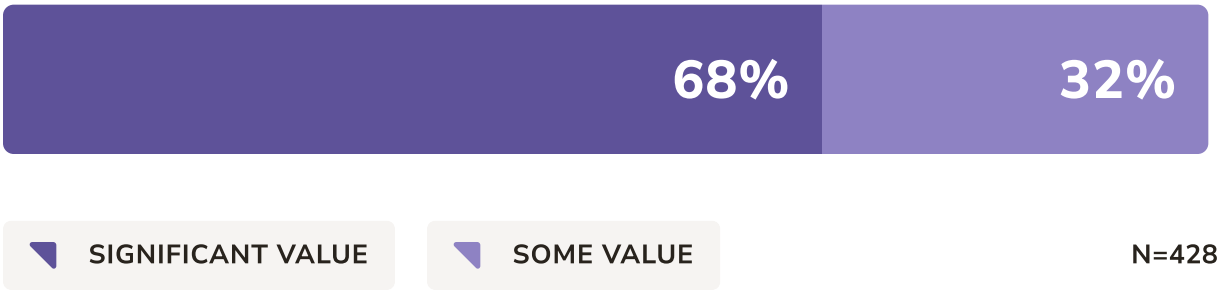
Engineers orchestrate agent teams that deploy independently with minimal human-in-the-loop handoffs. The traditional gates and checkpoints of the software development life cycle have been restructured around agent workflows.

Nearly every leader says AI is creating value

Nearly all 428 respondents reported that AI is creating value in their organizations. More than two-thirds (68%) described that value as "significant," while 32% reported "some value." Only one respondent also indicated that AI had contributed negative value.

The value of AI is no longer speculative. There's remarkably little disagreement about whether AI is helping. For most leaders, the more immediate concern is understanding how much it's helping and what that improvement means for the organization.

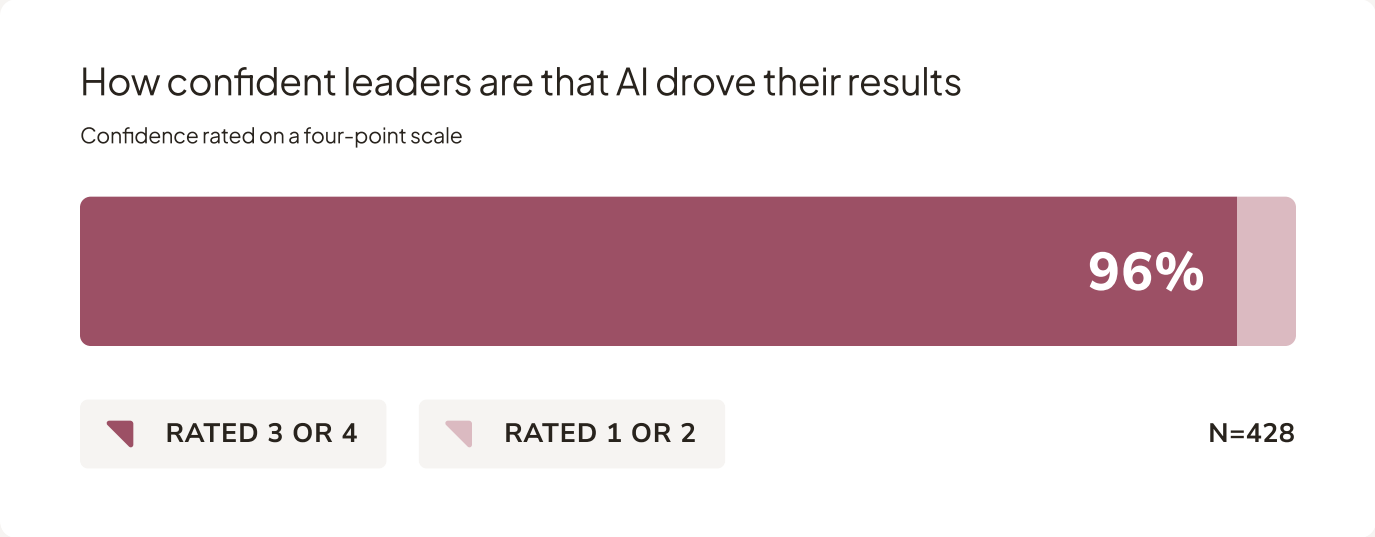
How much value has AI created in your organization so far?



Leaders are sure AI is the cause

Value is notoriously hard to identify and track. There are many different metrics and many ways to approach them. Similarly, there are many different ways to adopt AI and measure its effectiveness.

Despite this diversity, our research again finds consensus. We asked engineering leaders how confident they were that AI specifically was driving the results they observed, rather than other changes such as team growth or process improvements.



That confidence extends across seniority levels, although VPs are particularly certain: 76% described themselves as very sure AI was responsible. Among C-level respondents, that figure was 37%, although the smaller sample (n=19) makes this comparison directional.

This represents a striking contrast with our 2025 research, when 36.8% of leaders identified “difficulty isolating team performance from other factors” as their biggest measurement barrier.

A lot can change in twelve months. AI adoption has expanded, engineering organizations have gained experience, and leaders have had more opportunities to observe its effects. But has their ability to measure those effects improved at the same pace?

“Sure” is far more common than “Checked”

There are several ways to estimate AI’s contribution to engineering performance, each with its own limitations. Controlled comparisons between groups using AI and those not using AI can provide stronger evidence of causation, although finding a meaningful control group becomes difficult as AI adoption spreads.

Only 40% of leaders reported using a controlled comparison between teams with and without AI. The remaining 60% relied on other approaches or hadn’t attempted to isolate AI’s contribution.

So while 96% are ***sure*** AI is driving value, fewer than half took the first step to check and measure.

The gap between confidence and measurement remains large, and confidence appears to be outpacing measurement. Why? Our 2026 research digs into the reasons and explores the potential consequences and implications. At the end, you can find out how you stack up.



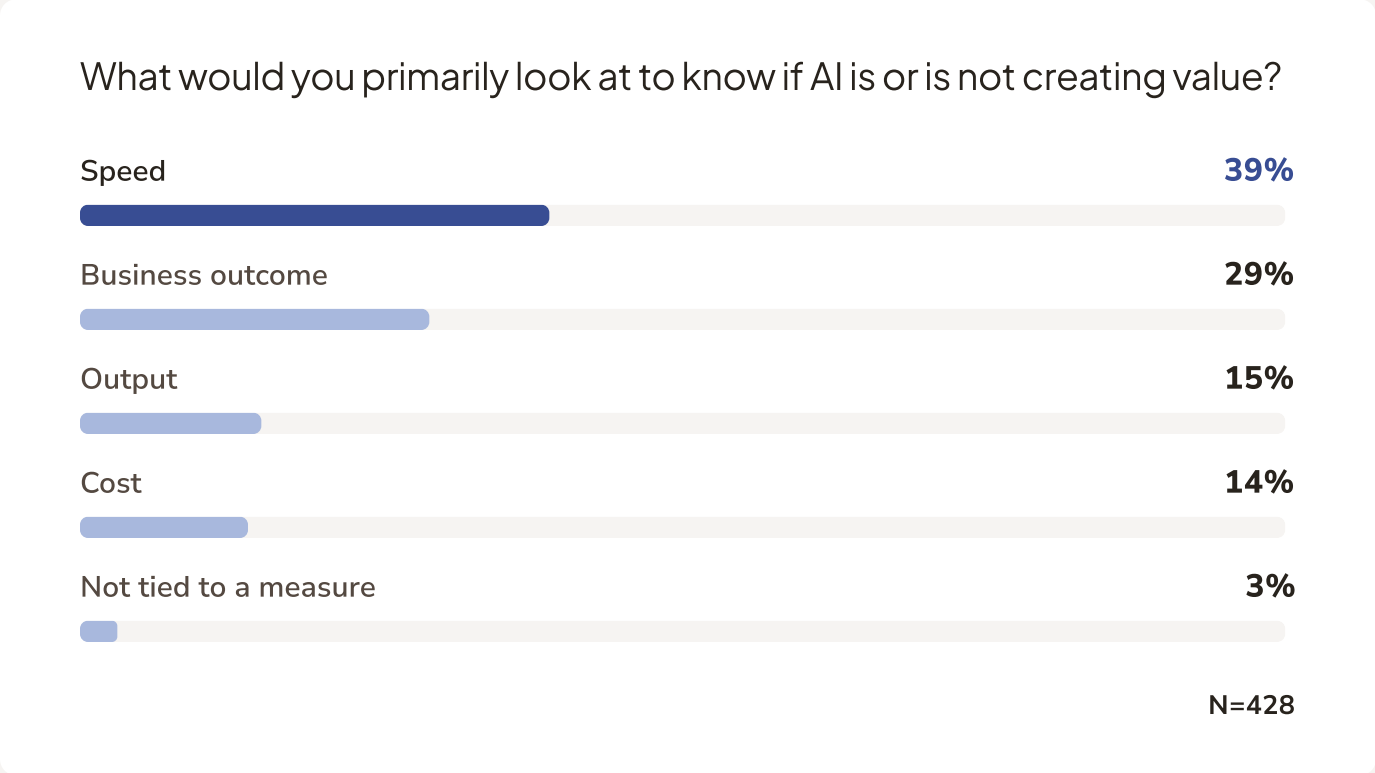
Defining AI's value **through productivity**

