KPMG

AI-ready data:

Close the five gaps preventing enterprise AI from scaling

A CDAO guide to the searchability, context, trust, governance, and operating model gaps

KPMG.com

Why enterprise AI needs AI-ready data, not just good data

Company leaders are no longer asking for smarter chatbots; they're asking for agents that reason, solve problems, and take action. But most enterprise data environments were built for people reading dashboards—not for machines that must search, interpret, and act within policy.

To move beyond AI pilots, CDAOs need data that agents can use, not just data people can analyze. That means data must be searchable across the enterprise, contextual enough for machine reasoning, and trusted enough for governed action.

These gaps include searchability hurdles caused by fragmented or unconnected data. A second area is related to context. AI can retrieve data, but without context, AI cannot understand its meaning, lineage, or business significance. The final gap is trust. To achieve true transformation with autonomous agents, AI must act independently within a governed, logical framework.

By closing gaps, organizations can produce AI-ready data that is searchable, contextual, and trusted—data the way AI agents need to reason and solve complex challenges.

What is AI-ready data?

AI-ready data is data that AI systems can find, interpret, and use safely. It's searchable across structured, unstructured, and dark assets; contextual through semantic layers, ontology, metadata, and knowledge graphs; and trusted through lineage, permissions, policy logic, and auditable controls. It is not simply clean data. It is data designed for retrieval, reasoning, and governed action.

AI data readiness diagnostic:

Can your data support search, reasoning, and governed action?

For one priority AI use case, ask whether the data foundation can support the three things agents need to do:

Search

- Can AI find the relevant structured, unstructured, and dark assets?
- Can the organization identify the authoritative source for each critical business entity?

Reason

- Are definitions, relationships, exceptions, and business rules encoded in a machine-readable way?
- Can AI interpret meaning consistently across domains?

Action

- Are lineage, permissions, and policy constraints available at runtime?
- Can outputs be traced back to the data, rules, and controls that shaped them?

If the answer is unclear, the organization does not yet have an AI-ready data foundation. It has data that still requires humans to compensate for what agents cannot search, interpret, or govern on their own.

Five AI data readiness gaps keeping enterprise AI in pilot mode

Most organizations have data useable for dashboards, reporting, and advanced analytics. It's good data that falls short for AI models that need AI-ready data to search across the enterprise, interpret business meaning, and act within policy.

AI-ready data has three requirements. It must be searchable, so AI can find and connect the right data across systems, domains, and formats. It must be contextual, so AI can understand meaning, relationships, and business rules instead of simply retrieving records. And it must be trusted, offering clear lineage, machine-readable permissions, and decision logic that makes AI actions traceable and auditable.

Here are the five gaps in data management programs preventing AI from escaping pilot purgatory.

GAP 1

Fragmented enterprise data creates an AI searchability gap

Teams refuse to give up their spreadsheets because so-called authoritative sources can't be relied on, or new AI use cases get bogged down in data discovery.

In most large enterprises, data accumulates—system by system, acquisition by acquisition, initiative by initiative—in CRMs, ERPs, data lakes, legacy platforms, and SaaS applications that never agreed on definitions and were never required to. The result is a fragmented data landscape where "customer" means one thing to Sales, another to Finance, and something else entirely to Marketing. Lineage is murky, mappings are brittle, and every new AI use case requires someone to reestablish connections.

For AI, fragmentation is not just inefficiency. It creates an incomplete evidence base, forcing agents to reason from partial signals, conflicting definitions, and uncertain lineage. Fragmentation doesn't just slow AI, it makes agentic AI structurally impossible. If data is fragmented, an agent needing a unified view of a customer or risk without a human intermediary will fail.

Addressing fragmentation is the prerequisite for everything that follows. Solving it produces a shared semantic foundation: a single source of truth for each critical business entity, resolvable across systems, accessible to machines and humans. When fragmentation is addressed, ambiguity is replaced by a consistent foundation on which trust can be built, context can be formalized, and governance can stick. Data becomes searchable enough for AI to ground outputs on a more complete view of the business.

