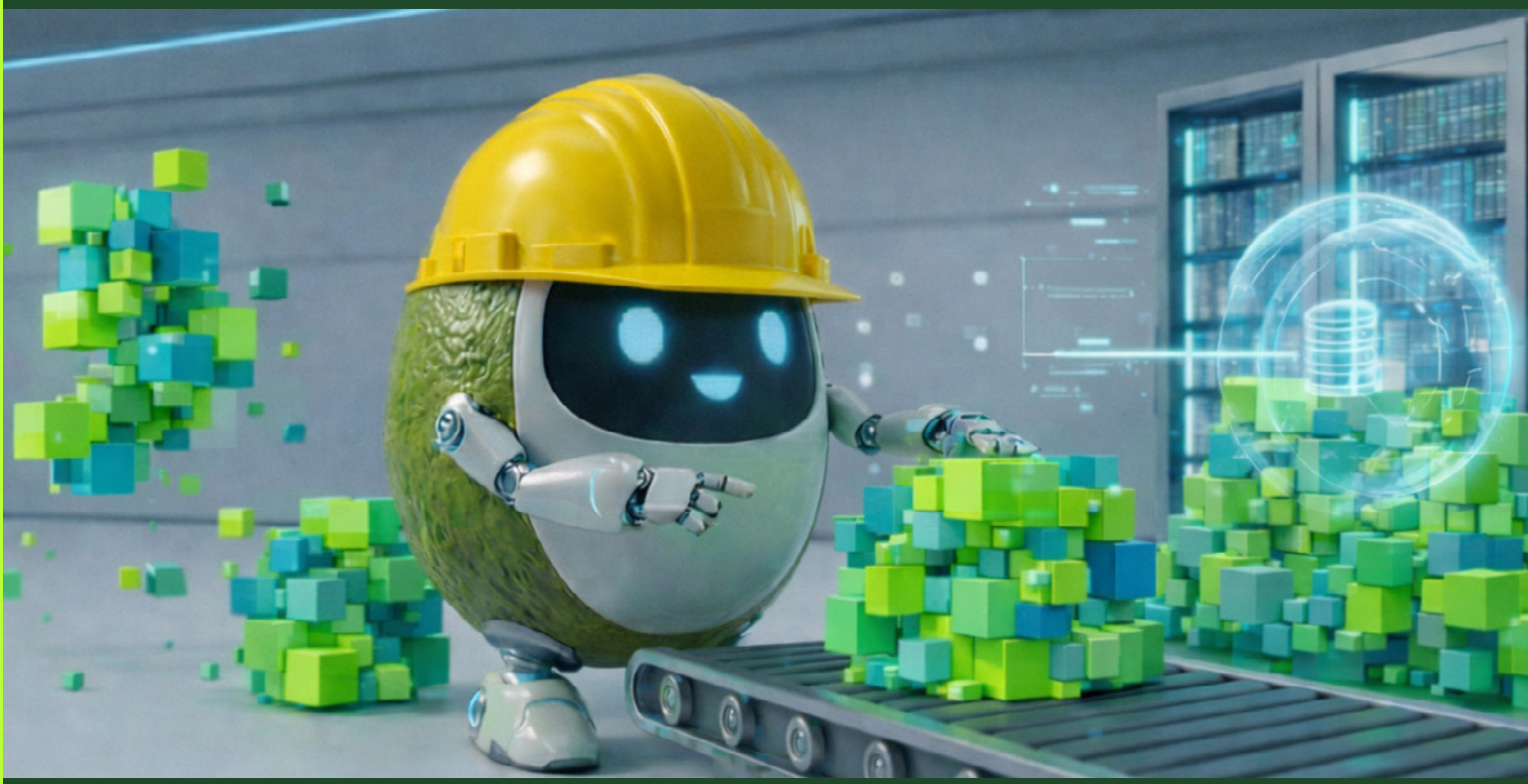
Arango

The AI Context Gap

Why Production AI Is Outgrowing App-by-App Infrastructure



New survey finds that 80% of AI leaders say fragmented data infrastructure is a barrier to scaling trusted AI.

