DigitalOcean

CURRENTS

AI Agents, Inference, and Implementation

DigitalOcean

CURRENTS

AI Agents, Inference, and Implementation

Currents is DigitalOcean's ongoing research report on trends impacting the growth of digital native enterprises around the world—from Italy to India. For the latest edition, we built on what we learned about how growing tech businesses are using AI since the February 2025 Currents report, with updated insights on emerging trends like AI agents and inference.

At DigitalOcean, we've spent over a decade making cloud infrastructure straightforward for developers. Now, we're bringing that same approach to AI, by providing AI-native businesses and digital native enterprises with the infrastructure and platform tools they need to build the next generation of applications. As the industry's center of gravity moves from building models to running them, DigitalOcean is purpose-built to be the Inference Cloud for teams that need to scale AI in production. Right now, AI implementation is messy—full of hype, hidden costs, and tools that don't quite work together. We conduct this research to cut through the noise and share peer insights on what's actually happening on the ground so that businesses have the insights and tools they need to get ahead of the game.

This report gathered responses from 1100+ developers, CTOs, and founders at a range of global technology companies. We dig into the specific challenges that remain and opportunities that make implementation worthwhile for companies that have boldly moved forward with integrating AI into their production workflows. You'll also find a breakdown of top AI models, tools, and infrastructure trends—including the shift from training models to running inference. This represents a defining moment for the market: while training builds the model, inference runs the product.

Key findings:

- **Organizations have shifted from exploring AI to actively implementing solutions**. The percent of companies actively implementing AI solutions, optimizing AI performance, or treating AI as a core component of their business strategy has grown to 52%, compared to 35% who said the same in 2024.
- **The cost of inference is the #1 blocker for scaling AI**. 49% of respondents identified the high cost of inference as a top challenge.
- **Developers focus on integration over training models**. The vast majority of AI activity does not involve training models from scratch (15% of respondents). Instead, 64% of respondents are integrating third-party AI APIs into applications, and 61% are using a hybrid or niche tools approach rather than a single integrated stack.
- **The future is agentic**. A clear majority (60% of respondents) considered "Applications and agents" to hold the most long-term value in their AI stack. On a related note, agentic spending is increasing—"Applications and agents" is the top-ranked area for AI budget growth in the next 12 months, according to 37% of respondents.

- **AI agents aren't just hype anymore**—they're driving real productivity gains. 53% of companies using AI agents have observed productivity/time savings for employees, and 44% of respondents report the creation of new business capabilities as a direct outcome of using AI agents. While 14% of respondents have yet to see a benefit, the vast majority (67%) have experienced productivity improvements from using AI agents.
- **Single-provider setups are rare—multi-tool complexity is the norm.** Only 23% are using a single cloud provider that combines models, data, and infrastructure. For organizations using multiple tools, the top challenges are all related to complexity and cost.

Our findings around the perceived cost, complexity, and barriers to adoption of AI in enterprises relative to potential ROI underscore the need to simplify the entire AI workflow, providing developers with the tools to build agents and run inference without complexity.
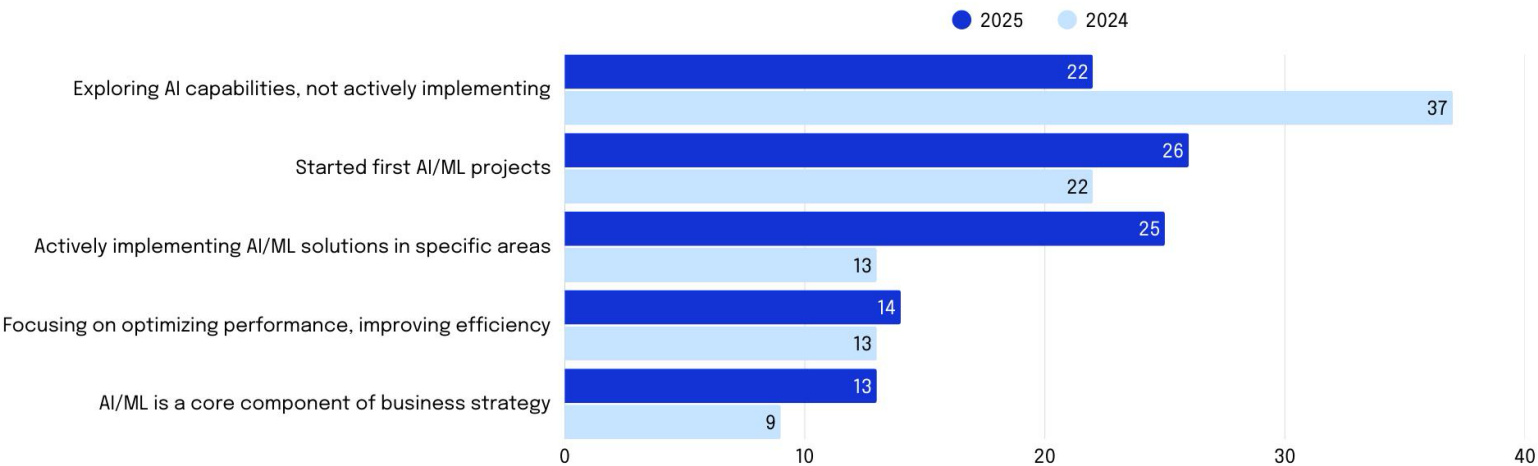
Read on for the full report and results.

