State of AI Data Connectivity Report: 2026 Outlook

Over 200 data and AI leaders say data infrastructure is the biggest barrier to AI success

The Top Line

The August 2025 MIT report, The GenAI Divide: State of AI in Business 2025,[1] made waves among business leaders and AI product owners largely due to its headline statistic: 95% of generative AI pilots at companies are failing. With the unprecedented scale of investment and the high expectations for enterprise applications of large language models (LLMs), both GenAI evangelists and skeptics were quick to weigh in on the disappointing outcomes of these early experiments.

While the accuracy of that specific statistic continues to be debated, the core issue it surfaces is not: a large share of companies are failing to realize meaningful ROI from their AI efforts. The more important question is, why?

We surveyed 200+ data and AI leaders, both from enterprises with internal AI adoption initiatives as well as software companies embedding AI copilots and agents into their products. And here's what we learned: enterprise AI is no longer limited by models. It's constrained by data infrastructure and enterprise context.

The strongest predictor of AI success in 2026 is the maturity of the underlying data infrastructure that delivers enterprise context to these models.

In fact, 60% of companies at the highest level of AI maturity also have the most mature data infrastructure. And the inverse is also true: 53% of companies with immature AI have immature data systems.

In this report, AI maturity refers to the extent to which an organization has operationalized AI, moving beyond experimentation to measurable business impact. Our framework considers dimensions such as model deployment, data integration maturity, governance, and ROI tracking. We categorize maturity in a five-stage progressive framework that draws from EY-Parthenon's AI maturity model: *experimenting, implementing, scaling, optimizing, and leading.*[2]

> "The paradox of AI readiness is that our data infrastructure becomes more powerful not through endless adaptability, but through intentional semantic boundaries that give LLMs the predictable contracts they need to orchestrate complex workflows. Without this deliberate architecture of constraints, we're left with systems that burn tokens on ambiguity rather than delivering value."
>
> **— Carlisia Campos, AI Software Engineer, Grokking Tech**

[1] "The GenAI Divide: State of AI in Business 2025", MIT NANDA, Aug. 18, 2025

[2] "How a Top-Down Holistic Strategy Can Maximize GenAI ROI", EY-Parthenon, June 18, 2024

As anyone using enterprise AI tools like ChatGPT, LangChain, or Agentforce can attest, it's no surprise that context plays a defining role in AI maturity. Large language models depend heavily on it for accurate, reliable, useful outputs. What *is* surprising is how few organizations are actually set up to deliver that context.

Other findings from the research highlight the specific challenges standing between intention and execution. Across both enterprises and software providers, we found:

Finding	Implication
71% of AI teams spend more than a quarter of their implementation time on data integration—including modeling data, implementing ETL pipelines, configuring connectors.	When significant resources are tied up in data integration, attention is pulled away from strategic product development and innovation.
46% of organizations require real-time access to six or more data sources for an average AI use case.	Each AI use case requires connecting to multiple systems, which adds architectural complexity and increases the burden on data teams.
AI-native software providers are 3x more likely to require more than 26 external data integrations in product, as compared to traditional providers (46% vs. 15%).	Modern software companies are architecting for scale from day one, exposing integration gaps in more traditional providers.
100% of organizations say real-time data is necessary for AI agents and customer service automation. While 80% of enterprises have begun implementing real-time integration, most are still in the early stages of scaling it effectively.	There is a significant real-time integration capability gap that could limit the adoption of AI agents and automation at scale.
All high-AI-maturity ("leading") enterprises have built centralized, semantically consistent data access: 80% of low-maturity ("experimenting") enterprises haven't even started.	Semantically consistent data access is not just a best practice, it's becoming an AI imperative. Software providers and enterprises that lack it will struggle to keep up.
58% of respondents prioritize structured data sources (organized, schema-based formats like databases and APIs) for AI features, while only 11% primarily rely on unstructured data (free-form content such as documents, chat logs, and media files).	There's lots of discussion about unstructured data, but structured data remains the core building block for most AI applications.
Only 9% of respondents rank AI model acquisition or development as their top investment priority, but 83% are implementing or planning a centralized, semantically consistent data access layer.	The market is prioritizing data infrastructure over model building, signaling that data access is the real bottleneck in AI progress.

The survey results point to a sobering truth: generative and agentic AI aren't bottlenecked by the capabilities of foundational AI models, but by access to connected, contextualized, controlled data. And the AI landscape is rife with data integration issues, from fragmented systems to a lack of connectors and real-time infrastructure.

That's the bad news. The good news? There are enterprises and software providers that are getting it right, and the survey surfaced the key initiatives, priorities, and investments behind their success. If you're an enterprise looking to self-assess your AI maturity or the current state of out-of-the-box agentic AI solutions, this report offers valuable insight. If you're a software provider aiming to benchmark yourself against industry leaders and better understand enterprise investment priorities, you'll also find practical guidance here.

The report is made up of two major parts:

1. **Enterprise AI adoption and data challenges:** A deep dive into how enterprise organizations are deploying AI and what infrastructural blockers are slowing progress.
2. **Product AI strategy among software providers:** An exploration of how product leaders are embedding AI into their platforms and why data integration remains a critical dependency.

Together, these sections form a comprehensive picture of how data connectivity, infrastructure maturity, and integration strategy dictate AI success in both enterprise and product contexts.

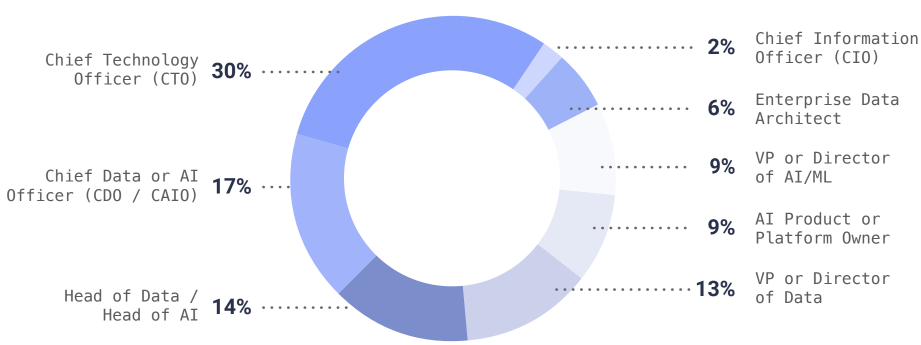
Table of Contents

Survey Methodology and Respondent Demographics

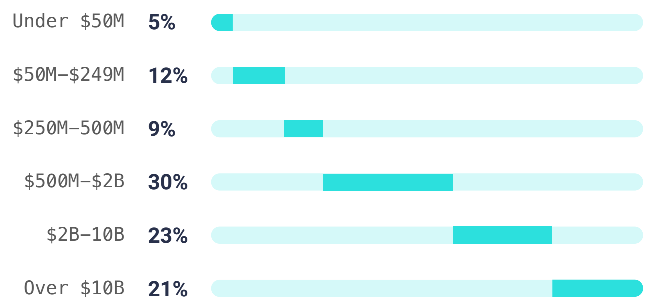
The insights in this report draw from two complementary surveys conducted in 2025; one capturing the perspective of enterprise AI implementation leaders, and the other from product leaders at software providers. Together, they offer a dual view of how organizations are adopting and operationalizing AI: from enterprises embedding AI into their operations, to software providers building AI directly into their products. Each survey aimed to uncover the current state of AI adoption, the infrastructure challenges shaping progress, and the investment priorities defining the next phase of AI maturity. Accordingly, Part I of the report focuses on enterprise AI adoption and the data infrastructure gap, while Part II examines the software provider perspective and the evolving strategies behind AI-powered product development.

Part I methodology: We used an independent research firm to blind survey 100 enterprise data and AI leaders, across industries and sizes ranging from startup to over $10B in annual recurring revenue.

Nearly half (49%) of the respondents were C-level executives responsible for technology, IT, data, and AI functions. Including the 22% VPs and directors who responded to the survey, the data set is strongly representative of enterprise leaders with decision-making authority and a mandate to drive organization-wide impact through the adoption of AI.

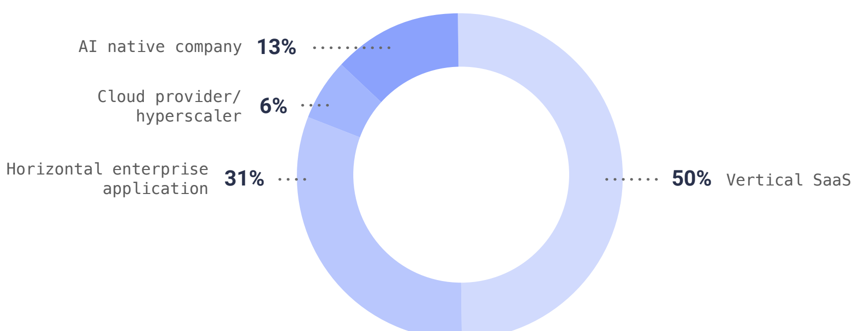


Seventy-four percent of respondents were from companies with more than $500M in annual revenue, while the remaining 26% belonged to mid-sized companies and startups. The dataset is thus skewed toward organizations that have bigger IT budgets and exposure to a wide swath of AI and data infrastructure approaches in the market.