39% of leaders look at velocity as the primary measure of success

Leaders define AI's value as efficiency

When ChatGPT launched in 2022 and Cursor gained prominence in 2024, code generation was among the most obvious use cases. Engineers could generate boilerplate in seconds and increasingly delegate coding tasks to agents. The immediate appeal was speed.

AI has since expanded into many other parts of software development, but speed remains the primary measure of success. Our research found that 39% of engineering leaders look first at how quickly work moves to determine whether AI is delivering value.



Business outcomes come next at 29%, followed by output (15%) and cost (14%).

Taken together, speed and cost account for 53% of responses, suggesting efficiency remains a dominant consideration. This echoes our 2025 research, when 53% of leaders identified operational efficiency as their most desired business outcome.

The preference for speed holds across most seniority levels. It's the top measure among managers (41%), directors (40%), and VPs (40%). C-level leaders differ: 37% prioritize business outcomes, 32% output, and just 11% speed.

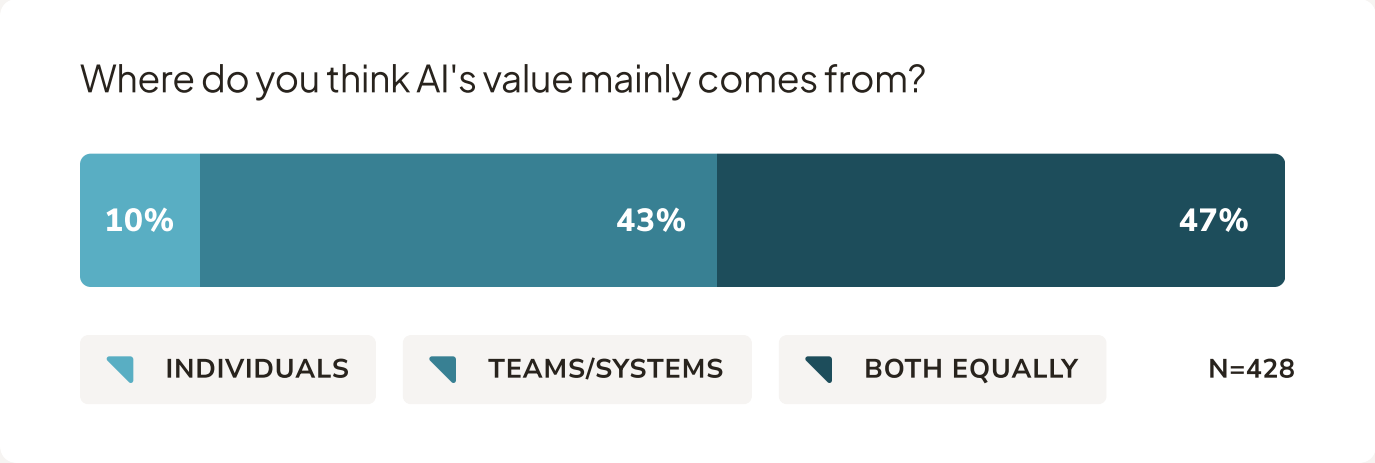
> “The idea that AI creates value by making coding faster remains deeply embedded in how engineering leaders evaluate its impact. We were told that time is money. But now that AI makes certain things faster, we see where value lives and where value delivery gets stuck.”
>
> 
>
> **Amy Carrillo Cotten, Director of Client Transformation at Uplevel**

Where leaders sit in the organization also influences what they measure. Speed is a natural concern for managers overseeing day-to-day engineering work, while executives are more likely to evaluate AI through broader organizational results.

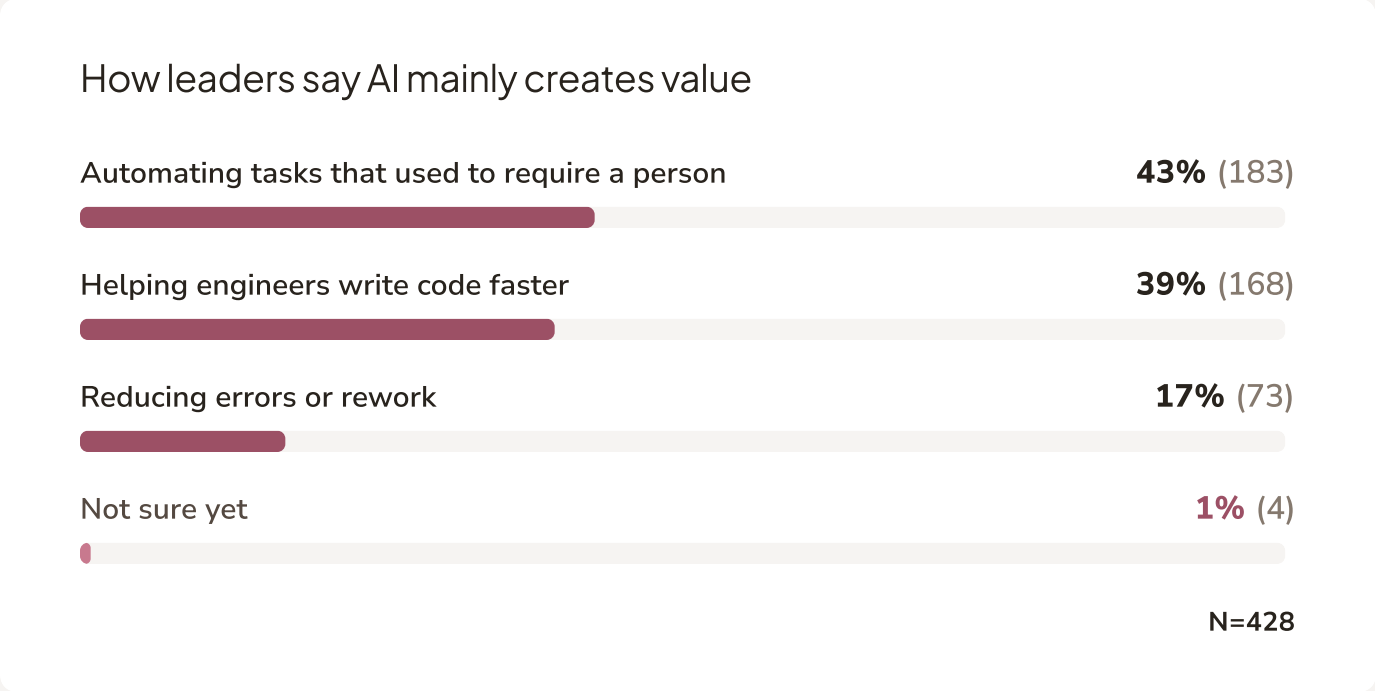
This stratification creates a potential disconnect. Only 29% of respondents primarily measure AI through a business outcome. Everyone else starts with a measure that must eventually be connected to business value. Different measures can lead leaders to different conclusions about the same AI investment.

AI’s value extends beyond code generation

When asked where AI’s value mainly comes from, 43% of engineering leaders credit teams and systems, while another 47% credit them equally with individual engineers. Only 10% attribute AI’s value primarily to individuals doing more on their own.



But when we asked ***how*** AI creates that value, 39% pointed to helping engineers write code faster. That complicates how few leaders attribute AI's value primarily to individual productivity. Leaders recognize the importance of the broader engineering system, yet faster coding (an individual task) remains a major part of how they understand AI's contribution.



Another 43% cited automating tasks previously performed by people, while 17% pointed to reducing errors or rework. These improvements can benefit the entire delivery system, depending on where they occur and how they affect downstream work.