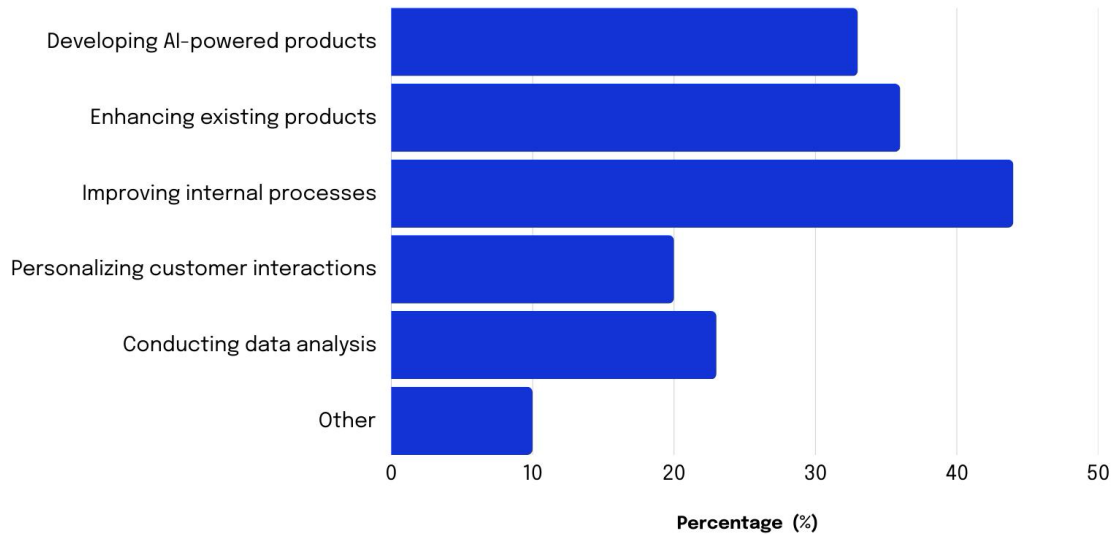## AI usage: From buzz to business value

As the AI market matures beyond the initial hype cycle, broad experimentation has given way to more focused implementation. This has led to a deeper integration of AI for internal purposes, with a clear focus on developer productivity.

Twenty-five percent of respondents are now actively implementing AI solutions, compared to 13% in 2024. Paradoxically, while AI implementation is maturing, the primary measure of AI adoption—the percentage of respondents actively using it—declined slightly from 79% in 2024 to 77% in 2025. This likely correlates with mass testing and high hype in late 2023 and 2024. Now that we're moving to actual usage, it's slightly lower. For organizations that are using AI, there was an increase across the board in leveraging it for internal purposes. This suggests a deeper integration of AI within adopter companies.

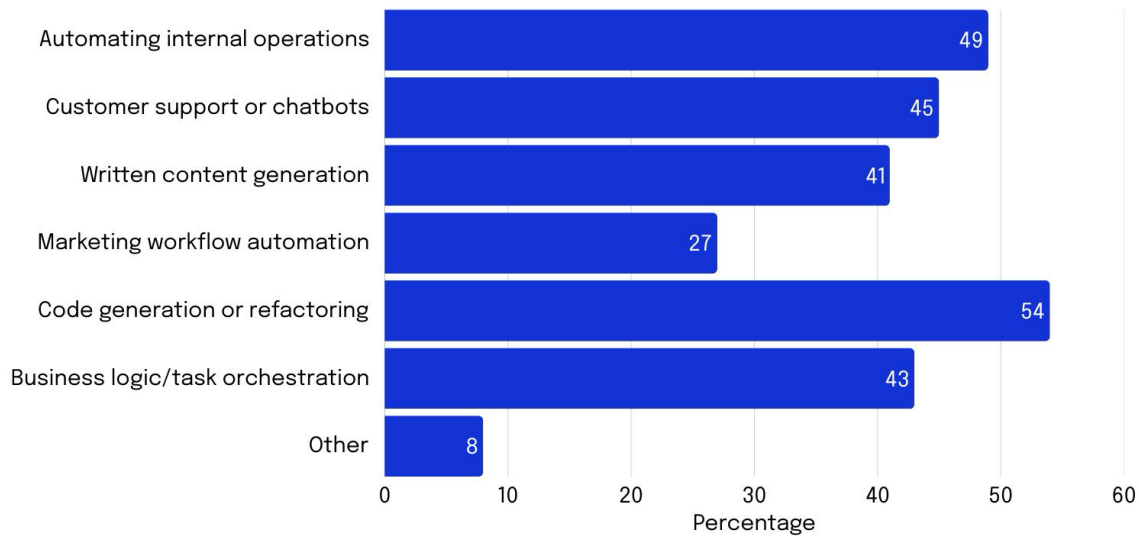## How would you rate your organization's current AI integration?