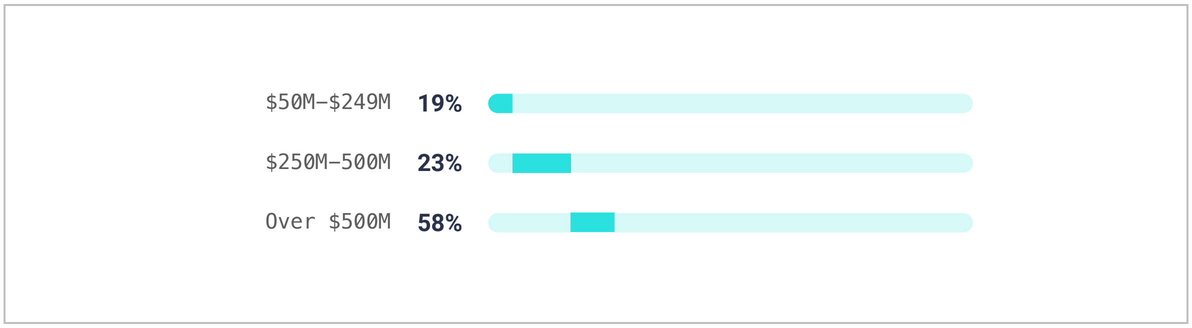


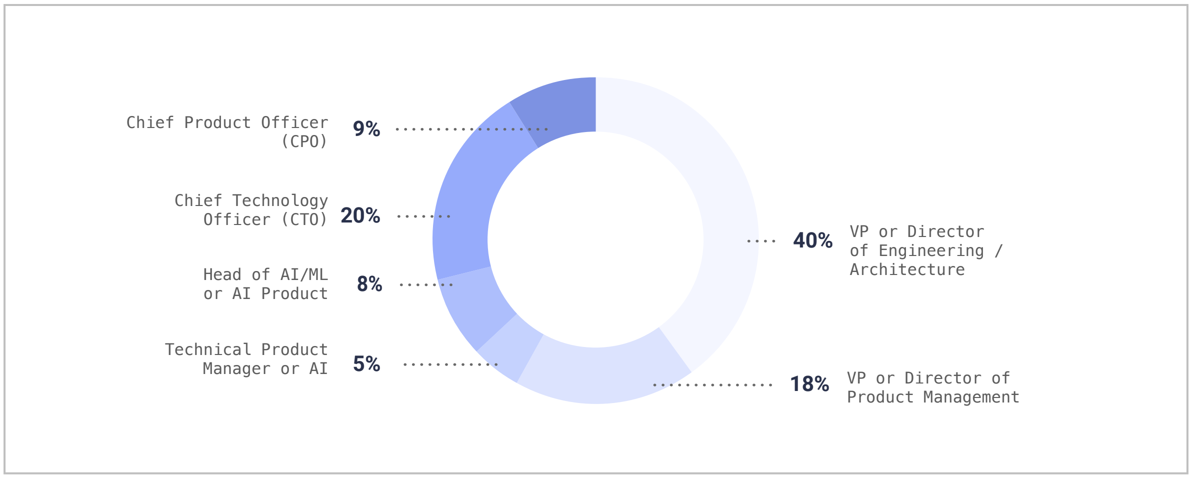
This respondent mix reflects a front-row view of how AI is being built and deployed today in the enterprise.

Part II methodology: This half of the report represents results from a blind survey conducted by an independent research firm of 100 product and engineering leaders from a mix of software companies, ranging from AI-native startups to some of the most established players in SaaS and enterprise platforms. This offers us a uniquely comprehensive view into how different product strategies intersect with AI readiness and integration approaches.



In this cohort, 58% of respondents are software providers reporting $500M or more in ARR. Titles include product leaders across functions: forty percent are VPs or directors of product, with significant representation from engineering, architecture, and AI leadership roles. Twenty-nine percent are C-level decision-makers (CTOs and CPOs), setting organization-wide priorities regarding AI implementation in product.





Definitions used in this report *(see glossary of terms on page 37 for additional definitions):*

Generative AI (GenAI)—AI-powered features built into products that help customers complete tasks by generating content, surfacing insights, or interacting with data, often using LLMs. This report focuses on two common GenAI applications:

- **AI Co-pilot**—An AI-powered assistant embedded in your product that helps users complete tasks by generating content, retrieving data, coding, or recommending next steps, but always requires human input to initiate or approve actions. Example: A copilot that summarizes recent customer activity and suggests follow-up actions, which the user reviews and approves.
- **Autonomous AI Agent**—Acts with minimal or no human prompting to complete tasks or achieve goals. These agents can reason, make decisions, and take action across systems or workflows on behalf of the user. Example: an AI agent that monitors pipeline activity, flags at-risk deals, and sends proactive alerts or messages.

Part I: Enterprise AI Adoption and the Data Infrastructure Gap

The findings below highlight key themes that emerged from our survey of enterprise leaders responsible for advancing AI adoption and maturity.

Key takeaways:

- **AI is already in production, not in pilot.** 78% of enterprises have moved beyond experimentation, embedding AI into operations, but only 17% are in advanced stages where ROI is measurable.
- **AI capabilities and model size are not the top blockers to adoption. Data and context are.** 73% of organizations cite data quality and integration as top blockers, and 71% spend over a quarter of AI project time just on data connectivity.
- **Scale and maturity go hand-in-hand.** Large enterprises with mature data infrastructure are pulling ahead, while 80% of firms under $50M in ARR remain stuck in early implementation.
- **Real-time, governed data is the new differentiator.** 60% rank governance and 42% rank real-time connectivity as top investment priorities, far surpassing investment in the AI models themselves (9%).
- **Fragmented tools demand unified integration.** 44% of organizations listed "lack of unified metadata and semantic context" among their top five current blockers to enterprise AI adoption, and 83% of organizations have built or are planning to build centralized, semantically consistent data access.

What follows is a deep dive into the priorities, roadblocks, and emerging trends shaping enterprise AI adoption, based on the survey results.

Enterprise AI isn't on the horizon: it's in production

Finding:

Beyond experimentation: 66% of companies are deploying GenAI and autonomous agents to augment human workflows.

Stuck in the middle: most enterprises are implementing and scaling AI, but very few are leading

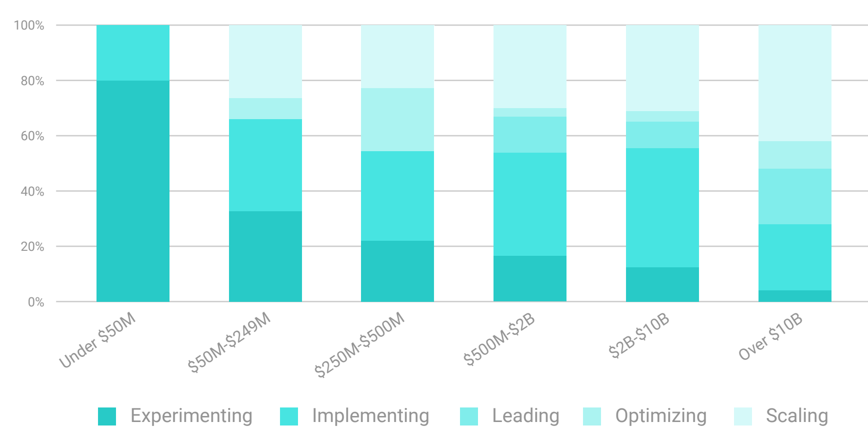
AI is not a future aspiration for most enterprises. It's here. In fact, 78% of enterprises are past the pilot phase, with AI use-cases already embedded in operations.

A majority of enterprises (68%) fall into the middle stages of AI maturity, between the "implementing" and "scaling" stages. However, only 17% are in advanced stages ("optimizing" or "leading") where ROI is measurable and AI is core to strategy.

Where would you place your organization on the AI maturity curve?

Stage	% of organizations
Experimenting (early pilots, proofs of concept, learning phase)	15%
Implementing (deploying initial production use cases, establishing governance)	37%
Scaling (expanding AI across multiple departments and use cases)	31%
Optimizing (AI integrated into core operations, measuring ROI and efficiency)	7%
Leading (AI drives competitive advantage and innovation strategy)	10%

The data also shows bigger companies are pulling ahead. Only 4.8% of enterprises over $10B in annual revenue are still in the early stage of experimenting with AI, while 80% of those under $50M in annual revenue remain stuck in early implementation.



Implication: Scale matters. Large enterprises have the data infrastructure and in-house talent to operationalize AI, while smaller firms are still laying the pipes to get pilots off the ground.

> "A year ago, we implemented AI assistants within all our call centers, fully in production. It is fully integrated with our backend data, so when a customer calls, it automatically recognizes their number, looks up the order, the delivery status, and answers the call, all before a human agent can even pick up the call, in real-time. The results were dramatic."
>
> **— SVP of Technology Portfolio, global retail brand**

Knowledge assistance and customer service automation are the most prevalent applications of enterprise AI

Early success stories focus on internal knowledge assistants and customer support automation. Code generation is close behind. These use cases thrive on access to both structured (databases, APIs, spreadsheets) and unstructured (images, emails, documents) data, but even more advanced capabilities (e.g., AI agents, decision support) are gaining traction.

Which use cases is your organization targeting with GenAI or agentic AI today or in the near future?