An engineer producing code faster may improve their own productivity without increasing the value delivered by the organization. Understanding AI's contribution requires looking at what happens to that work after it's generated.

As the next section shows, improvements in one part of the delivery process can change where work gets stuck elsewhere.



Faster development **exposes new bottlenecks**

29% name increased time reviewing AI-generated work as their biggest new friction

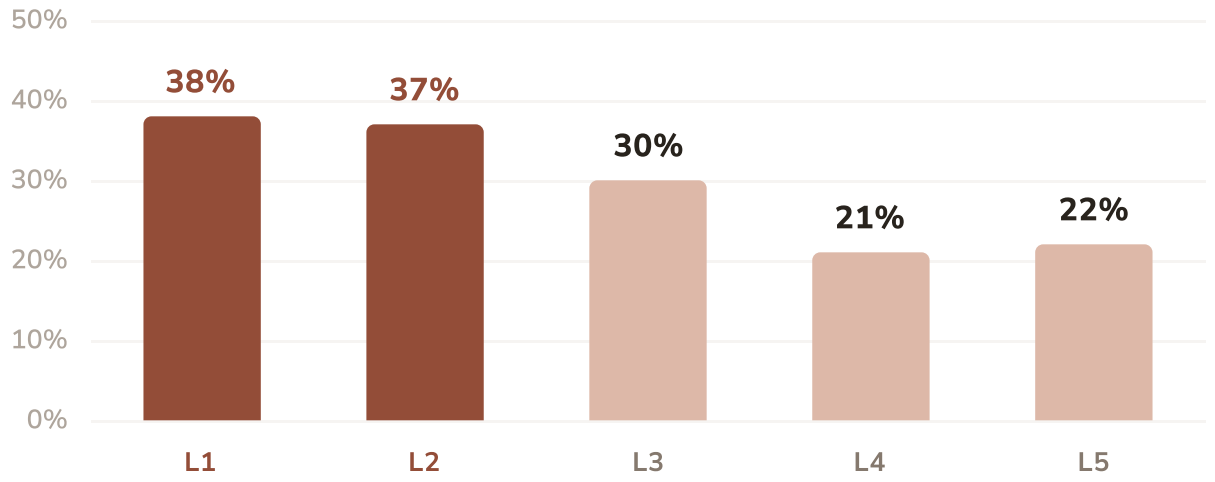
AI moves work into review

AI has made generating code faster, placing new demands on the rest of the delivery lifecycle.

Our research found that 29% of engineering leaders identify increased time reviewing AI-generated work as the most significant new friction introduced by AI in the last quarter.

Review burden varies considerably by AI maturity level, as defined on page 3.

In the last quarter, what's been the most significant new friction or trade-off AI has introduced for your team?



REVIEW BURDEN IS HIGHEST AT LOWER MATURITY LEVELS

N=428

The drop-off is nearly linear: as organizations mature, the review burden falls. More mature organizations either adapted their review practices or shifted the friction elsewhere. VPs are especially likely to identify the problem: 43% name review burden as their biggest new friction, the highest share of any seniority group.

> “Mature orgs typically have cofactors like strong continuous integration, which reduces the surface area that manual review has to cover. They have invested in platforms that ensure quality, in most cases.”



Amy Carrillo Cotten, Director of Client Transformation at Uplevel

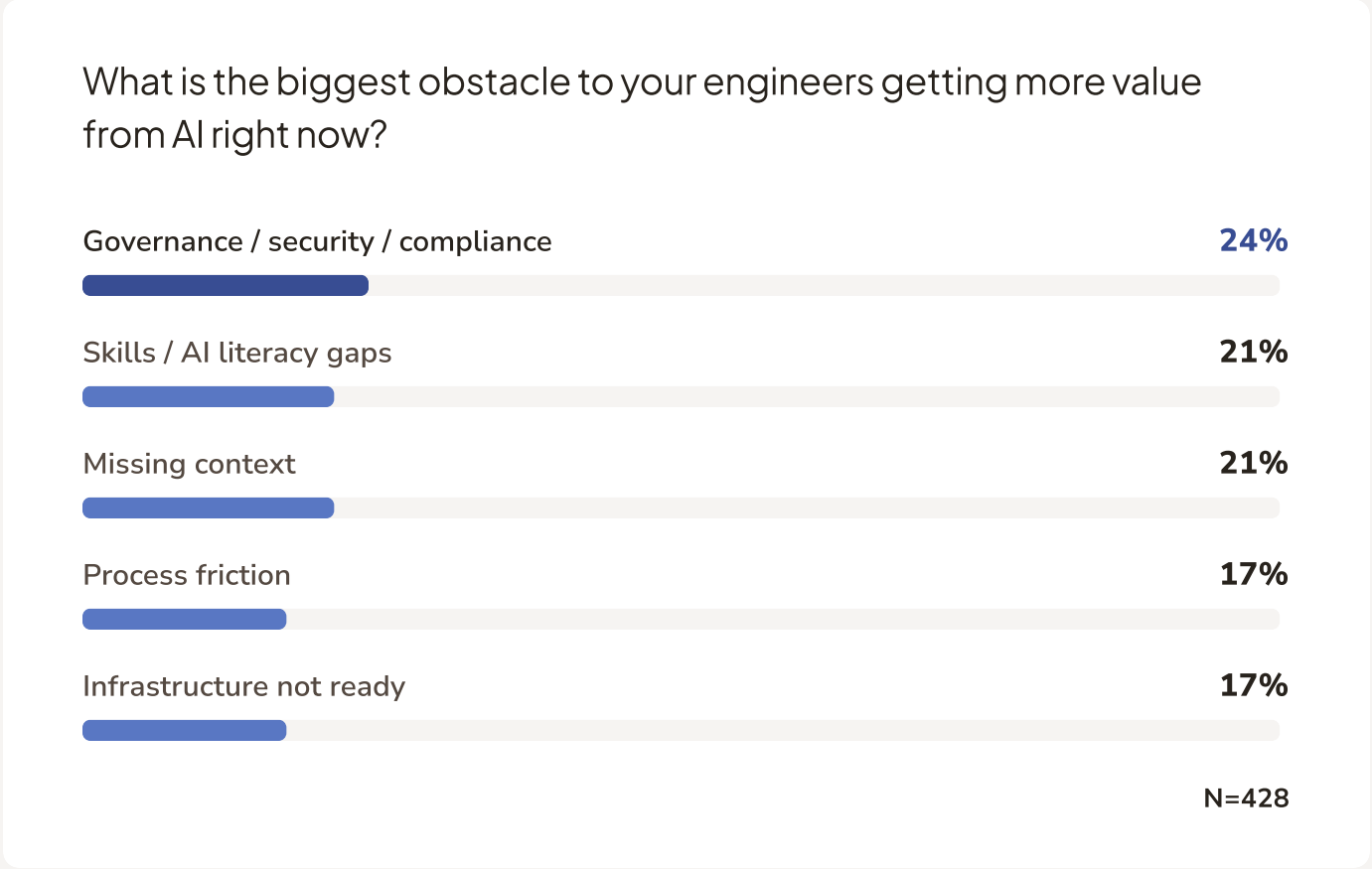
The finding also echoes concerns from our 2025 research, when 65% of respondents ranked QA and validation of AI outputs among their top two skill priorities, and 67% expected QA skills to become more important.

A year later, review is a prominent source of friction. The question is whether organizations have built the capabilities to handle increased output without creating new delays.