Original research commissioned by Arango | Conducted by TrendCandy



Introduction

Enterprise AI is moving into production. Leaders must now give applications and agents reliable business context while controlling the effort and cost required to operate them at scale. Business context includes the information an application needs to interpret a request, connect relevant records and produce a useful response. Maintaining that context requires attention to data freshness, permissions and relationships as well as retrieval quality.

Arango's research examines the gap between AI ambitions and supporting infrastructure. Many teams assemble data, integrations and retrieval separately for each initiative. Reusing context offers a practical path forward.

Abstract

77% of AI leaders are in production or moving applications there, but only 22% run multiple production applications. Deployment is advancing faster than enterprise-wide maturity.

80% identify fragmented infrastructure as a barrier to scaling trusted AI. 49% build separate or mostly separate context layers; only 19% mostly use shared infrastructure or have a shared foundation. Yet 71% expect shared enterprise context infrastructure within two years.

63% expect significant architectural changes or major redesign to support ten times more agents. 78% expect independently built context layers to create major technology debt. Reusable context is becoming an enterprise infrastructure priority.

Survey Methodology

Arango commissioned independent research firm TrendCandy to survey 260 AI and Tech Leaders about fragmented data infrastructure, enterprise context and production AI. The margin of error is +/-6% at the 95% confidence level. Respondents were sourced through a double-blind online survey using global panels complying with ESOMAR, MRS, CASRO, MRA, ARF, MRIA, AMA and AMSRO standards. All were compensated, and data quality was verified manually and automatically.

What you will learn

- Why fragmentation and separate context layers impede scale.
- How shared context can reduce repeated engineering work.
- Where infrastructure automation needs human oversight.
- How agent growth changes architecture planning.
- What leaders can do next.



Who this eBook is for

- Chief Technology Officers
- Chief AI Officers and Heads of AI or ML Platform
- Chief Information Officers
- Chief Data Officers and VPs of Data and AI
- VPs and Heads of Data, Data Platform and Engineering
- Enterprise and Data Architecture Directors

Respondent Breakout

- 63% of respondents were from the US
- 37% of respondents were from EMEA including Israel
- 100% of respondents were from organizations with at least 100 employees
- 87% of respondents were from organizations with over 1,000 employees
- Common titles represented: Enterprise architect, Platform engineer, AI/ML Engineer
- Common industries represented: Tech, Manufacturing, Healthcare, Business services

© Arango 2026 / arango.ai

1 AI production is ahead of architectural maturity

Production deployments do not automatically translate into scalable operations.

77% are in production or moving AI applications from pilot to production.

Only **22%** run multiple AI applications in production.

Getting one application into production differs from supporting AI across departments and workflows. Infrastructure for a single application often does not stretch to many.

KEY INSIGHT: One application in production isn't enterprise AI maturity.

WHAT IT MEANS FOR LEADERS: Make today's use cases productive while preparing their context for reuse.

2 Fragmentation blocks trusted AI at scale

AI needs context from customer records, databases, documents and other disconnected systems.

80% say fragmented data infrastructure is a barrier to scaling trusted AI.

KEY INSIGHT: When each application runs on its own mix of data and integrations, seeing the full picture of dependencies becomes difficult.

WHAT IT MEANS FOR LEADERS: Treat context as shared infrastructure with clear ownership, rather than something each team assembles independently.



Explore an AI context architecture assessment with Arango to prepare for applications beyond your first production deployment.

FURTHER READING: Arango's Building Trusted Business Context for Enterprise AI.