Use case	% of Orgs
Employee or agent co-pilot (e.g., internal knowledge assistants, agent augmentation)	79%
Customer support and service automation (e.g., virtual agents, chatbots, ticket deflection)	70%
AI-powered search or knowledge retrieval (e.g., RAG systems, semantic search)	61%
Code generation or augmentation (e.g., internal dev tools, LLM-driven refactoring)	60%
Intelligent document processing (e.g., summarization, extraction, classification)	58%
Marketing or content generation (e.g., campaign copy, image generation, personalization)	55%
Process or workflow automation (e.g., agent-triggered actions, RPA augmentation)	54%
Decision support or scenario analysis (e.g., contextual insights, what-if modeling)	52%
Predictive analytics for GTM, revenue, or customer retention	49%
Internal business intelligence enhancement (e.g., natural language dashboards)	47%
Predictive analytics for supply chain, logistics, or operations	37%
AI agent orchestration across systems (e.g., updating records, syncing workflows)	33%

AI chat assistants and agents are primarily deployed to augment human workflows, not replace them. The top use cases, employee/agent copilots (79%) and customer support automation (70%), signal that enterprises are focusing on human-in-the-loop augmentation. These use cases help knowledge workers operate more efficiently without handing over full control to AI.

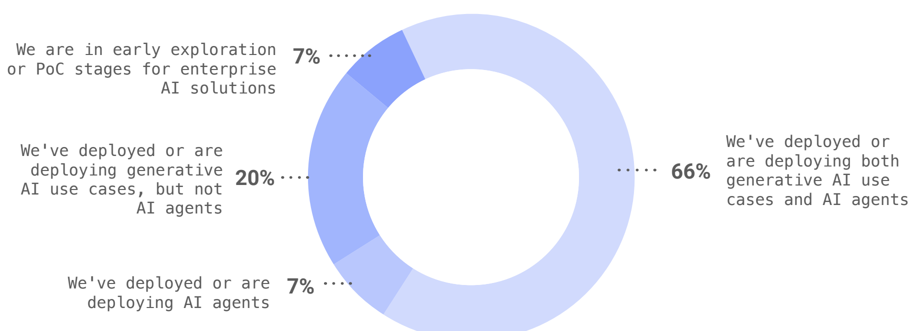
The impressive adoption of code generation (60%) and marketing/content creation (55%) shows that AI is now embedded in technical and creative workflows alike. These are productivity multipliers that are low risk but high impact, and are often early wins for AI adoption.

Implication: Enterprises are betting on AI to boost productivity, not replace people. The early focus on copilots, support automation, and code generation shows that adoption is centered on practical, human-in-the-loop use cases that deliver fast value with lower risk.

A majority of organizations have already deployed agentic AI systems

Generative and agentic AI adoption is prevalent, with a clear shift from experimentation to deployment, especially around agent-based use cases.

What best describes your organization's current internal engagement with generative and agentic AI?



Implication: AI agent adoption isn't theoretical. It's already happening at scale, signaling a fast-moving shift toward more autonomous, workflow-integrated AI.

> "Data is absolutely the lifeblood of agents actually being helpful for your enterprise. And so, having the right connections, the right fidelity, the right security, the right compliance around your data is all critical."
>
> **— Philip Stephens, Senior Staff Software Engineer, Google**

AI tool sprawl is fragmenting context at a time when context matters most

While 76% of organizations leverage foundational models in enterprise LLM platforms, the AI technology stack is not centralized. Enterprises also report significant use of BI-native copilots, agent platforms, and customer service AI.

Which AI applications or platforms are most important to your organization's use cases?

Platform category	% of surveyed organizations
Enterprise LLM platforms (OpenAI, Claude, Gemini)	76%
Enterprise data platforms (Snowflake, Databricks, etc.)	65%
Business intelligence AI (Microsoft Copilot, Tableau AI, etc.)	54%
Code generation AI (GitHub Copilot, Cursor, etc.)	48%
Enterprise AI agents (Salesforce Agentforce, Copilot Studio)	43%
Cloud AI services (Vertex, Bedrock, Azure AI, etc.)	34%
Customer service AI (Zendesk AI, ServiceNow AI)	31%
Custom AI applications (Built In-House)	29%
AI development frameworks (LangChain, LlamaIndex, etc.)	28%
Open source AI models (Llama, Mistral, etc.)	20%
Industry-specific AI solutions	14%

Implication: This sprawl creates integration complexity and context fragmentation that must be addressed through centralized, tool-agnostic semantics and integration.

> "Most enterprises, especially older companies with lots of history, have disparate systems that are cobbled together. Your ability to get value out of these data assets is largely a function of your data integration capability."
>
> **— Chief Data and Analytics Officer, Fortune 100 manufacturer**

The current state of data infrastructure powering AI

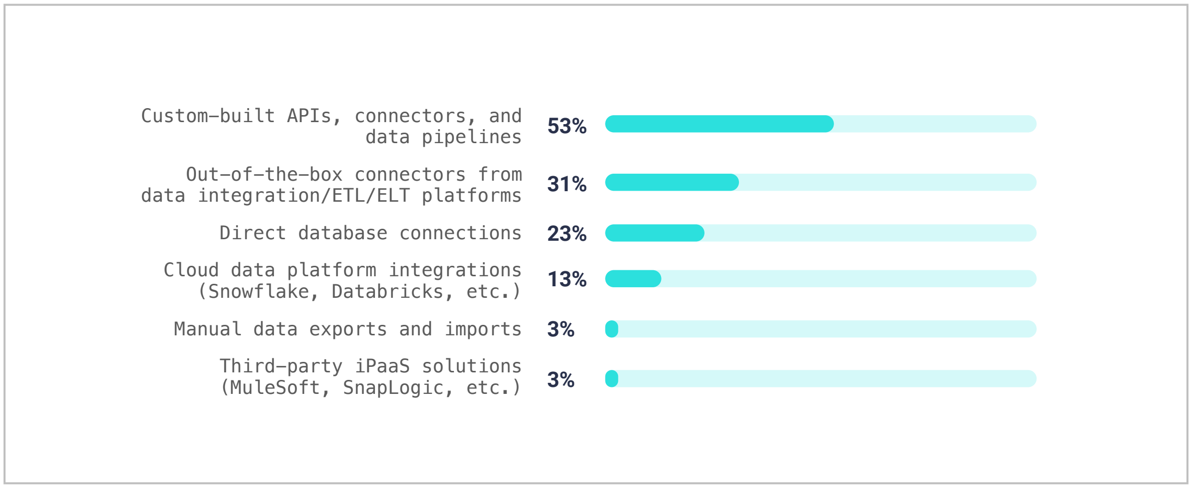
Finding:

Only 6% of enterprises are satisfied with their current data infrastructure for AI.

Enterprise AI leaders are largely unsatisfied with current integration architecture

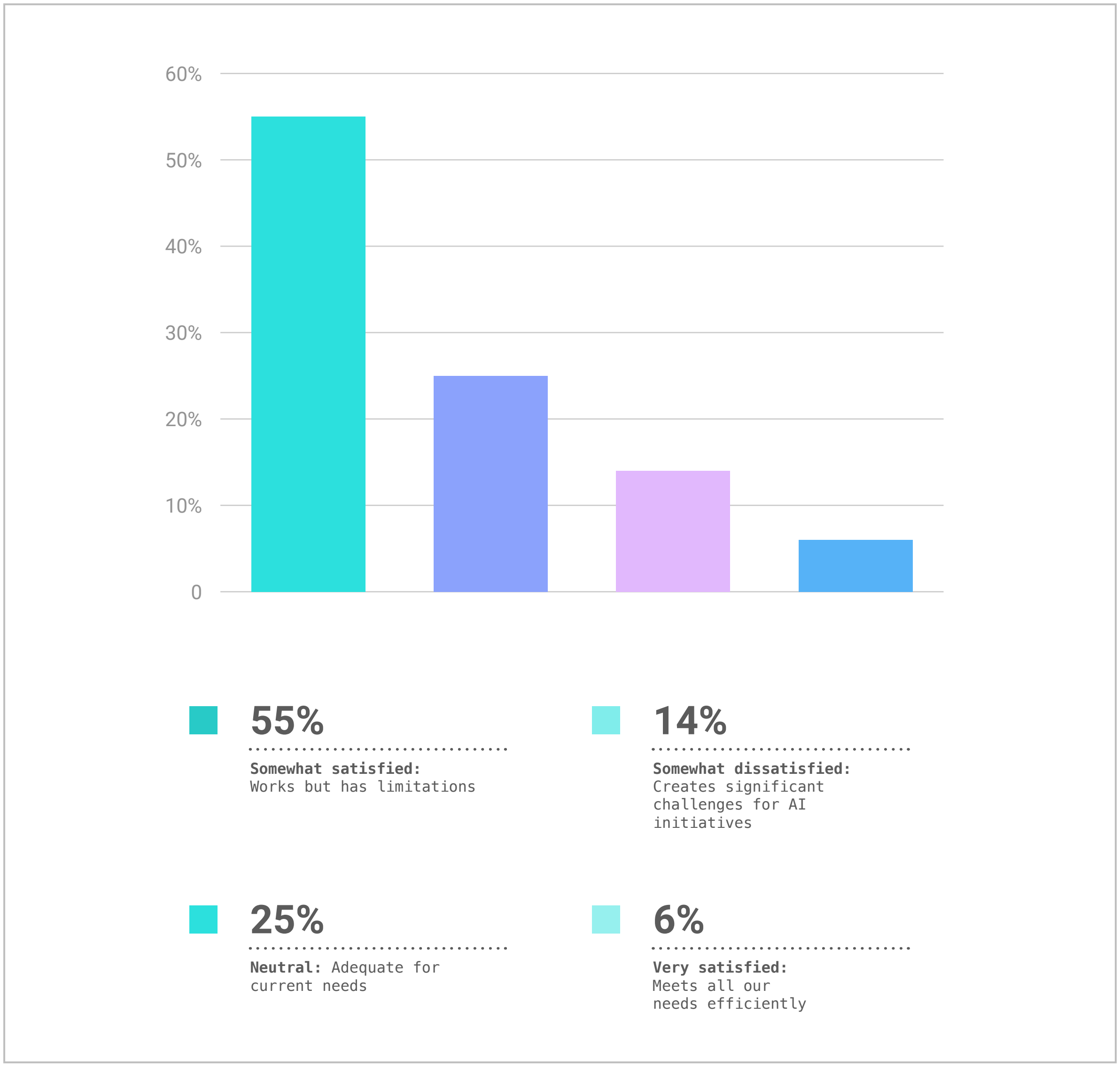
Most enterprises still rely on a mix of fragile or manual approaches, with 53% relying on custom-built APIs, connectors, and data pipelines to deliver enterprise data context to AI models.