No single obstacle dominates

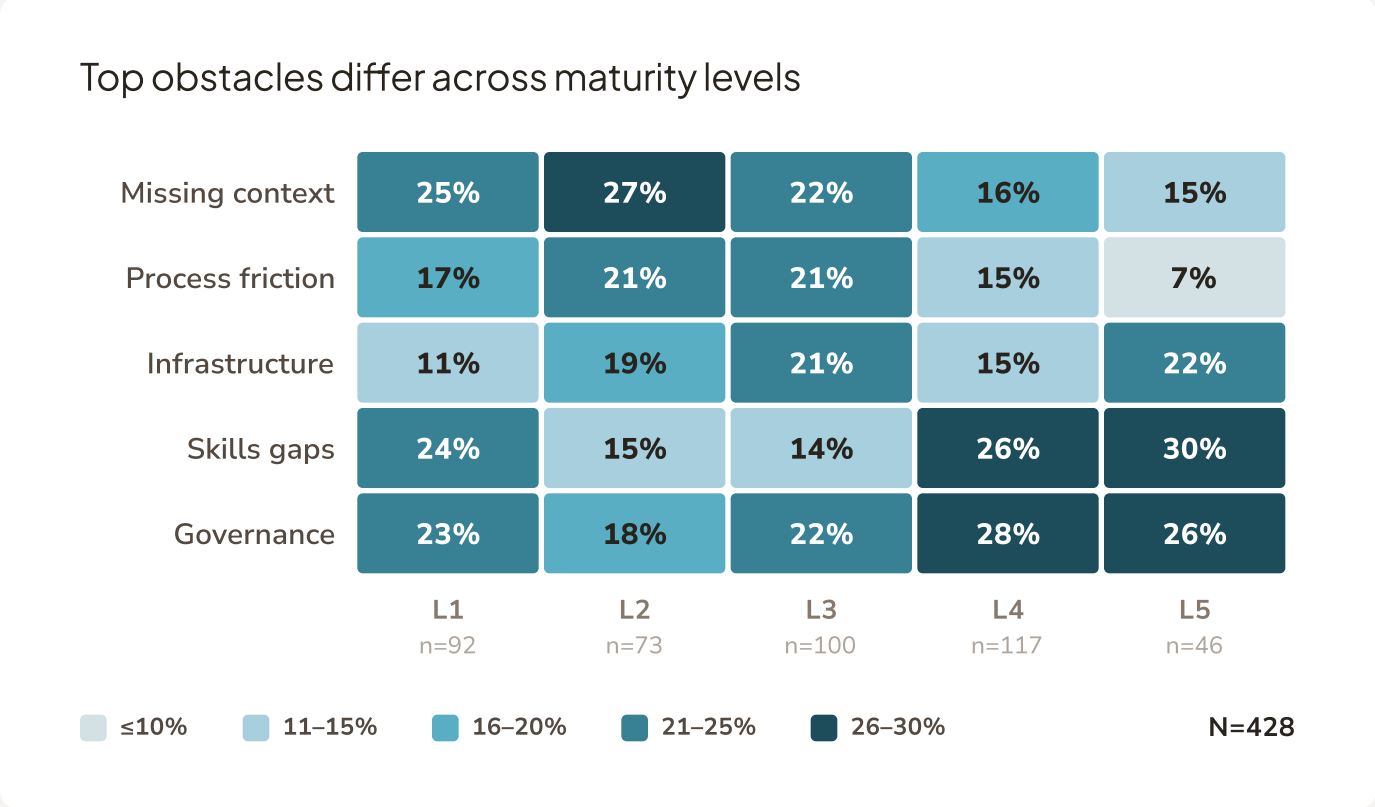
Although many organizations see the review burden as their new biggest bottleneck, surprisingly little consensus exists on what's preventing their organizations from getting more value from AI. Constraints on AI value shift as organizations mature, and there is no single bottleneck as this is a transformation that is more complex than previous transformations.

When asked to identify their biggest obstacle, responses were distributed across five categories:



Governance leads at 24%, but only seven percentage points separate the most and least frequently selected obstacles.

The distribution becomes more interesting when we break out these answers by maturity level. Missing context leads at Levels 1 and 2. Level 3 has a near-four-way tie between context, governance, infrastructure, and process friction. At Levels 4 and 5, skills gaps and governance become the most frequently cited obstacles.



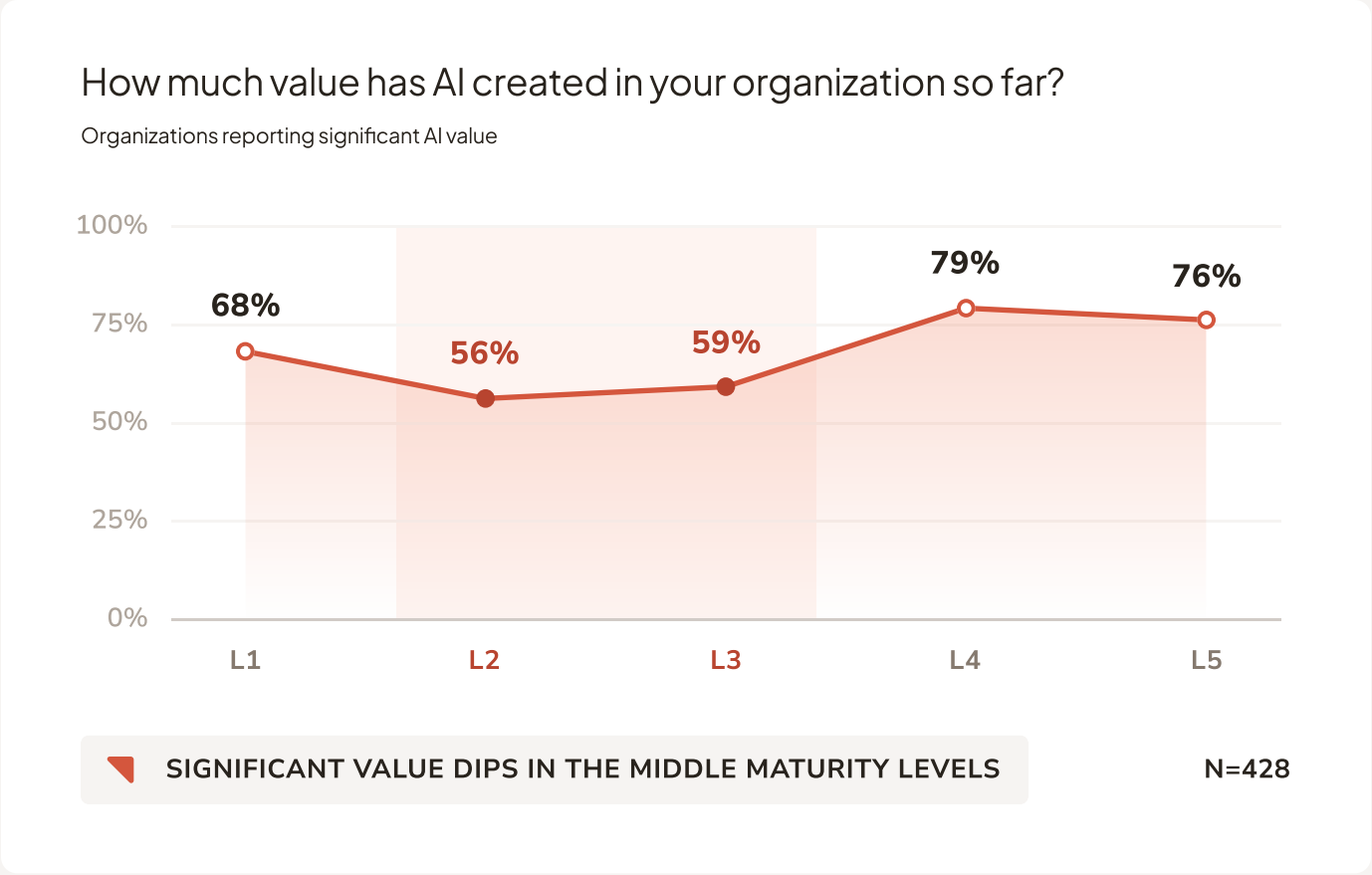
The same disagreement appears across seniority levels. Missing context reaches 38% among VPs, while managers are most likely to identify skills gaps (25%).