What stands in the way of shared enterprise context

73%	Data fragmentation
53%	Integration complexity
51%	Data governance
45%	Security requirements
43%	Legacy infrastructure
37%	Skills shortages
35%	Retrieval complexity
30%	Unclear ownership
29%	Budget constraints
2%	Other

© Arango 2026 / arango.ai

3 Shared context has support but limited adoption

49% build separate or mostly separate context layers for AI applications. Only **19%** mostly use shared infrastructure or have a shared enterprise context foundation.

71% expect enterprise context to be managed as shared infrastructure over the next two years.

20% say they build context for individual AI applications mostly or entirely through shared infrastructure.

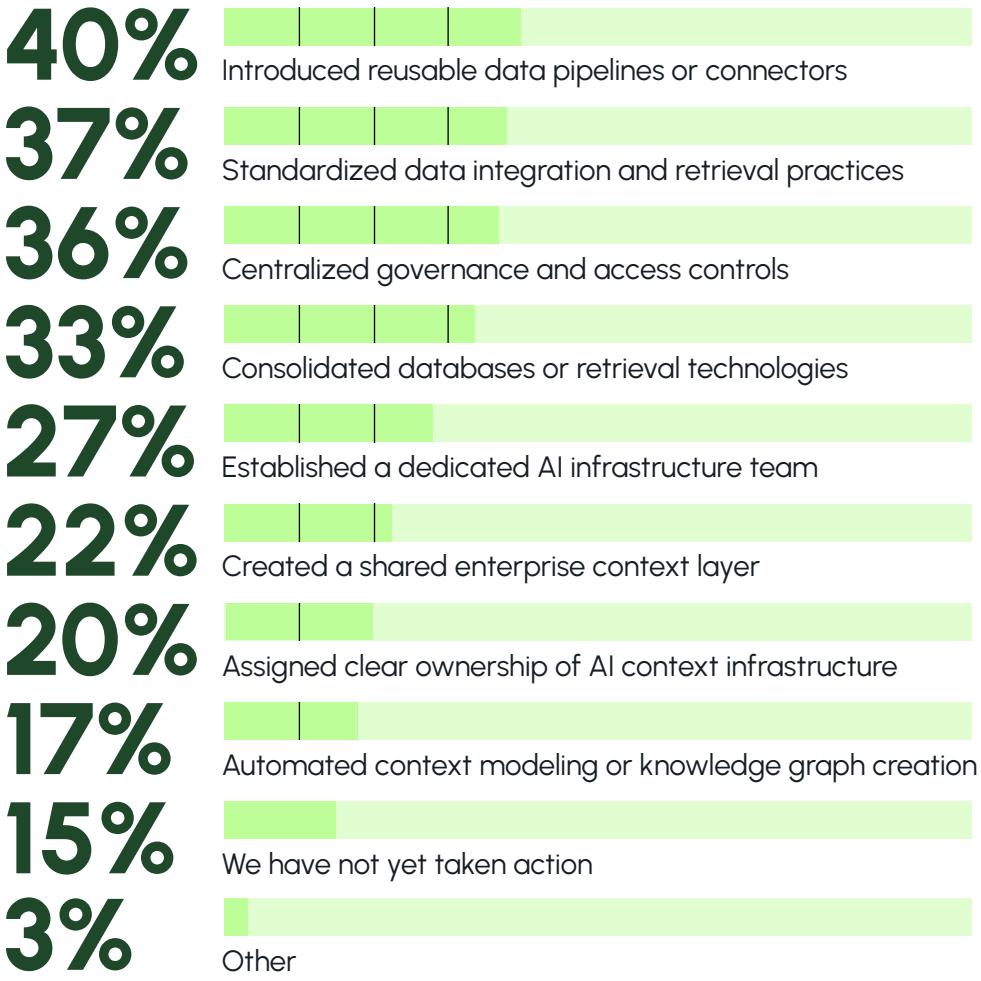
81% say a shared enterprise context layer is important to their organization's AI strategy.

70% favor shared AI data infrastructure over application-specific infrastructure.