How does your organization currently connect AI systems to enterprise data sources?



Overall, organizations report a high degree of pain and dissatisfaction with their current integration strategy and infrastructure. Only 6% reported they were "very satisfied" with their integration strategy. Fourteen percent reported their integration strategy creates significant challenges for AI initiatives.

How satisfied are you with your current data connectivity approach for AI initiatives, including ingestion of data from source systems, context injection for GenAI models, real-time data integration, etc.?



> "AI technology has advanced faster than organizational data capabilities, creating a critical bottleneck for AI adoption. While sophisticated AI models are readily available, most companies struggle with poor data quality, fragmented systems, and inadequate data preparation processes. Ultimately, AI success depends more on having high-quality, well-prepared data than on having the most advanced models."
>
> **— Harshit Kohli, Sr. Technical Account Manager, AWS**

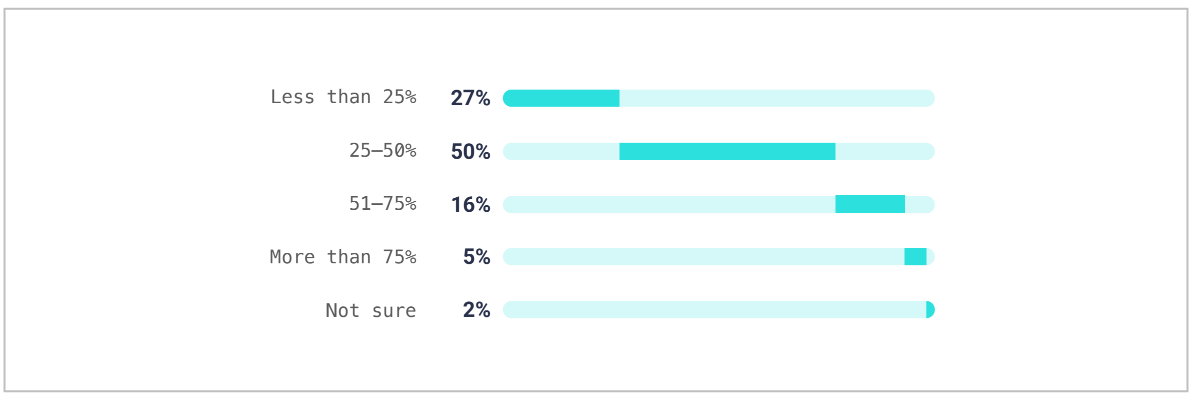
Implication: The cost of bespoke integration is high; not just in dollars, but in delays and fragility. The MIT Report on Enterprise AI adoption indicates that custom-built solutions result in a significantly higher rate of failure for enterprise AI initiatives. The overall dissatisfaction expressed by enterprise data and AI leaders is likely reflective of underlying maintenance overhead, delays, and technical limitations that accompany the lack of a systematic approach to connectivity and integration.

When data connectivity becomes the AI bottleneck

> "Historically, about 80% of the effort in data operations was manual, which meant humans had to manually fix pipelines and handle intermittent cloud outages. That's where most of our time went, just keeping things running. It's fragile and reactive, and when you're feeding AI systems, that brittleness becomes a real blocker."
>
> **— Global data and analytics executive, Fortune 50 CPG enterprise**

Organizations were asked to estimate what percentage of AI implementation was spent on data connectivity and integration challenges. The responses indicate that 21% of AI implementation teams spend a majority of their AI implementation timeline on data connectivity and integration. And 71% of AI implementation teams spend over a quarter of their project timeline on data connectivity and integration.

What percentage of your AI development time is currently spent on data connectivity and integration challenges?



When asked to rank the top challenges to enterprise AI adoption, respondents strongly indicated that data governance and connectivity are the leading barriers to enterprise AI adoption.

Data governance challenges are the top priority blocker:

- 83% of respondents cited data quality or security among their top five challenges.
  - 73% ranked data quality as a top blocker.
  - 66% ranked security and privacy as a top blocker.

Connectivity gaps are exacerbating the problem:

- 50% of respondents ranked data connectivity gaps among their top five challenges, and 29% listed them as the #1 blocker to enterprise AI success.

Together, these findings show that data governance and connectivity define the readiness gap for enterprise AI.

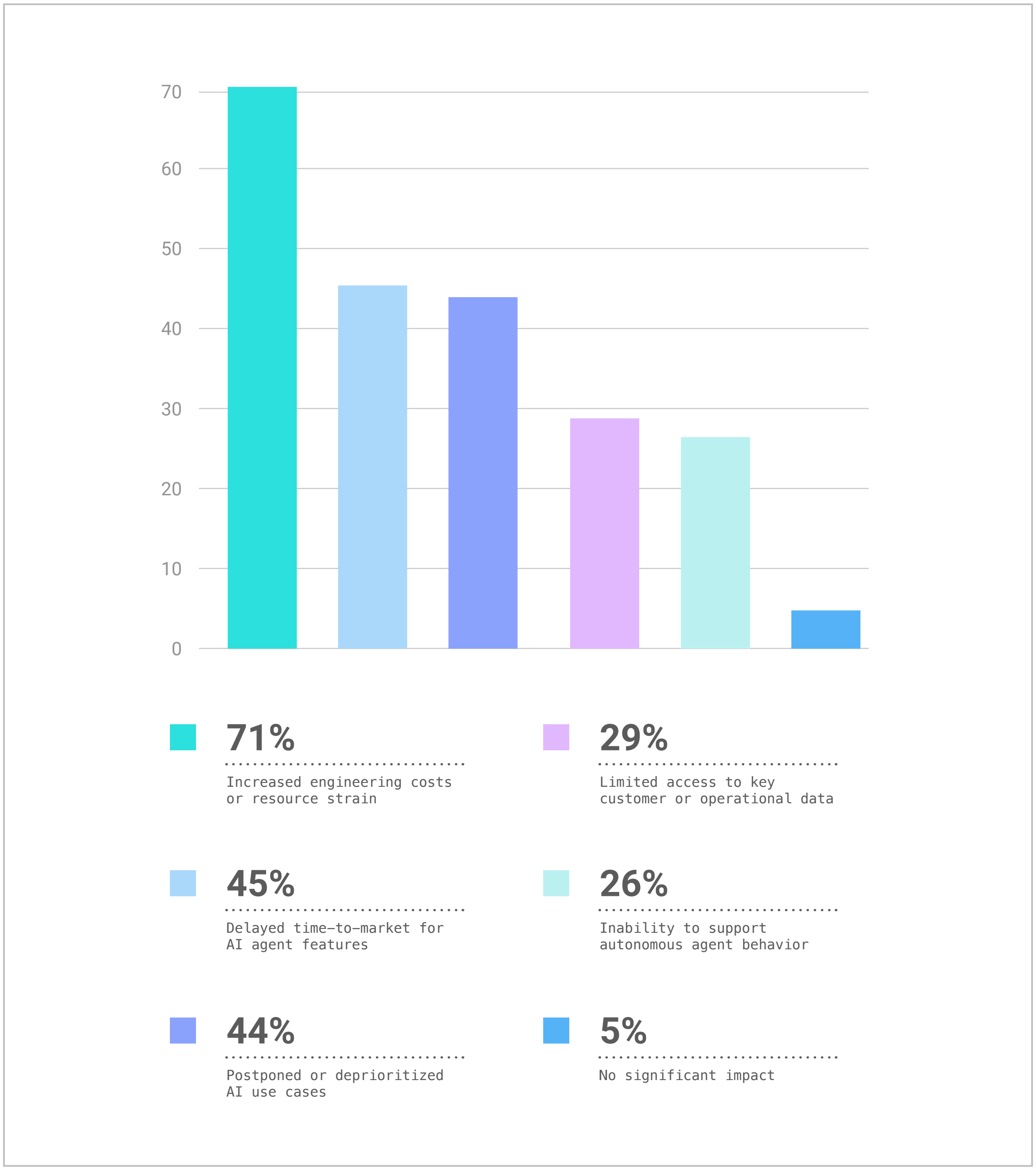
> "With global replication and legacy systems, data quality becomes a huge issue. AI models can be poisoned by even a small sample of bad data. Executives make decisions on dashboards, and if the data's wrong, the decision's wrong."
>
> **— SVP of Technology Portfolio, global retail brand**

What are your biggest challenges with connecting AI systems to enterprise data sources?

Challenge	% of orgs ranking problem in top 5
Data quality and consistency issues across sources	73%
Security and compliance concerns with data access	66%
Lack of pre-built connectors for specific enterprise applications	50%
High maintenance overhead for custom integrations	43%
Limited real-time data access capabilities	36%

Respondents were also asked about the impacts of integration challenges on enterprise AI initiatives. Only 5% reported no significant impact, with most organizations citing more than one detrimental impact from data integration challenges. Seventy-one percent reported increased cost and resource strain, and 45% reported delayed time to market.