> "Orgs that worked on certain fronts earlier may still have blind spots. They may have done a decent DevOps transformation but didn't really do observability, so they are now struggling with context capture and AI measurement. AI maturity requires work on many fronts at once."
>
> 
>
> **Amy Carrillo Cotten, Director of Client Transformation at Uplevel**

Reported value dips in the middle of the maturity curve

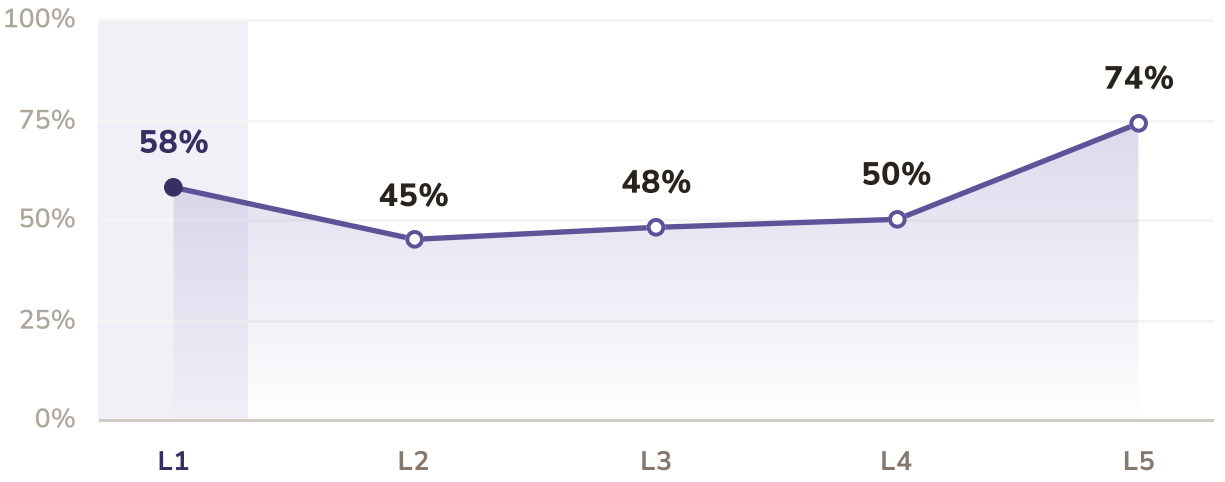
Greater AI maturity doesn't correspond to a steady increase in reported value.

The share of leaders reporting significant AI value is relatively high at Level 1 (68%), then falls at Levels 2 and 3 (to 56% and 59%, respectively) and then rises again to 79% in Level 4 organizations and 76% in Level 5 organizations.



The same pattern generally holds for confidence that AI drives new value generation — but not always. **At Level 4, 79% report significant value, yet only 50% are very sure AI is responsible.** That's the widest separation across the five maturity levels. By Level 5, the two figures nearly converge: 76% report significant value and 74% are very sure AI caused it.

If you look at a metric as evidence of AI's value, how sure are you that AI specifically is what's driving it?



HIGH CONFIDENCE APPEARS EVEN AT LEVEL 1

N=428

One possible explanation is that early AI adoption delivers accessible productivity gains, while more advanced use cases expose constraints that require additional investment. Even when those investments begin to pay off, attributing the resulting improvements specifically to AI may remain difficult.

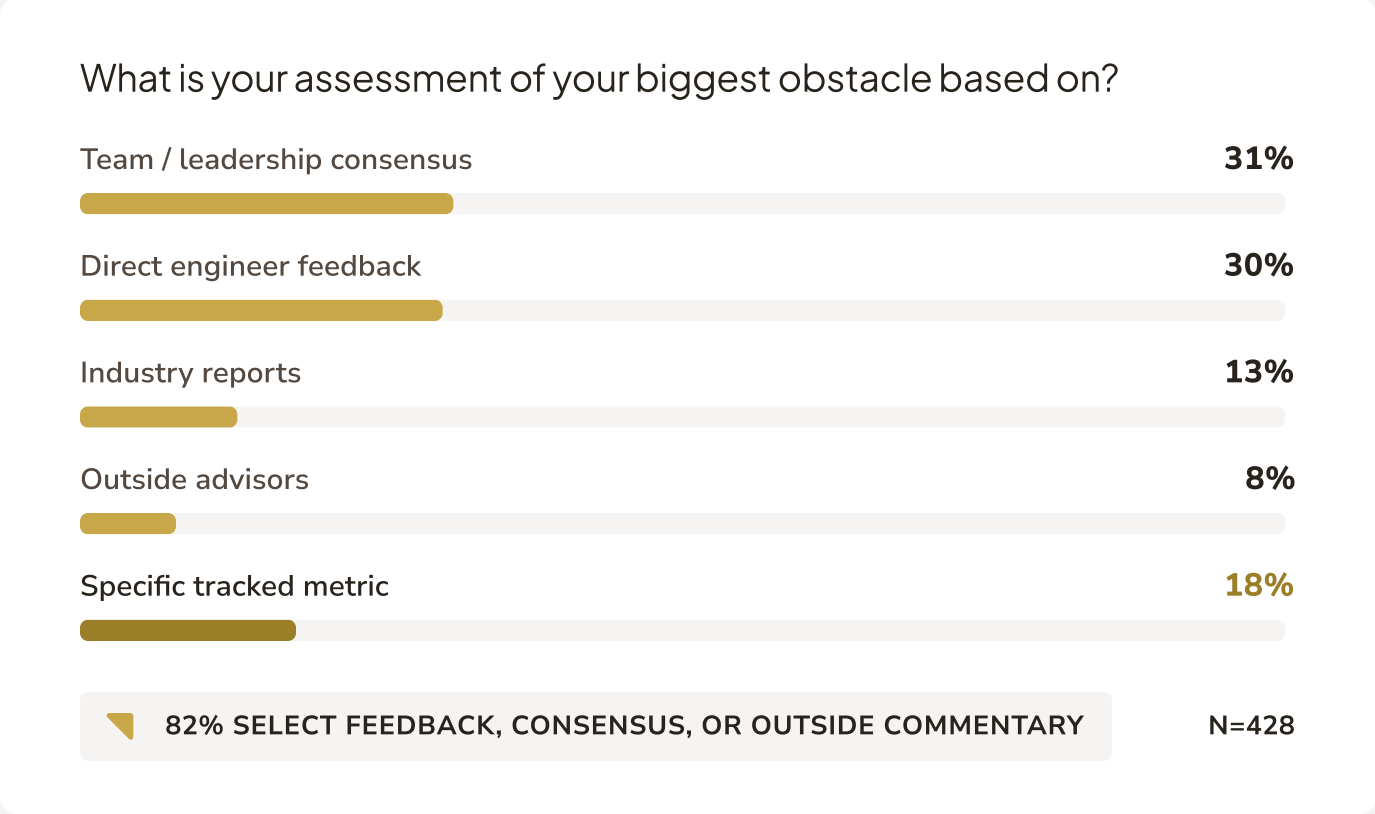
> "It's easy to get some quick wins at Level 1, then the going gets tough. It gets better at the other end, because AI is usually resting on top of previous successful transformations like DevOps, observability, and cloud."
>
> 
>
> **Amy Carrillo Cotten, Director of Client Transformation at Uplevel**

The maturity findings suggest that realizing value and confidently attributing it to AI may develop at different rates. Greater maturity alone doesn't guarantee either.

Most leaders diagnose bottlenecks without a tracked metric

The previous findings point to a recurring challenge. Organizations face different constraints depending on their AI maturity, and the value they report doesn't increase steadily as they advance. So how are leaders deciding which obstacle is holding them back?

Our research found that only 18% base their diagnosis on a specific tracked metric. The remaining 82% primarily rely on team or leadership consensus, direct engineer feedback, industry reports, or outside advisors.



Seniority also influences the diagnosis: 44% of VP/C-level leaders rely on team and leadership consensus, compared with 29% of managers and directors.

The findings raise a concern for organizations trying to improve AI performance. If leaders can't reliably establish what's limiting progress, how will they know when it's time to address something else? Vibe coding is getting all the attention, but vibe measurement might be even more widespread.



Proving business value is another matter

74% extended a pre-AI ROI approach to AI, while only **4%** are building an AI-specific approach.