About four in five see shared context as strategically important; about one in five report shared infrastructure. Separate layers repeat integration, governance and upkeep.

Where teams are starting

Most are building components first: pipelines, integration and governance.



Respondents could select multiple actions. Only **22%** have created a shared layer; **17%** have automated context modeling or knowledge graph creation, and **20%** have assigned ownership. Without ownership and automation, scaling remains a manual effort.

KEY INSIGHT: The gap is execution, not intent.

WHAT IT MEANS FOR LEADERS: Name an owner for enterprise context and automate its modeling to avoid adding separate layers with every application.

Discuss your next production AI use case with Arango and plan for context reuse.

FURTHER READING:
Context Is an Architectural Imperative.

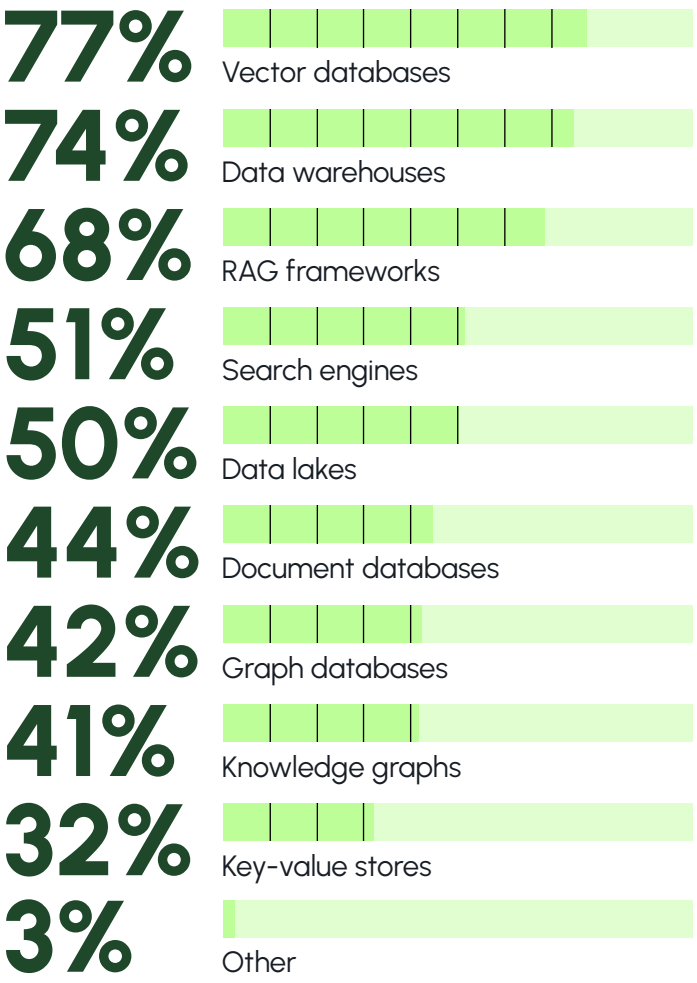
© Arango 2026 / arango.ai

4 Infrastructure adds engineering effort and cost

Models attract attention, but the infrastructure feeding them can consume more engineering effort than value-delivering AI work.

72% use three or more data technologies to give AI agents context.

The context technology mix



Different systems serve legitimate purposes. Repeated integration, maintenance and coordination add complexity and cost.