● 2025   ○ 2024

	2025	2024
Exploring AI capabilities, not actively implementing	22	37
Started first AI/ML projects	26	22
Actively implementing AI/ML solutions in specific areas	25	13
Focusing on optimizing performance, improving efficiency	14	13
AI/ML is a core component of business strategy	13	9

How is your company currently using AI?

What agent use cases is your business pursuing?

When asked what agent use cases they are pursuing, the top answer for 54% of respondents was code generation/refactoring.

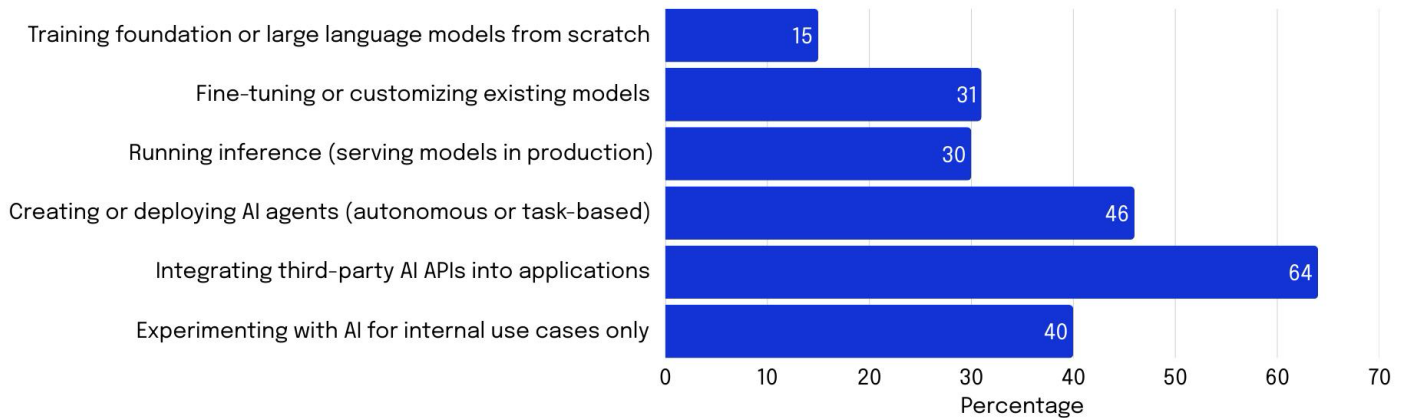Developer tools lead the pack for AI products being built (36%), with business intelligence/analytics (32%), customer support automation (30%), and productivity assistants (30%) close behind. All four speak to the same priority: doing more with less.

How are organizations currently engaging with AI? Sixty-four percent of respondents are integrating third-party AI APIs into their applications, while 46% are creating or deploying AI agents (autonomous or task-based).

Which of the following AI activities does your organization engage in?

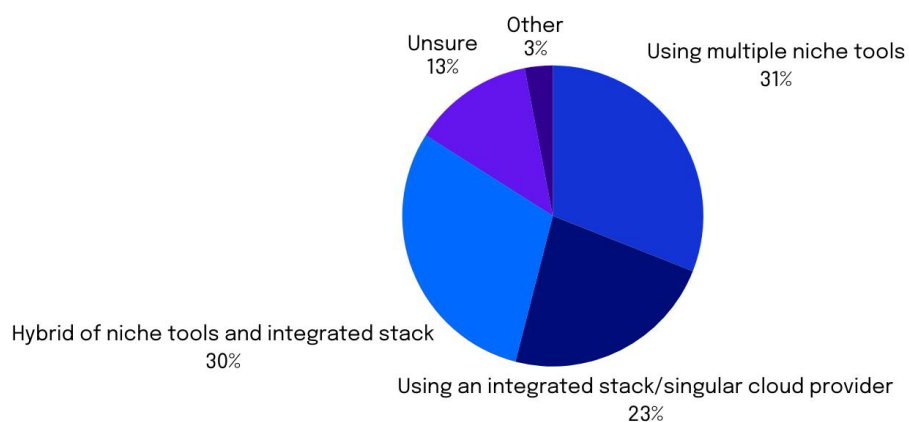| Activity	Percentage
Training foundation or large language models from scratch	15
Fine-tuning or customizing existing models	31
Running inference (serving models in production)	30
Creating or deploying AI agents (autonomous or task-based)	46
Integrating third-party AI APIs into applications	64
Experimenting with AI for internal use cases only	40