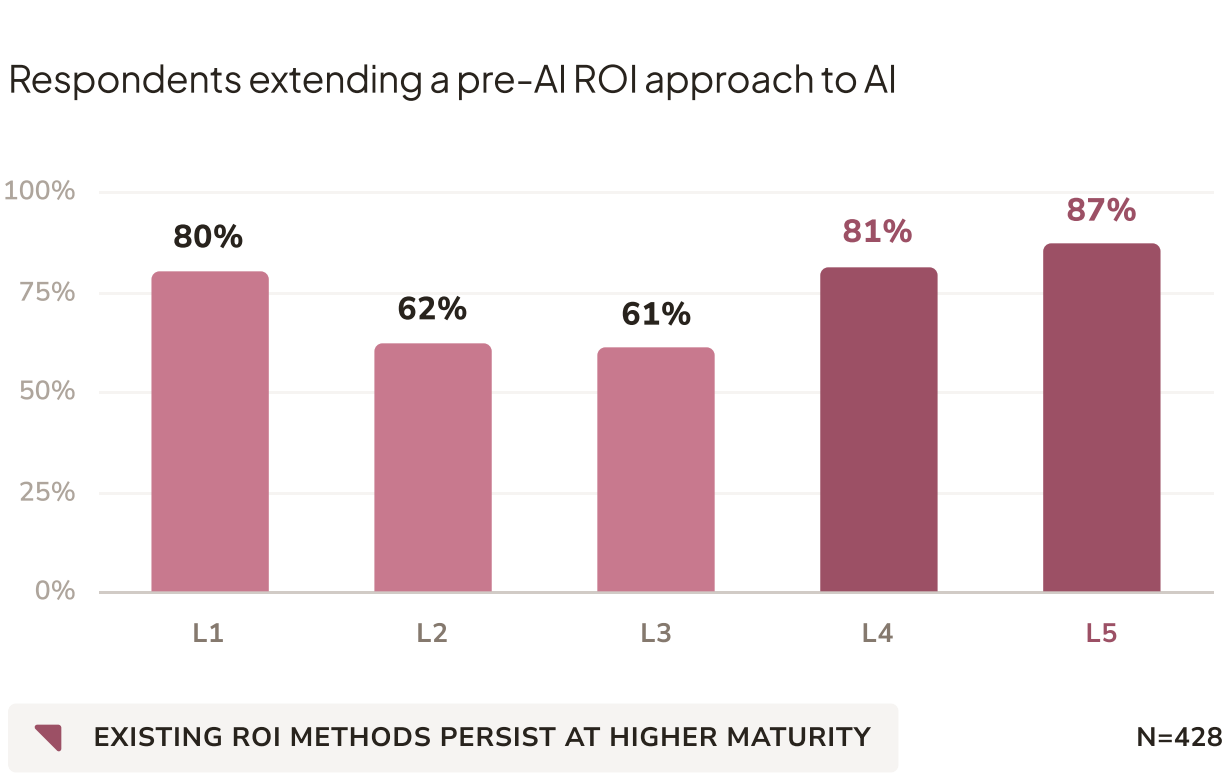
Leaders extended pre-AI ROI approaches

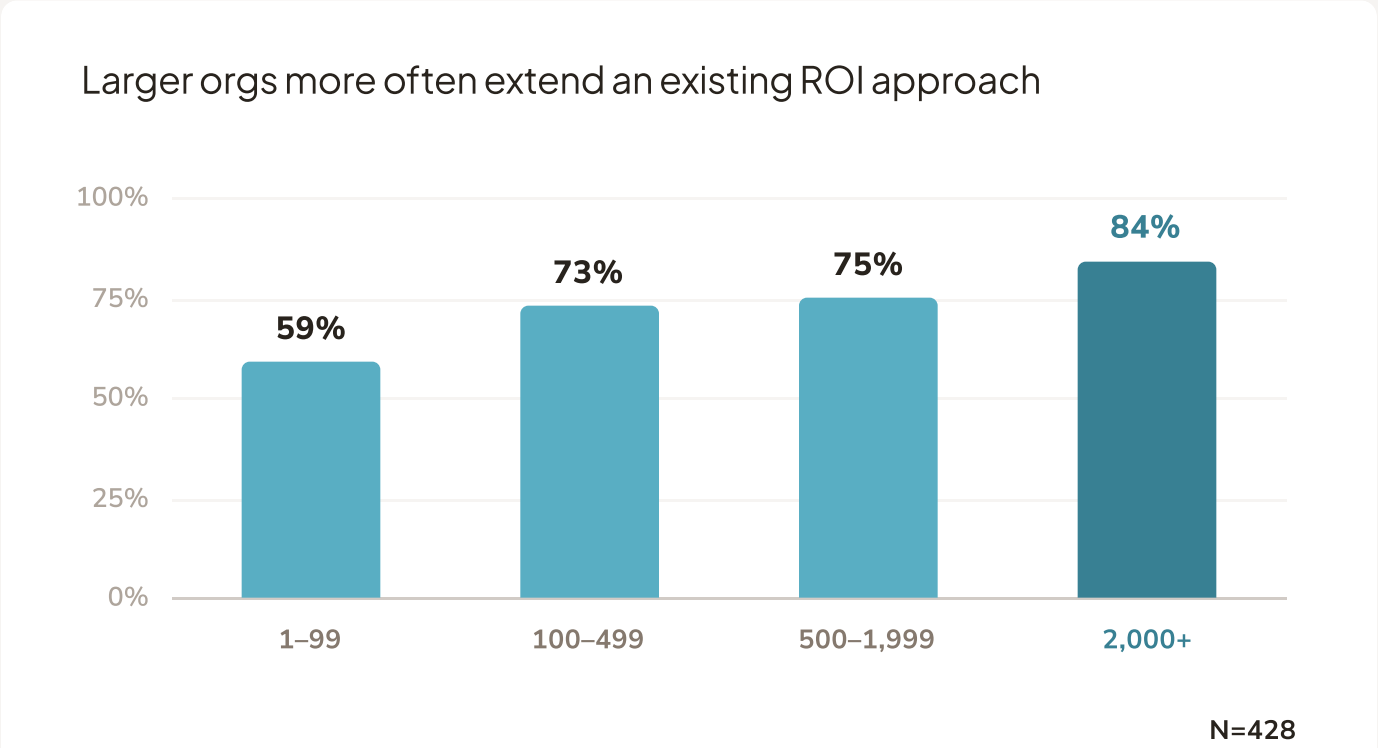
Engineering leaders are confident AI is increasing value, and they attribute that value to increased speed. But if speed is a leading indicator, can they translate early signals into real proof of business value? In other words, is increasing speed with AI actually worthwhile?

Traditional engineering metrics were designed around a delivery process in which writing code was the bottleneck. AI changes that balance. Teams can generate more code without a proportional increase in engineering capacity, while review and validation may require more attention. A team might produce twice as much code without delivering twice as much value.

Yet as transformative and valuable as AI is purported to be, our research found that 74% have simply extended the ROI measurement methods they used before AI existed. Only 4% are building an AI-specific approach.

Extending pre-AI methods is most common in the largest organizations, but also at those with the highest levels of self-reported AI maturity. At Level 5, 87% have extended a pre-AI approach, the highest of any maturity level. Among organizations with 1–99 engineers, 59% extended an existing approach, compared with 84% of organizations with 2,000 or more.





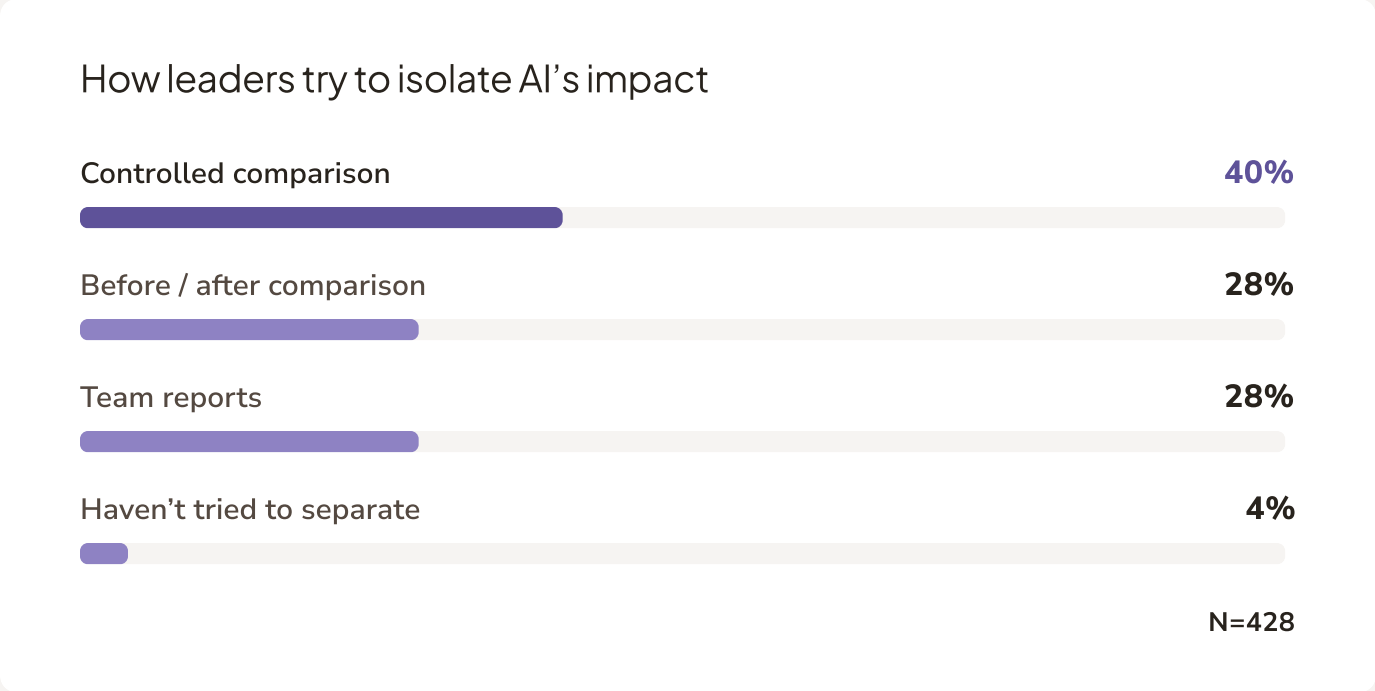
Existing metrics still have an important role. Cycle time, delivery performance, and business outcomes remain useful measures. But AI changes what drives those numbers and how they should be interpreted. Leaders need to account for the work AI accelerates and the additional demands it may create elsewhere in the system.