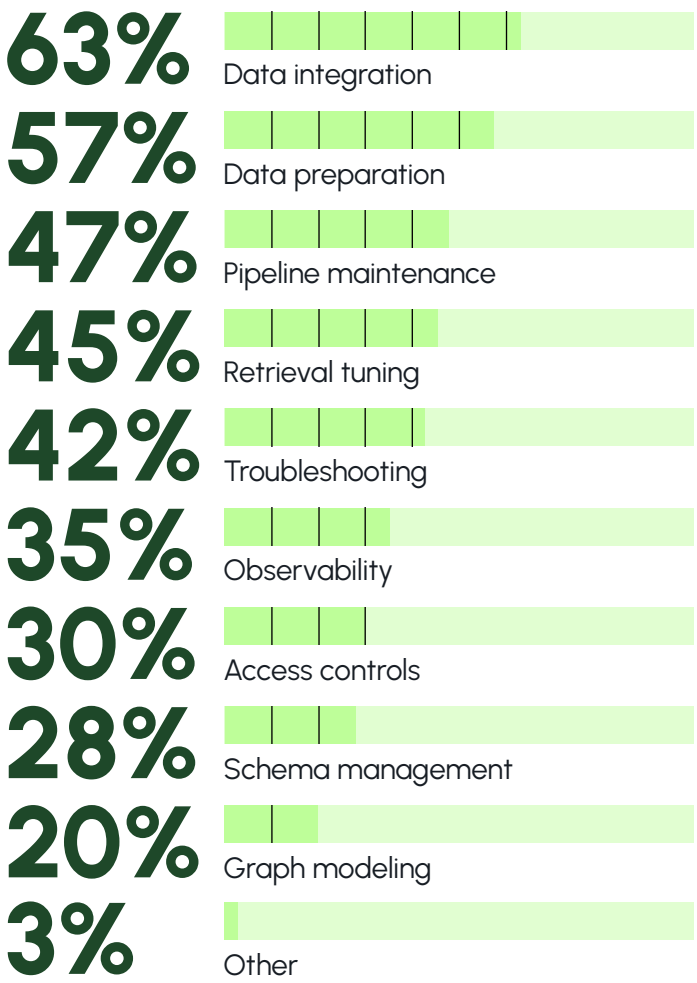
Where the effort goes

69% say managing disparate data for accurate responses takes more engineering effort than the AI investment delivering value.

78% of organizations with generative AI in production report increased infrastructure management and operations (n=196).

76% believe fewer supporting data and infrastructure systems would accelerate development.

Engineering activities consuming time



The top three activities prepare data before an agent produces an answer. Meanwhile, **23%** have not identified an effective approach to deploying AI at scale (n=251).

Key insight: AI moves faster when teams spend less time making data usable.

What it means for leaders: Count the systems behind each application. Consolidate where practical to reduce integration, upkeep and failure points.

© Arango 2026 / arango.ai

5 Automation needs transparency and control

Production retrieval still needs tuning; knowledge structures and data connections need maintenance.

The manual burden

69% say managing and tuning retrieval is a significant operational burden.

65% say retrieval requires too much manual tuning.

55% say knowledge graph creation is a significant operational burden.

Where leaders see the next gains

65% expect infrastructure automation to matter more for developer productivity than further model improvements over the next two years.

The trust gap

44% would be comfortable letting automated systems decide how enterprise data connects for AI; 33% would be uncomfortable.

Key insight: Leaders expect productivity gains from infrastructure automation. Transparent, reviewable processes can help address adoption concerns.

What it means for leaders: Automate retrieval tuning and knowledge graph creation. Choose tools that explain data connections and let teams review and adjust results.

Explore Arango's approach to automated context creation and retrieval.

FURTHER READING: Arango's AutoGraph guide.

6 Ten times more agents and only 1 percent ready

The scale is coming

61% say agents are forcing an infrastructure rethink to a great extent.

60% expect to operate ten or more agents at scale.

Building app by app won't scale

62% say this approach looks unsustainable as agents multiply.

- **80%** agree reusable context will become a strategic requirement as agent counts grow.
- **Today's architectures aren't built for it**
- **63%** would need significant changes or major redesign to support ten times more agents.
- Only **1%** could support that growth without significant changes.

These are expectations, not observed future outcomes.

Governance, security and data consistency rank above performance. Agents operating across the enterprise need trusted, governed context.