AI infrastructure: The stack is sprawling

The AI space is competitive, but it's not winner-takes-all when it comes to the infrastructure (compute, storage, GPUs) powering digital native enterprises and AI-native businesses.
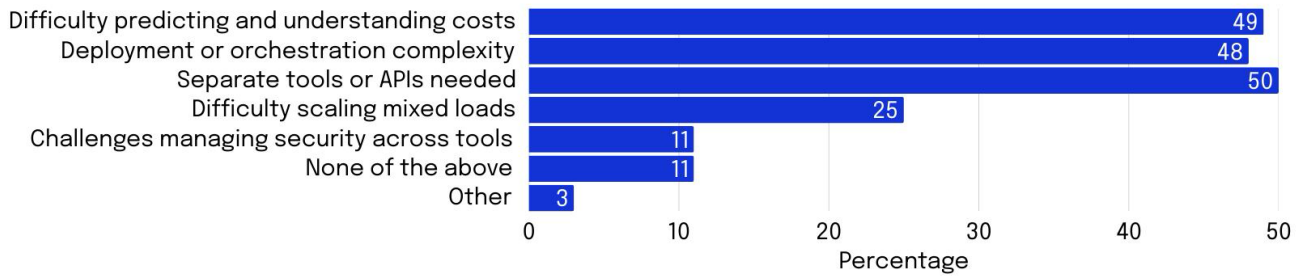
Sixty-one percent of respondents are using multiple tools stitched together for their AI infrastructure, or a hybrid of multiple tools alongside an integrated stack. Only 23% of respondents are using a single cloud provider that combines models, data, and infrastructure. The data shows that for most, the AI stack has become a management burden. To solve this, the market is shifting toward a dedicated Inference Cloud approach—where infrastructure, orchestration, and cost predictability are treated as a single, integrated system rather than a collection of disconnected tools.

Pricing (75%) and ease of use (61%) are important factors when selecting AI infrastructure. For organizations using multiple tools, the top challenges are all related to complexity and cost, with survey respondents specifically identifying: separate tools or APIs needed (50%), difficulty predicting and understanding costs (49%), and deployment or orchestration complexity (48%), and challenges managing security across tools (34%).

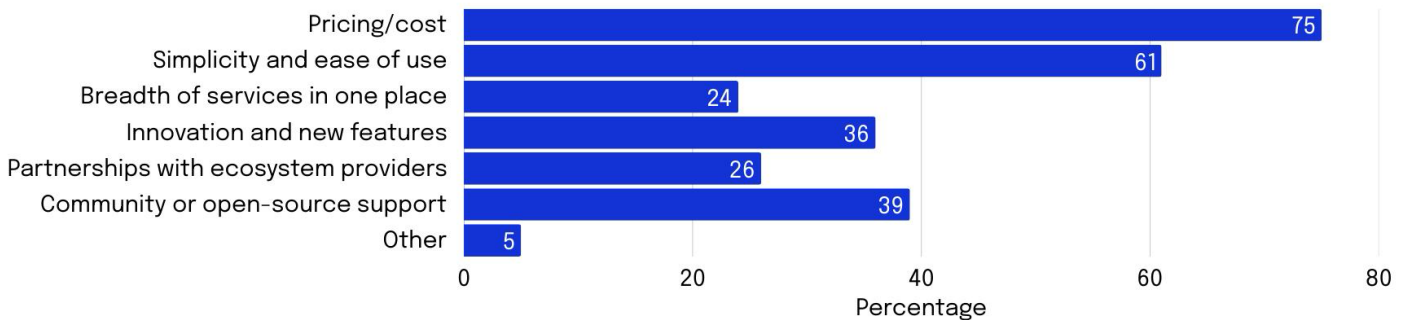## Which of the following best describes your organization's current approach to building AI solutions?

- Other: 3%
- Unsure: 13%
- Using multiple niche tools: 31%
- Hybrid of niche tools and integrated stack: 30%
- Using an integrated stack/singular cloud provider: 23%

What challenges arise from using multiple tools and/or clouds to build your AI solutions?

Challenge	Percentage
Difficulty predicting and understanding costs	49
Deployment or orchestration complexity	48
Separate tools or APIs needed	50
Difficulty scaling mixed loads	25
Challenges managing security across tools	11
None of the above	11
Other	3

Percentage

When selecting AI infrastructure or platforms, what factors are most important?

Factor	Percentage
Pricing/cost	75
Simplicity and ease of use	61
Breadth of services in one place	24
Innovation and new features	36
Partnerships with ecosystem providers	26
Community or open-source support	39
Other	5

Percentage

AI models: OpenAI leads, but the race is tightening

The market for the foundational models that power AI agent development remains fairly consolidated around a few key providers.

OpenAI, Anthropic, and Google are the three most popular LLM providers for AI agent development. OpenAI holds the early mover advantage, used by 72% of respondents, followed closely by Anthropic and Google, used by 47% and 50% of respondents, respectively.
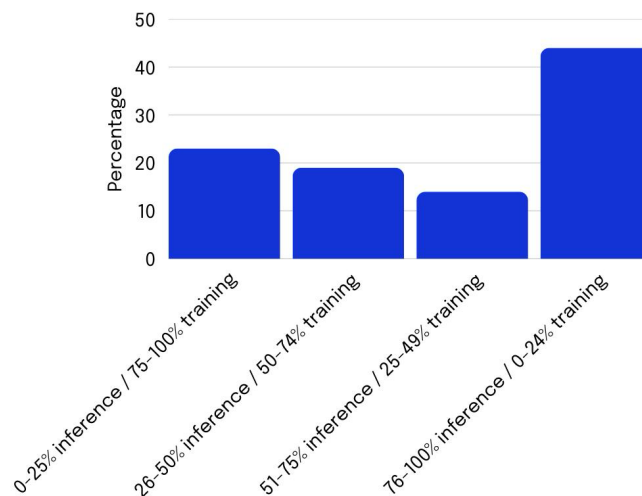
That said, open-source models are carving out real ground. Meta's Llama and DeepSeek—both open source— are each used by 21% of respondents, respectively, offering teams more flexibility and control over their AI stack. DeepSeek's rise is especially notable given that it only hit the market in late 2024 and has already nearly caught up with Llama in adoption.