In our 2025 research, we joked that organizations were measuring AI's impact like it was 1999. A year later, most are still working with the same measurement toolkit.

Separating AI impact from other changes

Measuring AI's contribution is complicated by everything else changing inside an engineering organization. If a company reorganizes engineers into smaller teams to support AI experimentation, for example, and delivery speeds up, how much improvement came from AI, and how much came from the new team structure?

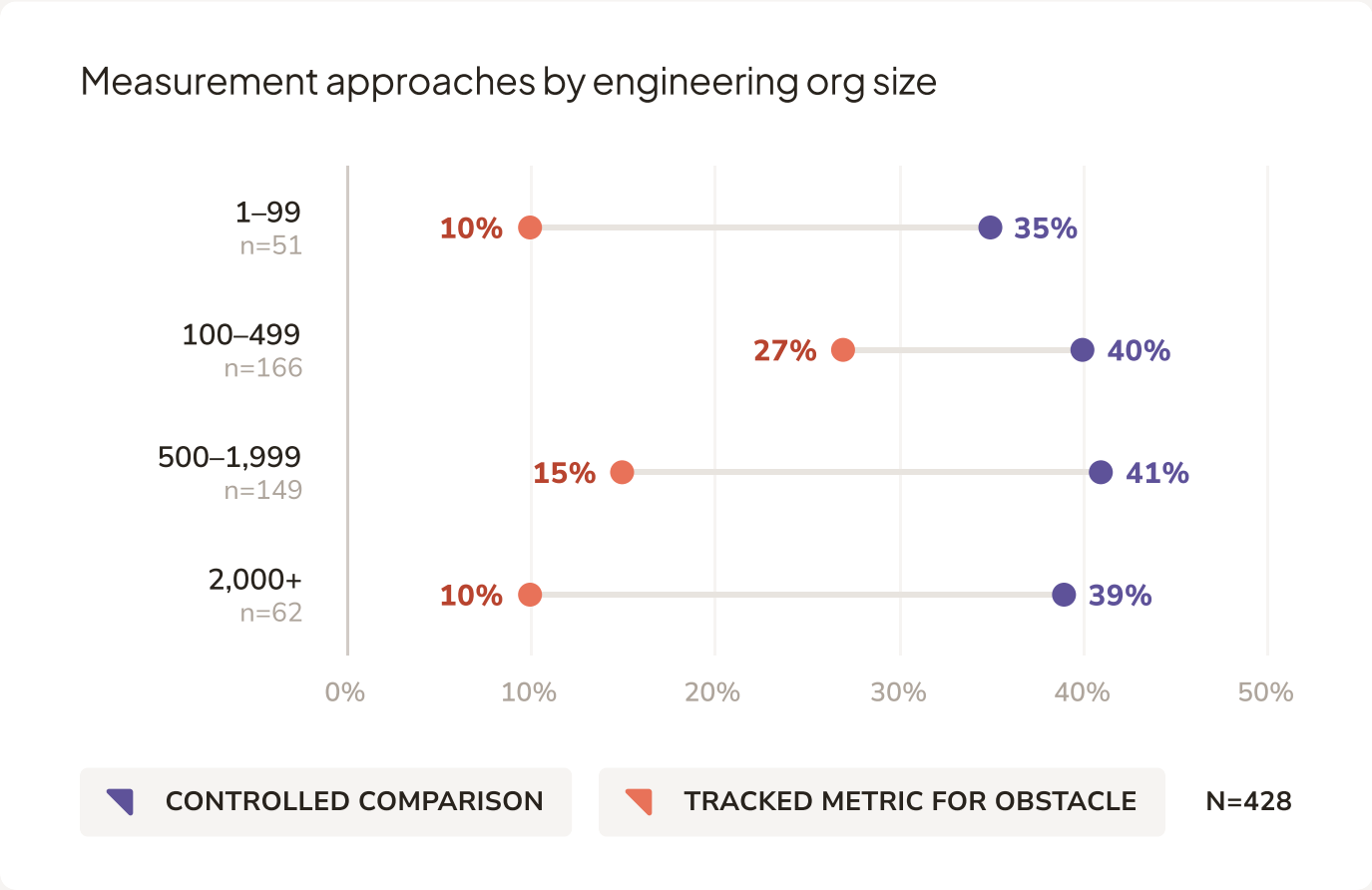
Leaders report using several approaches to make that distinction:



Controlled comparisons can provide stronger evidence of causation, but their quality depends on how the comparison is designed. With AI use this widespread, what serves as the control group? With shadow AI widespread and growing, do organizations really have reliable access to control groups?

Even if the 40% claim is accurate, that still leaves six in ten attributing AI's effect without a controlled comparison attempted at all.

Findings also show little relationship between engineering org size and measurement rigor.