Which of the following impacts have data integration challenges had on your AI agent initiatives?



Implication: The disproportionate time investment in data integration, as well as the significant detrimental impact of integration challenges on AI roadmaps, serve to reinforce the notion that connectivity and integration are the biggest blockers to enterprise AI initiatives. They also support the idea that investing in robust, automated data integration can significantly accelerate time-to-value/ market and the success rate of enterprise AI adoption initiatives.

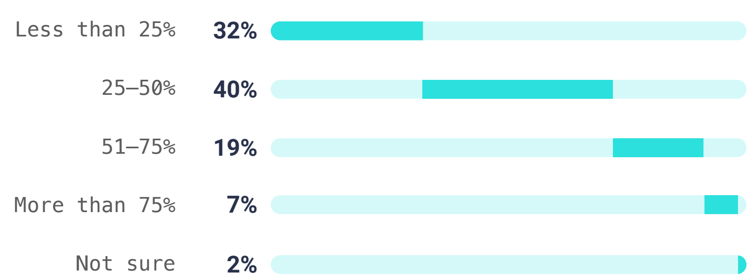
“Organizations at the AI frontier are differentiating themselves by scaling new patterns of work enabled by AI. To do this, they first need to solve for the triple constraints of data—data delivery through semantic integration to bring interoperability and context for AI agents, data quality to ensure reliable insights, and security to mitigate the real risks of an AI-first environment. Through this trustworthy and efficient foundation for AI, enterprises can outperform peers in productivity, innovation, and speed to value.”

— Anoop Tripathi, AI Business Solutions Scale Leader, Microsoft

Real-time integration is a maturity marker

Respondents were asked what percentage of their AI use cases require access to real-time or near-real-time data. Sixty-six percent of enterprises indicated that over a quarter of use cases require live data access, highlighting the importance of this capability over more static models of context engineering such as retrieval augmented generation (RAG).

For your AI use cases, what percentage requires real-time or near real-time data access?



When asked about their integration mix (between batch and real-time patterns), 80% of respondents indicated they support real-time or hybrid (real-time and batch) integration capabilities.

Does your AI strategy include integration with your enterprise data systems?

Realtime vs. batch integration mix	% of respondents
Yes – we support a combination of batch and real-time integration	33%
Yes – we support exclusively batch integration	20%
Yes – we support exclusively real-time integration	47%

When overlaid against respondents' AI maturity, real-time integration emerged as a predictor of an enterprise's success with AI. All the most AI-mature ("Leading") enterprises support real-time integration capabilities, whereas only 40% of those in the lowest tier of AI maturity supported real-time integration. Of the enterprises leveraging AI for customer service, decision automation, or agentic use cases, 100% of respondents support real-time data access capabilities, suggesting that these use cases are particularly driven by access to live data.

Implication: Organizations that support real-time integration are far more likely to be in advanced stages of AI maturity. Real-time infrastructure isn't just a technical preference: it's a strategic necessity for agentic AI use cases.

Beyond models: the architecture and capabilities of AI readiness

Finding:

Data infrastructure determines AI success. The most AI-advanced enterprises also have the most mature integration infrastructure.

AI maturity correlates with integration maturity

A full 60% of the most AI-mature enterprises also report mature data infrastructure. Conversely, 53% of the lowest-maturity AI orgs have the weakest integration capabilities.

This strongly suggests that enterprises can't scale AI without first solving data integration.

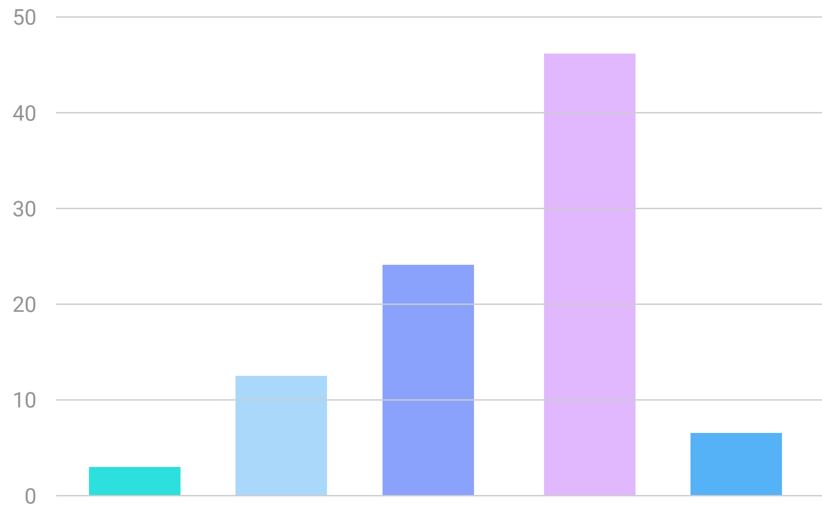
> "Unless you invest in your integration layer up front, you'll have unpredictable challenges surfacing data to AI models or making the right inferences. We learned that if you try to scale AI without fixing the plumbing first, you just amplify bad data and technical debt."
>
> **— Chief Data and Analytics Officer, Fortune 500 advisory firm**

Centralized, semantically enriched data access is a prerequisite for scalable AI

> "Neuro-symbolic AI thrives on data that carries meaning—not just values. When real-time, context-rich data connects to the symbols of how a business operates, AI can move from predicting outcomes to explaining them. That's how enterprises unlock true decision intelligence from their data."
>
> **— Tooba Durraze, Founder and CEO, Amoeba AI**

Eighty-three percent of organizations have built, or plan to build, a centralized data access layer. The overwhelming majority of enterprises view this investment as foundational to scalable AI, signaling a broad shift away from fragmented, ad hoc data connectivity toward unified, governed access across systems.

Where is your organization today in building a centralized data access layer that delivers contextualized data from enterprise systems to AI models or agents?



3%
We have not yet integrated enterprise system data into AI models

14%
Federated: We rely on point-to-point integrations or pipelines without a central access layer

24%
Planned: We've identified the need and plan to build one in the next 12–18 months

46%
In development: We're actively building one as part of our AI architecture

13%
Fully deployed: We've built a functioning data access layer that provides unified, contextualized data across key systems for AI agents

Forty-four percent of organizations listed “Lack of unified metadata and semantic context” among their top five current blockers to enterprise AI adoption.

Implication: Context is no longer optional. Enterprises are prioritizing centralized, semantic data access because fragmented pipelines can’t deliver the context AI needs to scale reliably.

“With agentic AI, agents must capture data points across CRM, ERP and other systems, but AI agents can’t “understand” the data natively. So there has to be a metadata layer: data must be integrated from silos so agents understand context, fetch the right data, and enable actions.”

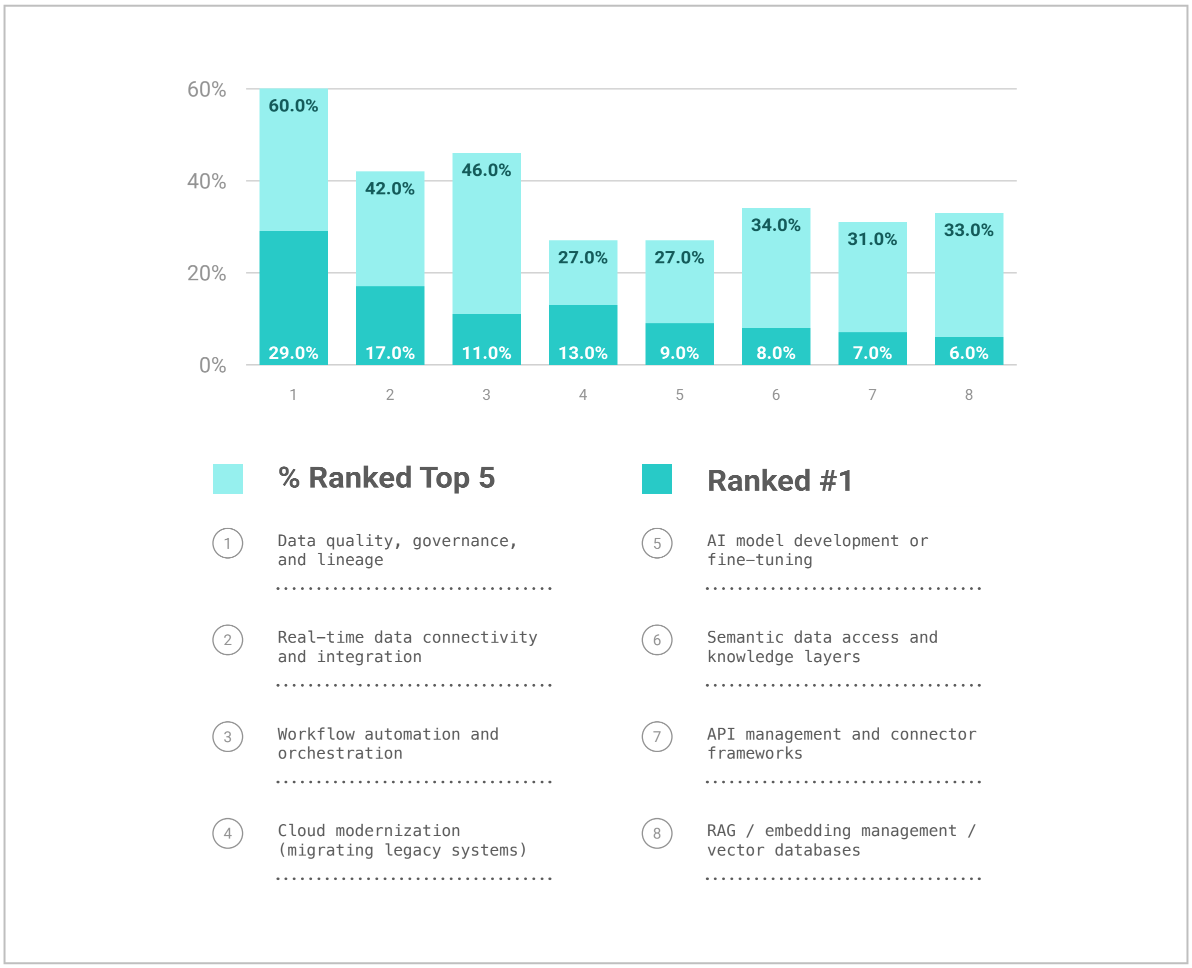
— Chief Data and Analytics Officer, Fortune 500 advisory firm