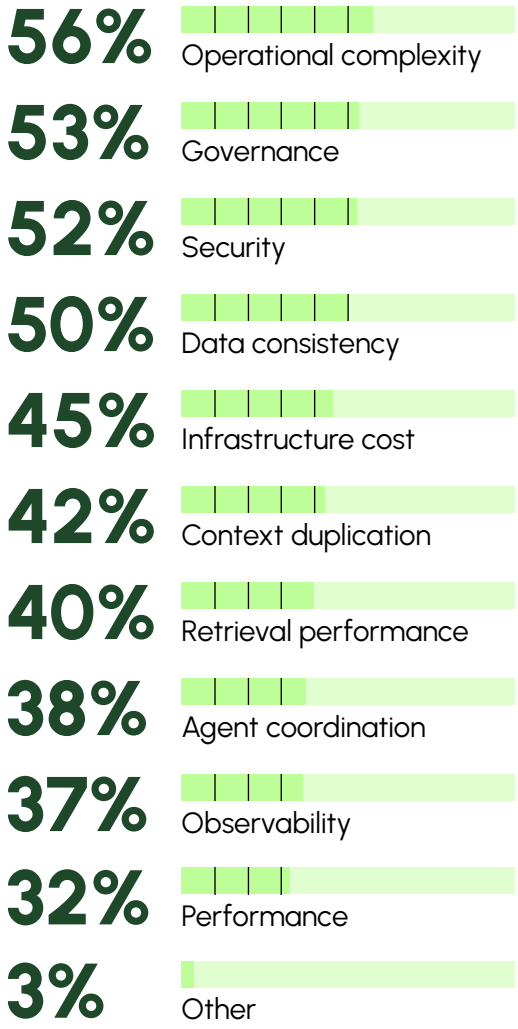
78% agree that team-by-team context layers will create major enterprise technology debt.

69% worry that more agents could create more complexity than productivity unless data architecture changes.

81% agree companies will scale AI by sharing context across the business rather than rebuilding it for each initiative.

What it means for leaders: Test architecture against ten times today's agent count. Build governance, security and consistency into shared context before retrofitting becomes expensive.

Concerns about scaling to hundreds of agents

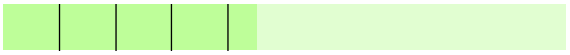


© Arango 2026 / arango.ai

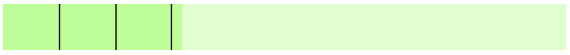
7 Nearly half would reconsider their AI stack

Stacks assembled during experimentation now face production demands. Many leaders question their earlier choices.

Second thoughts on the stack



45% would be unlikely to choose the same AI data stack if starting over



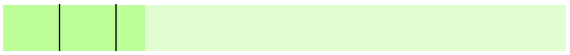
Only **31%** would likely choose it again.

These preferences do not establish replacement plans.

No one sees the full picture



53% lack confidence that anyone has a complete view of the technologies supporting their AI applications.



Only **25%** are confident someone does.