Inference: Integration beats building from scratch

The AI industry's center of gravity has shifted from training models to the post-training phase of inference—putting AI models to work to generate predictions and power applications—where the majority of budgets are now spent. This practical focus on integrating and deploying models has revealed a new primary obstacle to growth: the significant cost of running inference at scale.
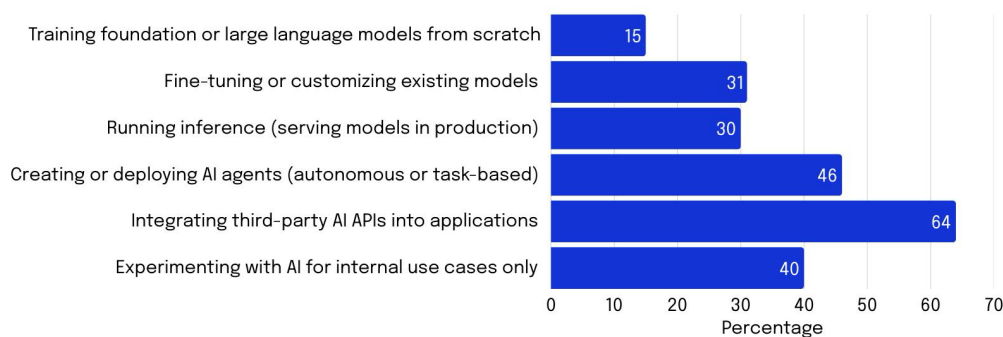
Nearly half (44%) of respondents report spending the majority (76-100%) of their AI budget on inference, not training.

Roughly how much of your AI spend goes to inference versus training?

This has rewritten job expectations for developers, where success is now measured as much by their ability to integrate as their ability to build. Our research shows that the majority of AI work happens after the initial training phase, as only 15% of respondents focus primarily on training models from scratch. Instead, the most common AI activities that organizations currently engage in include, for 64% of respondents, integrating third-party AI APIs into applications.

Which of the following AI activities does your organization currently engage in?
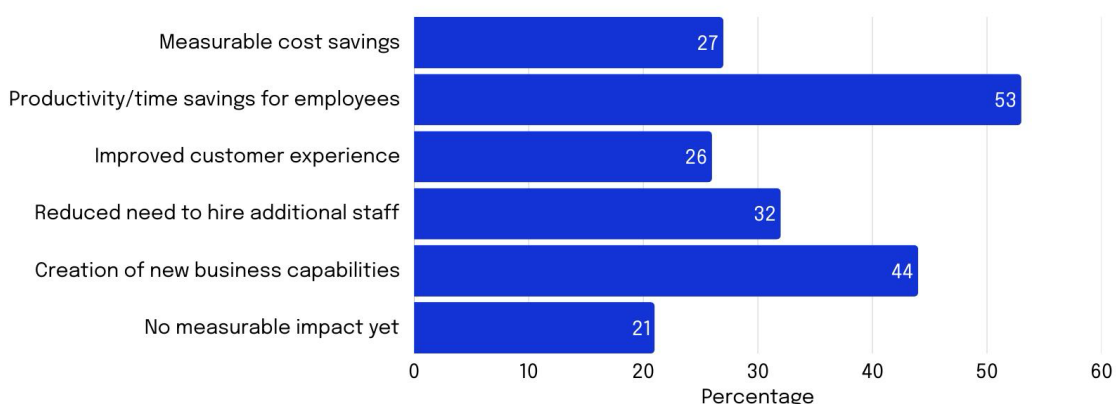

When asked what limits their ability to scale, 49% of respondents identified the high cost of inference at scale as the #1 blocker to scaling AI, followed by lack of internal knowledge and resourcing (37%), and the complexity of deploying/serving models in production (36%).
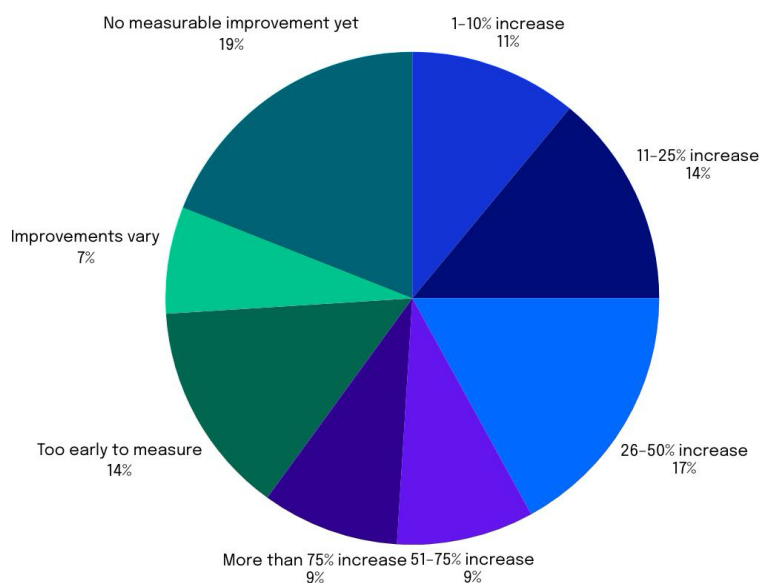
AI agents: Real gains, early days