GAP 2

Low trust in data keeps AI agents stuck in human supervision

Two dashboards, two numbers, zero trust.

For AI, the trust problem is not only whether a number is correct. It is whether the organization is willing to let a machine act on that number. Executives who revert to gut instinct and analysts who spend more time reconciling data than generating insights are the result of an organization that lacks trust in its data. It's explainable.

When data is scattered across systems that never agreed on definitions, reconciliation becomes manual, perpetual, and exhausting—and the longer it stays that way, the more confidence erodes. Not just confidence in the data, but confidence in the teams responsible for it. This dynamic predates AI entirely. Agentic AI makes it an existential risk. Humans previously caught errors. Agents act on what they're given at speed, at scale, without hesitation, generating confident, plausible, yet completely wrong outputs.

With traditional analytics, a human is always in the loop to catch the error. AI agents don't do that. Low trust forces humans back into the loop—not to add judgement but to compensate for missing what AI needs to resolve the task on its own. When wrong AI outputs reach customers, regulators, or the board, it's a business risk that must be managed manually for accuracy's sake, which slows execution.

Trusted data for AI is about execution, not just accuracy. When lineage is visible, sources are declared, and quality is monitored continuously by agents. AI outputs inherit more of the trust built into the data foundation. AI finally has what it needs: AI-ready data.

GAP 3

Missing context prevents AI agents from reasoning over enterprise data

AI can process fields, rows, and documents at speed, but that does not mean AI understands these data points from a business context. Data without context lacks the exactness of numbers, which makes AI capable of retrieving information but it's not interpretable and actionable. AI agents produce outputs that are polished, confident, and fluent — and sometimes completely wrong in ways that aren't obvious until the damage is done. This is what happens when an agent acts in good faith on data that isn't wrong exactly — it's just stripped of the context that would tell the agent what it truly means.

Inaccurate AI output can sound utterly convincing. For example, a customer service agent offers a high-value customer a standard win-back discount — the kind reserved for lapsed accounts. It's perfectly logical based on the transaction data. The data was not necessarily wrong; it lacked the context that would have changed the action. The customer has been with the company for years, never lapsed. The company loses money on the transaction and puts the lifetime value of a loyal customer at risk. The point is, agents acting without meaning don't just make harmless mistakes; they scale these mistakes that can hurt the bottom line and tarnish the company's reputation.

The fix is to attach meaning to data— so the agent arrives at every task knowing what the numbers mean, where they came from, what rules govern their use, and what isn't permitted. Semantics, lineage, business rules, sensitivity flags, and ownership information are

embedded in machine-readable form and travels with data. When context is inherent with data, agents stop guessing. They reason. This is the core requirement of AI-ready data: not just access, but meaning, permissions, and context that AI systems can use safely.

GAP 4

Manual governance can't keep pace with agentic AI

Your best AI use cases are dying in the approval queue. High-potential AI use cases often slow down in governance review because Legal, Privacy, and Risk teams must reconstruct lineage, sensitivity, permissions, and intended use before approving deployment. Meanwhile, motivated AI champions grow increasingly frustrated when ideas for AI speed are consistently stalled, leading them to stop trying.

Traditional governance frameworks were designed for human-speed consumption—like monthly dashboard reviews, not microsecond machine decisions. The problem is compounded by human interpretation and human-paced decision-making: policies written in documents, risk assessments conducted manually, approvals dependent on individuals who must reconstruct intent, lineage, and sensitivity with every new use case. For traditional analytics, this was slow but workable. For agentic AI, it is a structural incompatibility. Human review loops reintroduced for safety nullify autonomy, forcing AI systems to slow down governance to human speed.

Governance that travels with data functions at agent speed. When this happens, governance moves from document to infrastructure. Policies encoded as rules travel with the data. Sensitivity flags and usage permissions are embedded and machine-readable. Lineage and certification status are available at runtime. When governance is built into the pipeline rather than bolted on at the end, the approval queue shrinks because the evidence reviewers need is present and inspectable. Compliance becomes integral to how AI runs, not a gate it must pass through.