Top investment areas for AI success

Sixty percent put data governance, quality, and lineage in their top five investment priorities for AI readiness, more than double those prioritizing model development. Forty-two percent of organizations rank real-time connectivity among their top five, and 17% call it their top priority. Only 9% rank investment in AI models as their top priority.

Thirty-four percent of respondents ranked semantic intelligence in their top five investment priorities.

Where are you prioritizing investments to support AI?



Implication: The emergence of investment in semantic intelligence indicates leaders are realizing that data readiness for AI hinges on semantic context. The future of AI belongs to organizations that can connect, govern, and understand their data at scale.

Part II: The Software Provider Lens on AI Product Strategy

The second half of the report explores how software providers are embedding AI into their product roadmaps and architectures. While AI is now table stakes for product strategy—73% say in-product AI is a mission-critical organizational imperative—the findings reveal a striking gap between ambition and infrastructure. Across the board, software providers cited data fragmentation, schema inconsistency, and integration complexity as their top blockers to building and scaling AI features.

Key takeaways from this section:

- **AI features are already in production** for 77% of surveyed software providers, but only a small minority (9%) have reached autonomous agent capabilities.
- **Only 9% of respondents are "very confident"** their integration strategy can support AI feature development.
- **Integration pain is acute.** Data infrastructure and integration challenges were cited as the top blocker holding back AI initiatives by 32% of the surveyed software providers.
- **AI-native companies are building for higher integration complexity,** requiring 3x as many external data sources as traditional software vendors. This suggests that AI features are inherently integration intensive.
- **Real-time data access and semantic intelligence** are emerging as critical capabilities for in-product AI, with 62% of software providers implementing or having already implemented semantic standardization.

These insights confirm that data integration has become the new competitive frontier for software vendors in the GenAI era. The following pages unpack what's holding them back, and what it takes to seize the opportunities in front of them.

> "We're past the cool demo stage. Customers expect us to use GenAI to solve real problems end-to-end, not just add a shiny feature."
>
> **— VP of Product Management, leading cloud-native SaaS analytics provider**

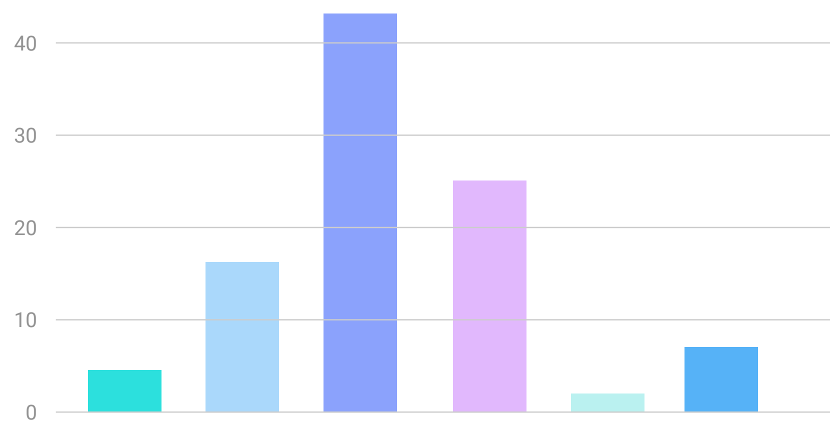
AI features are becoming table stakes for product leaders

Finding:

Most product teams are building AI features, but very few have deployed autonomous agents as a feature in-product.

Seventy-three percent of software providers say that implementing AI features is mission-critical to their long-term roadmap. That urgency is reflected in deployment: 77% already have AI copilots or agents live in their product, and 43% have deployed both. But while adoption is broad, full autonomy remains rare. Among those embedding AI agents, only 7% are scaling fully autonomous AI agents as defined in this report, signaling that most teams are still taking a measured, human-in-the-loop approach.

Where would you place your product or platform on the maturity scale for AI features?



5%

Exploratory: We're researching copilots but haven't implemented anything yet.

18%

Emerging: We've started embedding copilots into our product, but we're still defining use cases and architecture.

43%

Operational (human-in-the-loop): Copilots are live in select workflows and assist users through prompts or suggestions (e.g., summarization, guided insights, copilots).

25%

Adaptive (partially autonomous): AI agents can take limited actions such as updating records or triggering internal workflows, typically using internal data.

2%

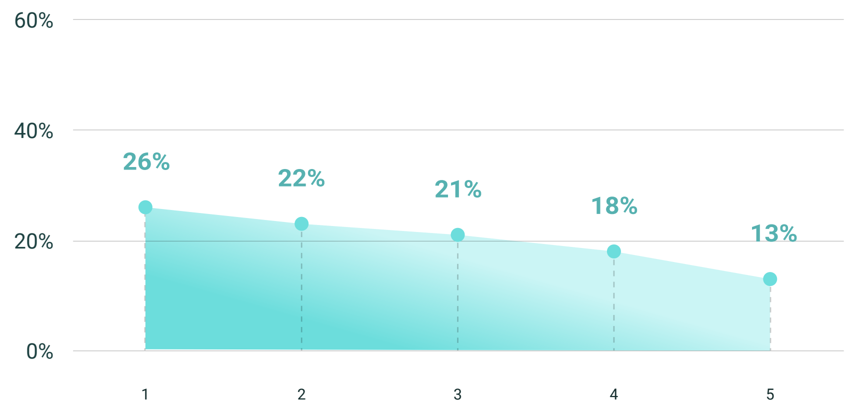
Autonomous: AI agents operate with meaningful autonomy, acting on behalf of users across systems, including third-party tools and dynamic, contextual data.

7%

Foundational: Autonomous AI agents are live and core to our long-term roadmap and product vision. We are actively scaling autonomous capabilities across the platform.

Thirty-nine percent of software providers have implemented insight-generation and knowledge-retrieval use cases, which do not involve action taking or workflow management. On the other hand, 61% are prioritizing workflow management and action taking.

If you’ve deployed—or are planning to deploy—AI features in your product, which best describes their most advanced use case?



1. **Limited action-taking (human-in-the-loop):** Agents can trigger system actions with user oversight (e.g., update CRM records, open tickets, send alerts).
2. **Workflow support (human-in-the-loop):** Copilots assist users in completing tasks but don’t act independently (e.g., guided form-fill, copilot UX, data lookups).
3. **Knowledge retrieval:** Copilots retrieve relevant documentation or contextual content to support users (e.g., onboarding guidance, help docs).
4. **Insight generation:** Copilots provide summaries, insights, or recommendations based on data (e.g., conversational analytics, summarizing dashboards, surfacing KPIs).
5. **Fully autonomous agents:** Agents can reason, decide, and take actions across systems—including read/write access and dynamic context—without step-by-step prompts.

Implication: The data shows a bias toward the implementation of AI features for action as opposed to synthesis and retrieval of information. This explains the rising importance of real-time integration models, as batch integration is insufficient for writing data back to operational systems.

Data fragmentation and integration is the biggest limiting factor for AI feature development

Finding:

Integration challenges have forced 55% of software providers to rethink data architectures and delay AI launches.

Despite recognizing AI's potential, most software products still lack the data infrastructure to deliver on it for customers. Integration strategies remain fragile and incomplete, a clear signal that even technically mature software providers are wrestling with foundational AI data integration in product.

Thirty-two percent of product and engineering leaders rated data quality and integration issues—including with internal data, customer data, and third-party APIs—as the biggest factor blocking the development of AI features. Tellingly, 84% of software providers indicated that data quality and integration are a top-five blocker to AI implementation in product.

What are the biggest factors limiting the development of AI features in your product(s)?

Limiting factor	% ranked as #1	% ranked in top 5
Poor data quality and integration	32%	84%
Defining a clear use case for AI copilots or autonomous AI agents	26%	56%
Customer concerns about the risk of automation	13%	58%
Security or governance	13%	59%
Lack of internal expertise or resourcing in our customer base to manage agents once deployed	11%	45%
Organizational resistance to AI	3%	13%
Confusion about what AI copilots and autonomous AI agents are	2%	17%

When asked to name their biggest technical challenges, product leaders' answers reinforced the same story: the barriers primarily relate to data quality and integration. Schema inconsistency, API integration complexity and lack of standards-based connectivity remain the toughest obstacles to scaling AI features.

What are the biggest technical challenges limiting your AI initiatives?