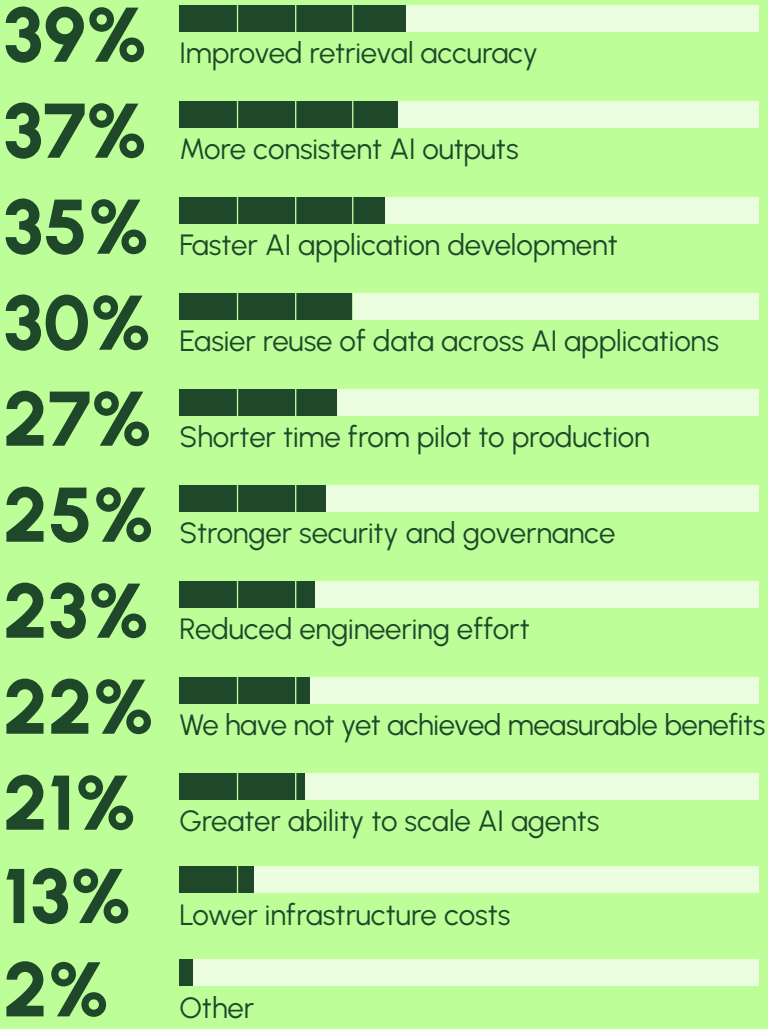
KEY INSIGHT: It is hard to fix a stack you cannot fully see. Visibility comes before architectural change.

WHAT IT MEANS FOR LEADERS: Before your next long-term commitment, map every system, dependency and owner, and identify opportunities for reuse. Use that inventory to make ownership explicit across teams. A shared view of dependencies can guide decisions about which components to retain, consolidate or make reusable before the next application adds further complexity.

8. Better context is already paying off

Organizations improving AI context report gains in quality and speed.

What stands in the way of shared enterprise context



Answer quality and development speed lead reported benefits. Reuse may extend gains to cost and agent scale.

The survey does not establish causation or the sequence in which benefits occur.



22% have not yet achieved measurable benefits.