Agents, autonomous AI systems that can perform single or multiple tasks independently, have risen in popularity; however, questions remain about their actual value. Our findings show a gap between those who find real productivity gains from agents and those who aren't seeing the benefits just yet, suggesting that it's still early days for overall agent adoption.

FIfty-three percent of respondents shared that using AI agents has led to productivity/time savings for employees at their organization, suggesting that AI agents are worth the investment. Although 14% of respondents have yet to see a benefit, the vast majority (67%) have experienced productivity improvements from using AI agents, with 25% reporting a 1-25% increase, 17% reporting a 26–50% increase, 9% reporting a 51-75% increase, and 9% reporting a massive increase of more than 75%. Going beyond time savings, 44% of respondents reported the creation of new business capabilities as a direct outcome of using AI agents.

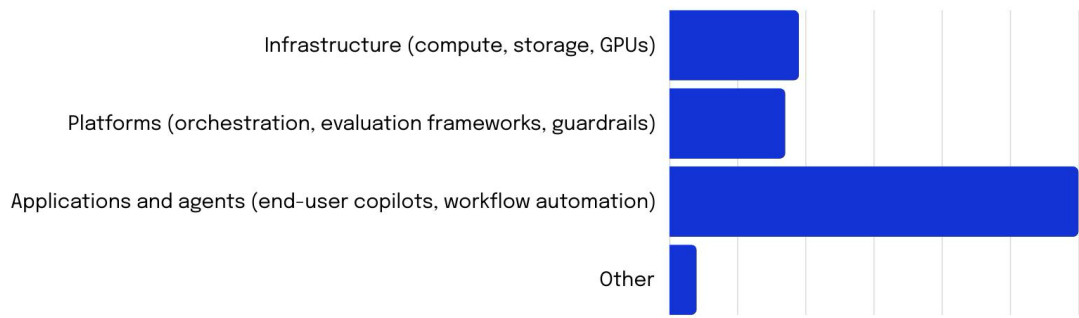## What outcomes have you observed from using AI agents?

Outcome	Percentage
Measurable cost savings	27
Productivity/time savings for employees	53
Improved customer experience	26
Reduced need to hire additional staff	32
Creation of new business capabilities	44
No measurable impact yet	21

What level of productivity improvement has your organization observed from implementing AI agents?

No measurable improvement yet — 19%
1-10% increase — 11%
11-25% increase — 14%
26-50% increase — 17%
51-75% increase — 9%
More than 75% increase — 9%
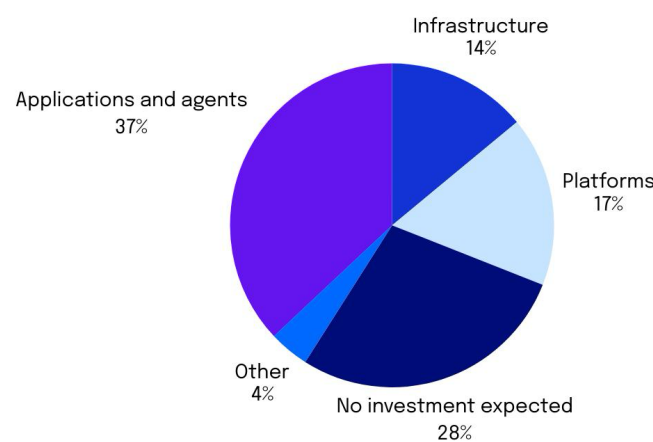Too early to measure — 14%
Improvements vary — 7%

There is a strong consensus on where the long-term value of the AI stack lies: 60% of respondents chose applications and agents, compared to just 19% for infrastructure and 17% for platforms.

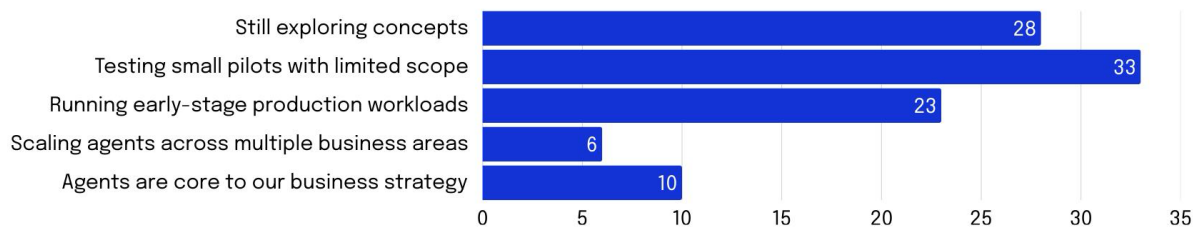### Where do you see the greatest long-term value and opportunity in the AI stack?