Biggest challenge to building AI features	% ranked #1	% ranked in top 3
Poor data quality or inconsistent schema definitions	35%	60%
Difficulty integrating with external SaaS platforms and APIs	18%	41%
Lack of support for standards (MCP, A2A)	14%	30%
Lack of live access to source databases (e.g., direct SQL access)	14%	30%
Conflicting data definitions (e.g., what counts as a "customer")	10%	36%
Limited observability into data lineage or source reliability	5%	30%
Lack of access to real-time streaming data (sensors, IoT)	4%	15%

When these challenges are not systematically resolved, they don't just slow development: they reshape entire product roadmaps. The data shows that integration issues have forced teams to re-engineer their architectures, delay releases, and even pause AI initiatives altogether.

Which of the following impacts have data integration challenges had on your copilot and autonomous AI agent initiatives?

Impact	% ranked #1	% ranked in top 3
Forced us to rethink our data integration architecture	33%	55%
Delayed time-to-market for AI features	28%	51%
Increased engineering costs or resource strain	18%	51%
Limited access to data that's critical to making agents or copilots work	9%	28%
Postponed or deprioritized AI use cases	3%	28%
No significant impact	9%	—

In short, when integration is not prioritized, it can lead to unpredictable downstream impacts that significantly hinder AI feature development and delivery. The data paints a stark picture: every software company at the highest stage of AI maturity has already established a centralized, semantically consistent access layer for both internal and customer data. In contrast, 80% of organizations at the lowest stage have yet to implement a central integration layer, though most plan to.

“

The key unlock for enterprise AI success is connectivity to internal tools and relevant information or data. This report highlights that without reliable connectivity and data quality, enterprise AI adoption won’t result in efficiency improvements, as humans will remain the bottleneck of relevant information. Data quality is just as important as data connectivity for AI, otherwise human employees will be left mediating what is the right information to input or share.

MCP is a big unlock here, and it’s not surprising to see so many enterprises actively exploring its adoption. That said, any connection between AI and internal tools creates inherent risks. Securing enterprise deployments of agents and connected language models is a huge area of risk and an opportunity to build meaningful safeguards and security.”

— Tobin South, Head of AI Agents and MCP, WorkOS

Implication: Enterprises and software providers alike face the same uphill battle: building AI features that depend on secure and real-time access to end-customer data spread across dozens of external platforms and APIs.

> "The biggest barrier is creating integrations. It's the #1 impediment to building product—especially for agentic experiences that must transcend multiple systems."
>
> **— Chief Technology Officer, high-growth vertical SaaS platform**

Most AI use cases need multiple integrations to customer data

Finding:

AI-native firms win on integration, and they're leaving legacy software behind.

The central importance of integration in AI use cases is borne out by the fact that 74% of product leaders need multiple external data integrations in product to support AI use cases, with 18% requiring more than 25 customer data integrations. Not surprisingly, there's a significant difference when comparing the average software provider to AI-native companies (comprised of startups and mid-sized organizations).

Number of Integrations	Software provider population	AI-native companies	Delta
51–100+	10%	**15%**	+5%
26–50	8%	**31%**	**+23%**
11–25	11%	**23%**	+12%
2–10	**45%**	15%	**−30%**
Only internal data/1 external	14%	0%	−14%
Don't yet know	13%	15%	+2%

While 45% of the general software provider population is working with a modest 2-10 external integrations, only 15% of AI-native companies are operating at that low integration scale. Instead, a majority (69%) of AI-native companies require 11+ integrations, with over 46% requiring 26 or more.

The significant delta of AI-native companies needing 26-50 integrations versus the general software provider population (31% vs. 8%) suggests that AI-native products are committed to deep integration into varied customer ecosystems. It also supports the notion that AI features and use cases are inherently more integration intensive, which has resulted in AI-native product leaders prioritizing integrations in a way traditional software providers haven't. Yet.

> "At some point it's all about the data and data acquisition…that's the barrier to scaling AI across the platform."
>
> **— VP of Product Management, leading cloud-native SaaS analytics provider**

Digging deeper into the specifics of these integrations, we asked product leaders which customer data sources are most important for the implementation of AI features.

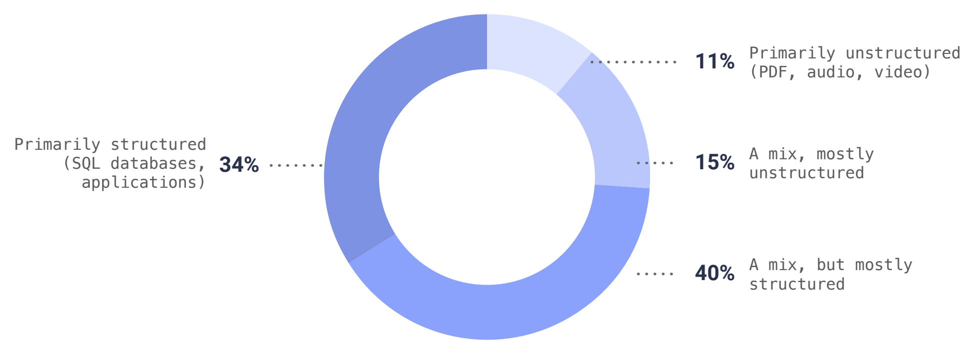
What are the most important customer data sources for your copilot or autonomous AI agent initiatives?

Customer data source type	Avg. rank* (lower #=more important)	% ranked in top 3
Databases or data warehouses (e.g., PostgreSQL, Snowflake, Redshift)	2.06	69%
CRM data (e.g., Salesforce, HubSpot)	3.17	33%
Internal application logs or telemetry data	3.50	26%
Customer support platform data (e.g., Zendesk, ServiceNow)	4.10	22%
Unstructured documents: PDFs (e.g., contracts, regulatory docs)	4.34	23%
Streaming data (e.g., IoT, Kafka, social media)	5.00	19%
File storage systems (e.g., SharePoint, Box, Google Drive)	5.03	16%
Collaboration tools(e.g., Slack, Microsoft Teams)	5.13	14%
ERP or financial platform data (e.g., NetSuite, SAP)	5.25	13%
Unstructured product manuals or customer-facing documentation	5.25	7%
External APIs or commercial data services (e.g., Clearbit, Dun & Bradstreet)	5.66	11%
Partner platforms and B2B integrations	6.35	8%
Unstructured Web content (internal or external)	6.79	6%
Marketing automation platform data (e.g., Marketo, Pardot)	7.56	5%
Unstructured voice data (e.g., notes, customer conversations, sales conversations, etc.)	8.26	3%
Unstructured video (e.g., for equipment diagnosis, security monitoring, etc.)	9.56	2%

*Ranking is on a scale from 1-10, with lower rankings denoting higher priority.

In addition, we asked respondents about the importance of structured data (e.g., relational databases, APIs, and spreadsheets) relative to unstructured data (e.g., documents, photos, and emails) for AI features embedded in product. Fifty-eight percent indicated that structured data sources are the main requirement for AI features, with only 11% of software providers primarily reliant on unstructured data sources.

For your most advanced GenAI initiatives (copilot or autonomous AI agents), the required data is:



Implication: Combined, the data shows that while unstructured data integrations grow in importance with respect to in-product integrations, structured data still remains dominant. Despite growing momentum around unstructured data, software providers are still focused on what AI actually needs to work in production: structured, relational data.

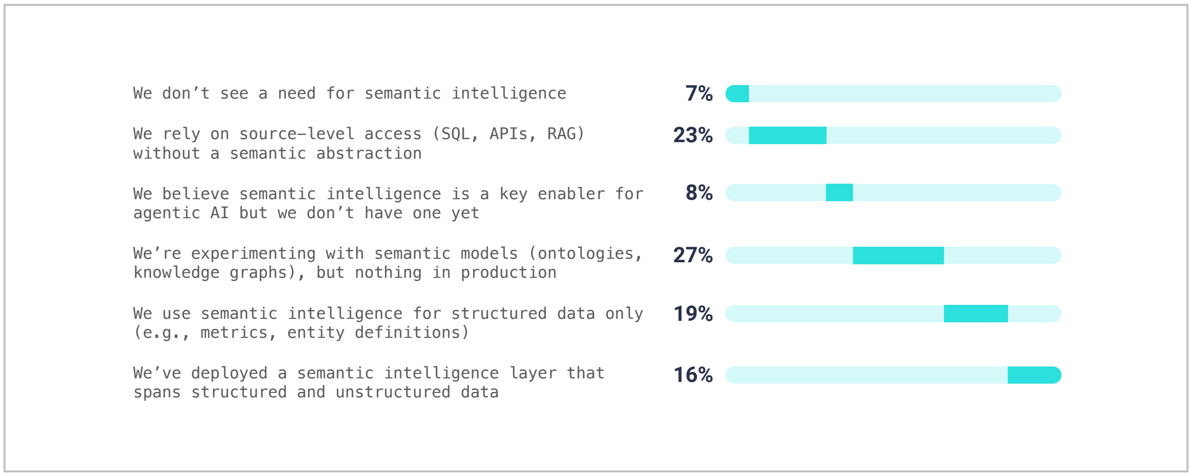
Semantic standardization and real-time integration demands from enterprises are shaping product roadmaps

Finding:

Over 75% of product leaders say real-time semantics is the key to building AI.

Just as in the enterprise segment, software providers are shifting toward building system-level semantic intelligence: the capability to understand the business context of data within each source application and how that context relates to data across other systems.

How is your organization handling semantic intelligence for copilots or autonomous AI agents (i.e., defining relationships, meanings, and structure across data sources)?