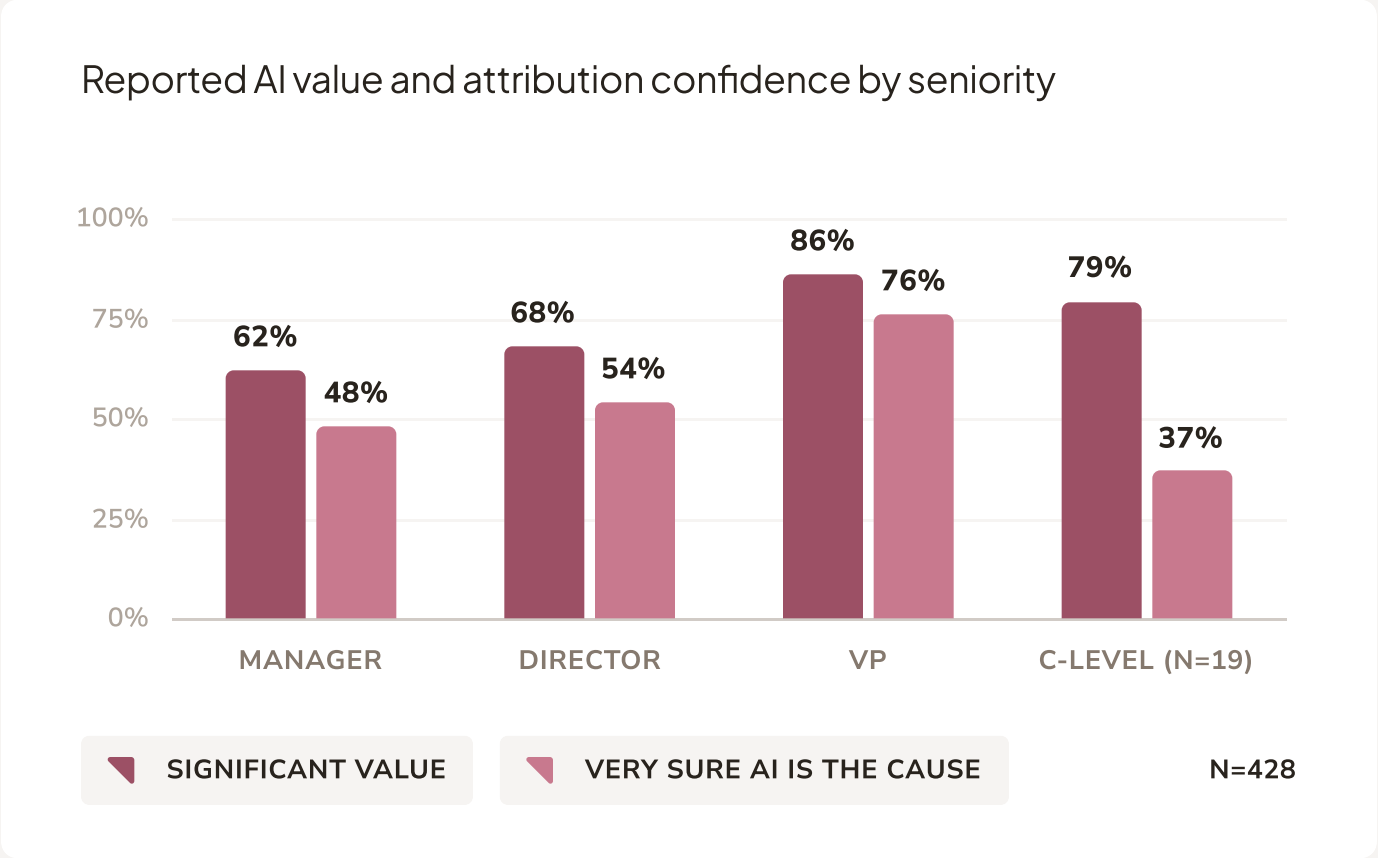


More headcount doesn't necessarily bring stronger measurement practices. The largest engineering organizations are about as likely as the smallest to use a tracked metric when identifying what's limiting AI's value.

> "Everyone is grasping at straws when it comes to understanding and measuring AI. A lot of the measurement skill right now comes down to leaders' experience, mental model, and organizational influence. That shouldn't be so. The industry needs a way to think about engineering value both pre- and post-AI and a way to measure engineering discipline now that some of the rules of coding have changed."
>
> 
>
> **Amy Carrillo Cotten, Director of Client Transformation at Uplevel**

VPs are AI's most confident advocates

VPs stand out for their conviction that AI is delivering results, although that confidence looks quite different in the C-suite.



Among VPs, 86% report significant value, compared with 62% of managers and 68% of directors. VPs are also the most likely to be very sure AI is responsible, at 76%.

That confidence drops sharply at the C-level, where 79% report significant value but only 37% are very sure AI caused it. The C-level sample is small (n=19), so this difference should be interpreted cautiously.

VPs also report using controlled comparisons more frequently than other seniority levels, at 55%. Across VP and C-level respondents, that figure is 54%, compared with 37% among managers and directors.

Yet their approach to diagnosing obstacles looks different. Among VP and C-level leaders, 44% rely primarily on team and leadership consensus, compared with 29% of managers and directors. They also rely less on direct engineer feedback.

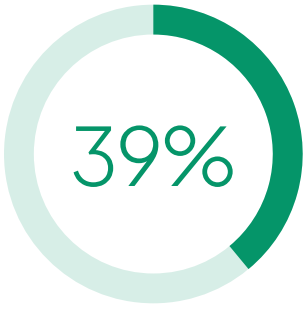
The distinction is revealing. Senior leaders report greater use of controlled comparisons when attributing AI's impact, but tend to rely on consensus when identifying what's preventing further improvement. Measurement rigor in one part of the decision-making process doesn't necessarily carry over to another.

> "VPs are often the most exposed because they are accountable for proving results from AI programs for the C-suite. They need AI's value to be true, whether or not it is. Often there's not yet a culture of psychological safety at the upper levels of organizations to say, 'We don't have the foundational conditions to achieve exponential AI acceleration, so perhaps we should invest in them.' That's been a non-starter since 2023. So they need AI to work — and it does — but perhaps not as well as they want it to."
>
> 
>
> **Amy Carrillo Cotten, Director of Client Transformation at Uplevel**

Three findings to take to your next board meeting

Engineering leaders are increasingly confident that AI is creating value. Our research suggests they should be asking for more of the evidence behind that confidence.

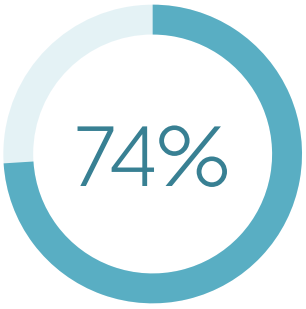
Three findings stand out:



primarily use speed to judge whether AI is working.



identify increased review time as their biggest new source of friction.



have applied their pre-AI ROI measurement approach to AI.

Together, these findings point to a problem with how organizations understand AI's contribution to engineering performance. Faster development can expose constraints elsewhere in the delivery process, while familiar metrics may struggle to explain what those changes mean for business value.

AI maturity isn't an easy button. Organizations at different maturity levels report different obstacles, and even the most mature continue to rely heavily on existing ROI methods. Waiting for teams to gain experience won't necessarily improve their ability to identify what's limiting performance.

Engineering leaders need to keep examining the entire delivery system as AI changes how work gets done. That requires measurement practices that can distinguish local productivity gains from improvements in overall performance, along with the judgment to recognize when the organization's biggest constraint has moved.

Would you recognize AI ROI if you saw it?

You may be confident AI is creating value in your engineering organization. But how well can you distinguish AI's contribution from other changes, and do your metrics capture the results that matter?

Answer three questions from our 2026 survey to see how your measurement practices compare with those of 428 engineering leaders.

Each answer earns 0–3 points, for a maximum score of 9. Higher scores reflect practices that generally provide stronger evidence of AI's contribution and place greater emphasis on business outcomes.

The scoring is Uplevel's assessment framework, informed by the research. Your score reflects how you approach measurement, rather than how much value AI is actually creating.

> "The journey to AI value is a hill-to-valley-to-peak trail with constantly shifting bottlenecks and limited visibility. Leaders need a compass to steer through this."
>
> 
>
> **Amy Carrillo Cotten, Director of Client Transformation at Uplevel**

The AI Reality Check

How do you separate AI's impact from other changes?

ANSWER	POINTS
A controlled comparison (a team using AI vs. one that isn't)	3
A before/after comparison over time	2
What engineers or teams report noticing	1
We haven't tried to separate these	0

What is the diagnosis of your biggest AI obstacle based on?

ANSWER	POINTS
A specific metric or data point we track	3
Direct feedback from engineers or teams	2
Team and leadership consensus based on day-to-day experience	1
Input from an outside advisor or consultant	1
What we're seeing reported by other companies, analysts, or industry research	0

How do you separate AI's impact from other changes?

ANSWER	POINTS
A business outcome (revenue, retention, customer satisfaction)	3
Cost (spend on tools, infrastructure, or headcount)	2
Speed (how quickly work moves compared to before)	1
Output (features shipped, PRs merged, tickets closed)	1
We haven't tied this to a specific measure yet	0

What the Data Reveals

0–3: Mostly unverified

13% of surveyed leaders scored in this range.

Your answers suggest that you're relying primarily on indirect signals or informal assessments of AI's impact. You may be seeing genuine improvements, but your current approach gives you limited evidence for attributing them to AI.

Start by identifying one outcome you expect AI to improve and establishing how you'll measure it.

4–6: Partially measured

63% of surveyed leaders scored in this range.

You have some evidence of AI's impact, but parts of your approach leave important questions unanswered. You may be tracking productivity without connecting it to business results, or relying on consensus to identify what's limiting progress.

Look at your lowest-scoring answer. That's a useful place to examine whether your current methods are giving you enough information to make decisions.

7–8: Strong measurement practices

22% of surveyed leaders scored in this range.

You're using several approaches that can provide useful evidence of AI's contribution. Your next step is to examine where the remaining uncertainty sits, whether that's attribution, obstacle diagnosis, or the connection to business outcomes.

A high score is a good starting point. The quality of the underlying measurements still matters.

9: Strong measurement foundations

Only 2% of surveyed leaders achieved this score.

Your answers indicate that you're using controlled comparisons, tracked metrics to diagnose obstacles, and business outcomes to evaluate AI's value.

Together, these practices provide a strong foundation for understanding AI's contribution. The next challenge is making sure your comparisons are meaningful and your metrics capture what happens across the delivery system.

You've scored your measurement. **Now examine the results.**

Your score tells you how your measurement practices compare with other engineering leaders. It can't tell you which obstacles are limiting AI's value in your organization.

That's what we built StackUp to explore.

StackUp is a free 45-minute session with an Uplevel expert. We'll use a short questionnaire to understand your organization, compare your situation with similar engineering teams, and identify one place to focus your improvement efforts.



Find your AI bottleneck →