Applications and agents are also the top-ranked areas for AI budget growth over the next 12 months, according to 37% of respondents, signaling where investment is going next.

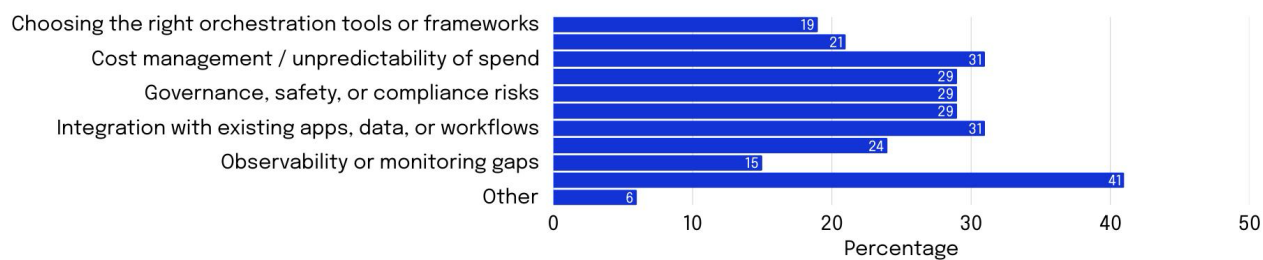### Over the next 12 months, where do you expect your AI budget growth to be focused?

When it comes to their organizations' readiness for scaling AI agents, 33% of respondents reported testing small pilots with limited scope—organizations are still exploring concepts (28% of respondents) and running early-stage production workloads (23% of respondents).

How would you describe your organization's readiness for scaling AI agents?

Alongside all of the positivity and potential are several challenges to consider. According to 41% of respondents, the #1 barrier for scaling the use of agents is reliability concerns, followed by integration with existing apps for 31% of respondents. The takeaway? Teams need platforms built for reliability and easy integration—not more duct tape.

What are the main barriers your organization faces in building and scaling production-level AI agents?

Bar chart showing percentages:
- Choosing the right orchestration tools or frameworks: 19
- (unlabeled): 21
- Cost management / unpredictability of spend: 31
- (unlabeled): 29
- Governance, safety, or compliance risks: 29
- (unlabeled): 29
- Integration with existing apps, data, or workflows: 31
- (unlabeled): 24
- Observability or monitoring gaps: 15
- (unlabeled): 41
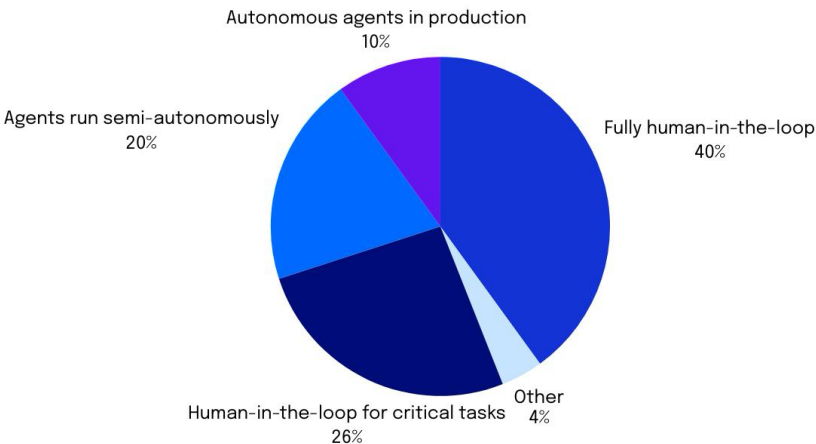- Other: 6

X-axis: Percentage (0 to 50)

While the use of AI agents is on the rise, fully autonomous production deployment is still at early stages of readiness. 50% of respondents say they are experimenting with or deploying AI agents. Of these, only 10% are scaling agents or see agents as core to their business strategy. Expect to see the expansion of AI agents in 2026, as 38% of respondents who haven't yet explored agents report that they will start experimenting with or deploying agents at that time.