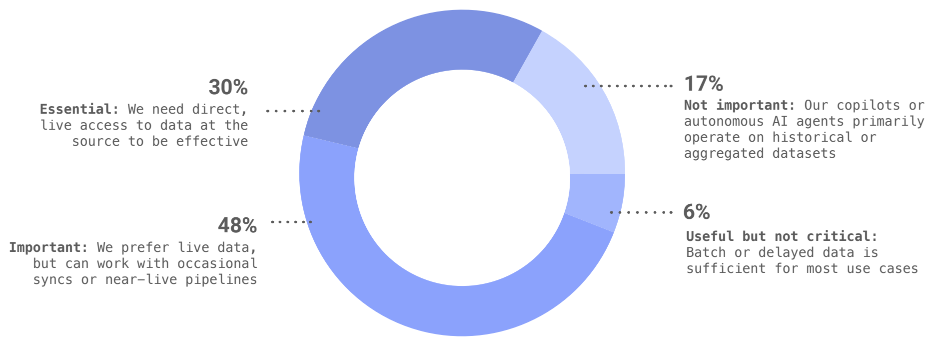
> "Data connectivity is the foundation for AI, and knowledge graphs are the framework for reasoning. As enterprises move from connecting data to understanding it, they're finding that semantics and relationships, not just access, are what make AI accurate, explainable, and ultimately more intelligent."
>
> **— Sudhir Hasbe, President & CPO, Neo4j**

Seventy-three percent of software providers are building or have already built semantic intelligence into their data infrastructure. Among organizations with autonomous agents, 100% have deployed semantic intelligence, reflecting the importance of semantic consistency for agentic AI. But among the general cohort of software providers, only 16% have data architectures supporting semantic intelligence that span all data sources, indicating that investment in this area will grow as software providers continue to advance their AI roadmaps.

> "Before you even think about agents, think about semantic intelligence. Execution will stall without it."
>
> **— Senior Director of Product Management Data and AI, global life sciences technology company**

Similarly, the demand for real-time data integration for enterprise AI use cases is reflected in product leaders' perceived importance of real-time data access. Confirming the consensus from enterprise AI leaders, real-time isn't just a nice to have. Seventy-eight percent of product leaders believe that real-time connectivity is important or essential for in-product AI features.

How important is real-time access to source data for your existing or planned copilots or autonomous AI agents (vs. relying on pre-processed or batch data)?



The survey also found that Anthropic's real-time context injection standard, Model Content Protocol (MCP), is rapidly gaining traction as the de-facto integration standard for in-product AI features, enabling real-time, low-latency access to internal and external data systems. Seventy-six percent of software providers are either exploring or have already implemented MCP as the connectivity standard for context injection into embedded generative AI models.

Is your organization exploring or implementing a Model Context Protocol (MCP) to structure how data is delivered to copilots or autonomous AI agents or models?

MCP adoption response	% of respondents
Not on our radar yet	11%
We use internal conventions for delivering context into models, but not a shared protocol like MCP	9%
We're interested in learning more and evaluating MCP	9%
We've evaluated MCP or similar concepts, but not implemented	31%
We're actively implementing or testing MCP	36%
We rely on ad hoc or prompt-based context injection only	4%

While enthusiasm for MCP is high, early adopters report practical implementation challenges that mirror the broader data integration issues highlighted throughout this report. Many teams note that MCP assumes a level of data consistency, metadata quality, and schema alignment that most organizations have not yet achieved. As a result, rather than resolving fragmentation, MCP can often surface underlying integration weaknesses such as inconsistent entity definitions, lack of semantic models, and limited observability across source systems. Ultimately, the rising prevalence of MCP accelerates the need for semantic governance: without a unified data model and reliable connectivity framework, even the most advanced context protocols risk compounding the very problems they aim to solve.

Implication: Software providers are quickly aligning their AI roadmaps to meet enterprise expectations for real-time, semantically rich data access. Semantic modeling and standards like MCP are no longer fringe experiments. They're starting to see rising adoption for delivering reliable, scalable, and intelligent AI features in product.

> "Before MCP, everything was API-by-API. MCP offers a more structured way to ingest information and handle authentication against the MCP server."
>
> **— Director of Product Management, customer and employee engagement software provider**

The Final Say: The AI Connectivity Imperative

Across both enterprise and software provider contexts, the same theme emerges: AI is ready, but the data isn't. The time and cost of custom integration work is delaying AI features, bloating engineering workloads, and negating business impact.

The companies that win the next phase of the AI race won't necessarily have the best models. They'll have the best plumbing.

This report should serve as a call to action: it's time to invest in the data connectivity, real-time infrastructure, and semantic intelligence that enable AI not just to answer, but to act.

Take our interactive AI readiness assessment

Are you interested in learning where your organization ranks in AI readiness against 200+ data and AI leaders? Take our two-minute assessment to find out:

- Where you fall on the AI maturity curve
- How you compare to leading organizations
- Which gaps in context, connectivity, or control are limiting AI impact today

Take the Assessment →

About CData

By working with CData, you can solve your thorniest data integration and semantic challenges immediately. CData Connect AI is the first managed connectivity platform built specifically for enterprise AI. With MCP, real-time integration, and virtualized SQL access through natural language semantics, enterprises and software providers can clear the most expensive bottleneck in AI: access to connected, contextualized, and controlled data.

Glossary of Terms

A2A (Agent-to-Agent) protocol

An open standard, popularized by Google, that enables communication and interoperability between different AI agents, allowing them to work together to complete tasks. Reduces custom integration work and fragility in agentic workflows.

Agentic AI

AI systems that can plan, reason, and take actions across tools and data sources. In this report, "agents" range from human-in-the-loop to fully autonomous.

API (application programming interface)

A programmatic interface to apps or services. Core surface area for connectors, agents, and copilots to read/write operational data.

Batch integration

Moving or syncing data on a schedule (e.g., nightly). Often reliable for analytics, insufficient alone for low-latency agent actions.

CDC (change data capture)

A data movement pattern that replicates database changes in near-real time. Supports low-latency context for agents and dashboards without full reloads.

Centralized data access layer

A governed, unified interface that provides consistent, contextualized access to multiple systems for AI and analytics. A key maturity marker in the report.

Connector/connectivity

Software that links the data from a source (SaaS app, DB, file store) to a destination or runtime (LLM, agent, data platform).

Copilot

An embedded AI assistant that suggests content or actions but relies on human initiation or approval, useful for knowledge retrieval, insight generation, and decision support.

Data catalog/metadata management

Systems that collect and manage technical and business metadata (schemas, lineage, owners). Enable discovery, governance, and semantic clarity.

Data governance

Policies and controls ensuring data quality, security, privacy, and compliance.

Data integration

The processes and tooling to connect, transform, and deliver data across systems.

Data lineage

Traceability of where data came from and how it changed. Critical for trust, compliance, and debugging AI outputs.

Data pipeline

Orchestrated steps that move and transform data from sources to targets (batch or streaming).

Data quality

Accuracy, completeness, timeliness, and consistency of data. Poor quality degrades AI outputs and user trust.

Data virtualization

Accessing data in place through a unified SQL or API layer without copying it. Useful for live, governed context with lower latency.

ELT (extract, load, transform)

The process of ingesting source data into a data store first, then transforming in place (e.g., in a warehouse/lakehouse). These transformations represent the data in a structure that is better suited for analytics, reporting, and modeling.

ETL (extract, transform, load)

The process of extracting source data and transforming it into a preferred structure, before loading into a destination data store.

Enterprise context

Business-specific data (entities, relationships, rules) that grounds AI outputs. This includes consistent definitions of entities and metrics like “ARR,” “customer,” “opportunity,” and how these entities relate to each other.

HITL (human-in-the-loop)

A control pattern where people review or approve AI suggestions or actions. Common interim step before higher autonomy.

iPaaS (integration platform as a service)

Hosted tools for building and maintaining data integrations and workflows.

Knowledge graph/ontology

Structured representation of entities and their relationships. Provides machine-readable context that improves grounding and reasoning.

Latency (real-time/near-real time)

Delay between data change and its availability. Agents and operational copilots often require sub-second to seconds-level latency.

LLM (Large Language Model)

A foundation model that processes and generates natural language (and often code). Its utility in the enterprise often depends on available context and control.

MCP (Model Context Protocol)

A standard for delivering structured, authenticated, real-time context to models and agents. Reduces bespoke, prompt-only integrations.

RAG (Retrieval-Augmented Generation)

Technique where a model retrieves relevant context (typically unstructured data in the form of documents) before generating an answer. This reduces hallucinations and improves model accuracy, but is unsuitable for structured data with frequent updates.

Real-time integration

Continuous data access from operational systems to power up-to-date insights and agent actions.

ROI (return on investment) tracking

Measuring business impact (e.g., time saved, revenue lift) from AI initiatives.

Semantic intelligence/semantic layer

A shared, business-friendly abstraction (entities, metrics, definitions) that maps source data to consistent meaning. Crucial for scaled AI and tool-agnostic context.

Semantic search

Search that uses embeddings and meaning rather than exact keywords. Often paired with RAG for enterprise knowledge retrieval.

SQL (Structured Query Language)

Standard language for querying and manipulating relational data. Remains the lingua franca for structured enterprise context.

Tool sprawl

Proliferation of overlapping AI and data tools. Increases integration complexity and fragments context.

Workflow orchestration

Coordinating tasks and dependencies across pipelines, connectors, and agent actions. Ensures reliable end-to-end outcomes.