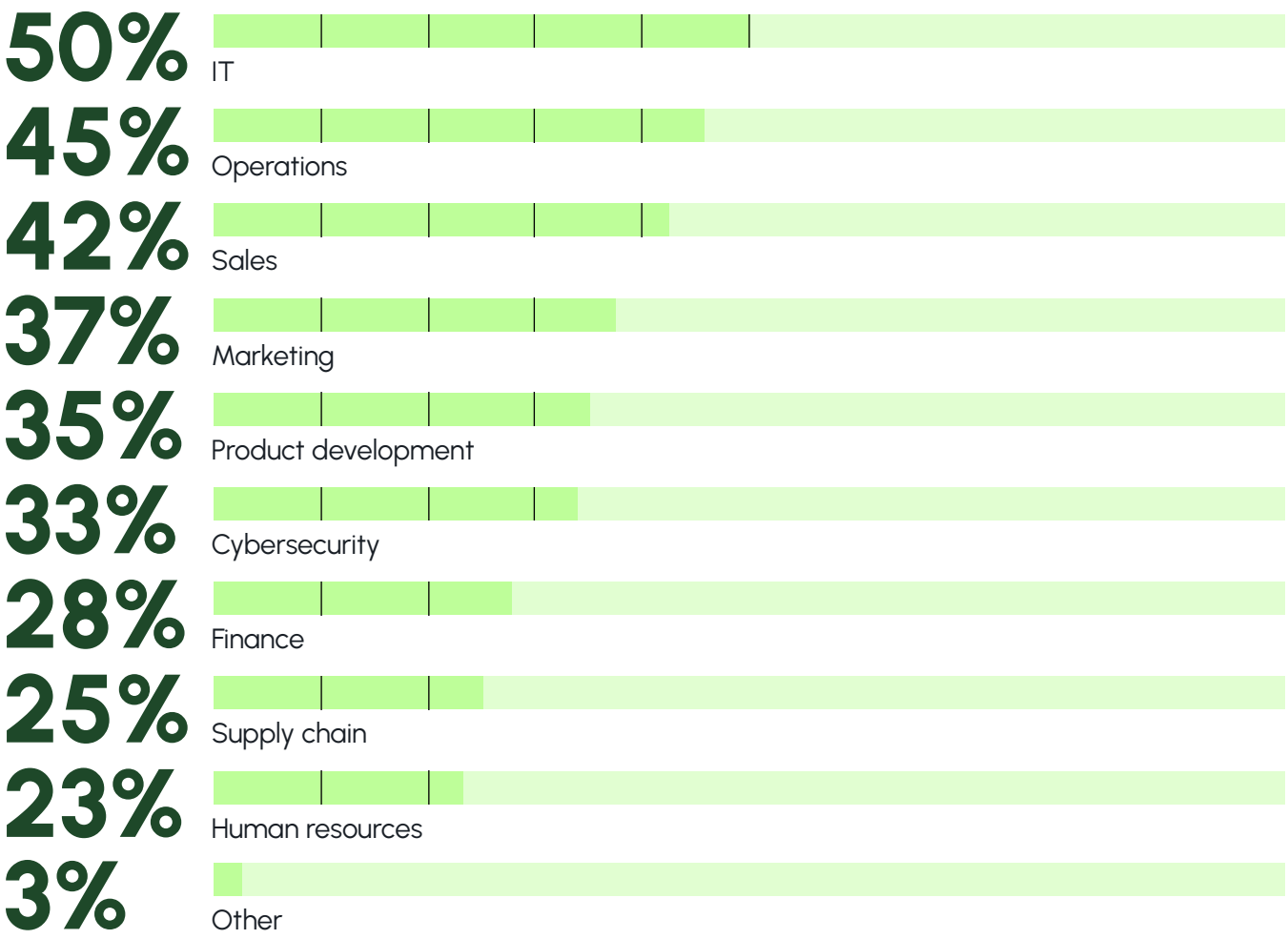
What it means for leaders: Set baselines for retrieval accuracy, output consistency and time from pilot to production. Measure gains to build the business case.

© Arango 2026 / arango.ai

9 The next agents will come from across the business

Leaders expect operations, sales and marketing to drive deployment alongside IT.

Functions expected to drive future deployment



When departments build context separately, they repeat work. Shared context lets them use the same trusted foundation.

Key insight: Shared context supports departmental innovation without rebuilding infrastructure.

What it means for leaders: Plan for the enterprise and provide self-service access to shared, governed context.



Connect with Arango to plan enterprise context around your next production AI use case.

Report Summary

Enterprise AI is progressing toward production, but the research reveals substantial concerns about whether today's infrastructure can support tomorrow's ambitions. Fragmented data systems, application-specific context layers, and growing engineering workloads are creating obstacles to scaling trusted AI. Although shared enterprise context infrastructure remains relatively uncommon, most leaders recognize its strategic importance and expect broader adoption. The expansion of AI agents adds urgency, raising concerns about operational complexity, governance, security, and technology debt. For business leaders, the findings suggest a practical direction: begin with valuable production use cases, develop reusable context capabilities, establish architectural visibility and ownership, and reduce the need to reconstruct business context for every new AI initiative. The organizations best positioned for the next phase of enterprise AI may be those that treat context as shared infrastructure rather than an application-specific afterthought.

https://arango.ai