Looking toward 2026, 26% of respondents said they have planned pilots/testing for AI agents, 44% said they don't expect their organization to experiment with AI agents, and 17% said they were unsure. For organizations without 2026 plans, this is particularly notable because we'd expect developers to be involved in discussions about potential experimentation regardless of whether a commitment to move forward has yet to be made.
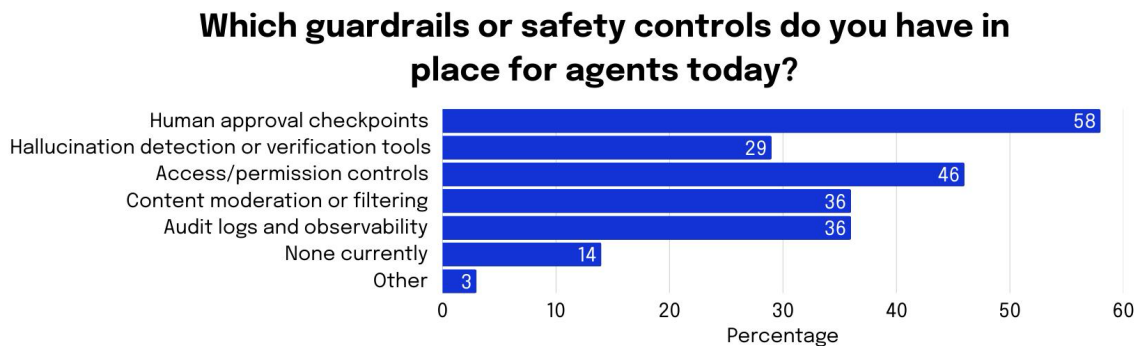
A failure to experiment with and ultimately ship AI agents is also a missed opportunity, because 32% of respondents shared that a top outcome observed from using AI agents was a reduced need to hire additional staff.

For those already using agents, the most common type of AI agent currently in use is the single-task agent (44%), followed by multi-task agents (29%) and multiple agents working together (17%). Demonstrating the early stage of production-level agents, 40% of respondents still have all agent outputs reviewed by a human, and only 10% have fully autonomous agents in production.

How do you currently manage the level of control over your AI agents?

Pie chart:
- Autonomous agents in production: 10%
- Agents run semi-autonomously: 20%
- Fully human-in-the-loop: 40%
- Human-in-the-loop for critical tasks: 26%
- Other: 4%

In terms of guardrails, 58% of respondents report using human approval checkpoints.

Which guardrails or safety controls do you have in place for agents today?

Guardrail	Percentage
Human approval checkpoints	58
Hallucination detection or verification tools	29
Access/permission controls	46
Content moderation or filtering	36
Audit logs and observability	36
None currently	14
Other	3

Conclusion: The future of building with AI—don't get left behind

Reinforcing our February 2025 Currents research, this report finds that there's still a large gap between the first-movers, who are actively building production workflows with AI, and the holdouts, who have yet to adopt AI use formally in their organizations. There's also an emerging middle-ground of organizations that are starting to experiment, but not yet realizing the productivity and business opportunity gains that come with using AI in production.

Inference is the substrate for this agentic future. As AI applications evolve into autonomous systems, the demand for high-throughput, low-latency execution will only intensify. Reliability of agent outcomes and the cost of inference, as well as finding the ideal integrated solution for handling all aspects of running AI/ML workloads at scale continue to be top challenges that can get in the way of realizing the full potential of AI in an enterprise organization.

However, the data suggests a big picture takeaway for enterprises to act on now: Don't wait until AI agents are widespread to start using them. At a minimum, start experimenting now. Then, start planning how to go from experiments to production workflows to hit the ground running in 2026.

The likely outcome of getting started now is achieving the productivity gains and new business opportunities that first-mover organizations are already benefiting from.

Methodology

This survey was conducted via an online Qualtrics-based survey between the dates of October 9, 2025 and November 21, 2025, which garnered 1100+ responses. The questionnaire was developed by DigitalOcean and was distributed via link through email to both DigitalOcean customers and non-customers.

Approximately 25% of respondents were full-stack developers, 17% were CEOs, founders, or owners, and 11% were back-end developers. CTOs made up 7%, while systems architects and system administrators accounted for 3%, each. All other roles, spanning from DevOps specialists to marketing professionals, summed up to 54%. The remaining 5% fell into the "Other" category, and include titles such as machine learning engineer, non-technical C-suite executives, data scientists, students/educators/academic researchers, and freelance technology consultants.

Respondents represented 102 countries, with 28% being in the US, 7% in the United Kingdom, 6% in Canada, 4% in India, 3% each in Germany and the Netherlands, and 2% in Italy.

Build with DigitalOcean's Gradient™ AI Inference Cloud

At DigitalOcean, we've spent the last decade plus building integrated infrastructure offerings for developers with diverse needs, including those that encompass the emerging trend of developers as integrators. Our Gradient AI offerings, including Gradient AI Platform and Gradient AI Infrastructure, is built specifically for AI/ML workloads and users at digitally native enterprises and AI-native companies, simplifying costs and reducing the complexity to get started.

DigitalOcean continues to improve and expand AI infrastructure offerings to support developers and organizations that want to take advantage of the game-changing possibilities with helpful guardrails for testing and deploying agentic workflows. Character.ai moved production workloads to DigitalOcean's Inference Cloud to manage over a billion queries per day. By using our platform powered by AMD Instinct™ GPUs, they doubled their production inference throughput and improved cost efficiency by 50%.

DigitalOcean has the AI/ML solutions you need in our Gradient™ AI Inference Cloud, from powerful GPUs to our Gradient AI Platform that makes AI accessible to anyone. Sign up or speak to our sales team about your AI/ML needs.