What runtime governance requires

Runtime governance means AI systems can determine, at the moment of use, what data they may access, which permissions apply, what policy constraints shape the action, and how the decision will be traced. This is not traditional governance moving faster. It is governance designed for AI execution.

GAP 5

Operating model gaps leave AI-ready data work ownerless

Operating model gaps become more visible in AI programs because AI-ready data requires new work that most organizations have not assigned to anyone. Ownership is unclear, accountability is contested, and incentives in business units work against the collaboration needed to support AI data readiness. IT claims responsibility for infrastructure, not quality or meaning—that's the business's job. The business claims it lacks the technical capability—that's IT's job. Treating data as a mere byproduct of IT applications is no longer a viable operating model. Data science sits in the middle, building models on a foundation neither side has agreed to maintain.

Fragmentation can't be resolved when nobody with domain authority is accountable for definitions. Context can't be formalized when the people who hold the meaning haven't accepted responsibility for attaching it to the data. This leaves the CDAO in an unwinnable position: accountable for the quality, meaning, and trustworthiness of data that only the business has the authority to own. It's not a technology problem. It's an organizational one that the operating model must solve.

An effective AI-ready data operating model assigns ownership for the work agents depend on: who certifies AI-ready data products, who maintains semantic standards, who owns ontology and business rules, who approves permissions, and who monitors decision logic at runtime. Without that clarity, the CDAO remains accountable for AI data readiness while the authority to define meaning, rules, and use cases remains scattered across the business.

AI-specific ownership that must be defined

- **Data discovery ownership:** who catalogs structured, unstructured, and dark assets for AI use cases
- **Semantic ownership:** who defines business meaning across domains
- **Ontology ownership:** who maintains relationships, rules, and exceptions
- **AI-ready data product ownership:** who certifies data for machine reasoning and governed action
- **Runtime control ownership:** who monitors permissions, decision logic, and auditability

Industry stakes: How AI data readiness gaps show up in regulated and fast-moving sectors

While data issues are universal, they manifest differently for industries due to specific regulatory and competitive environments. The same gaps show up differently depending on regulatory pressure, customer expectations, and the speed at which decisions must be made:

- **Asset Management / Financial Services:** In heavily regulated environments, AI agents cannot make a trading recommendation or assess risk without absolute data provenance and observability. A confident but poorly grounded AI recommendation can create model-risk, suitability, reporting, or compliance exposure.
- **Healthcare:** To manage siloed, HIPAA-governed data, an AI agent needs interoperability (FHIR/HL7) to bridge systems and privacy-preserving architectures to secure sensitive information. It requires "Trusted AI" governance with human-in-the-loop oversight to ensure clinical accuracy and specialized healthcare accelerators for medical logic. This ensures agents extract insights within secure environments without violating regulatory mandates or compromising patient privacy.
- **Retail/Consumer Markets:** The speed of consumer trends requires AI to act instantly. If the data operating model requires weeks of manual reconciliation between inventory and marketing systems, the AI's insights are useless by the time they are approved. Retail moves too quickly for AI outputs that require weeks of reconciliation before action.

What changes when enterprise data becomes AI-ready

When enterprise data becomes searchable, contextual, and trusted, AI systems can use it as a foundation for reasoning and governed action. Agents can search across the full evidence base instead of the narrow slice that happened to be indexed first.

They can reason in business terms, and they can act within traceable rules that define what they are allowed to do and how decisions are audited.

That is the dividing line between AI pilots and enterprise AI. The organization is no longer asking humans to compensate for hidden assets, missing context, and unclear controls after the fact. The CDAO is embedding AI-ready data requirements into the foundation itself: searchability, context, trust, lineage, permissions, and decision logic.

The shift is significant: once AI-ready data establishes a baseline of trust, agents can begin earning autonomy to execute processes independently. It's best illustrated with a before (data not AI-ready) and after (AI-ready data) example.

Before: An agent hallucinates a fix or forces a human to spend four hours digging through three different systems to verify the agent's suggestion.

After: The agent instantly accesses the contextualized, trusted data, understands the constraints, and autonomously executes the optimal resolution within safe governance boundaries.

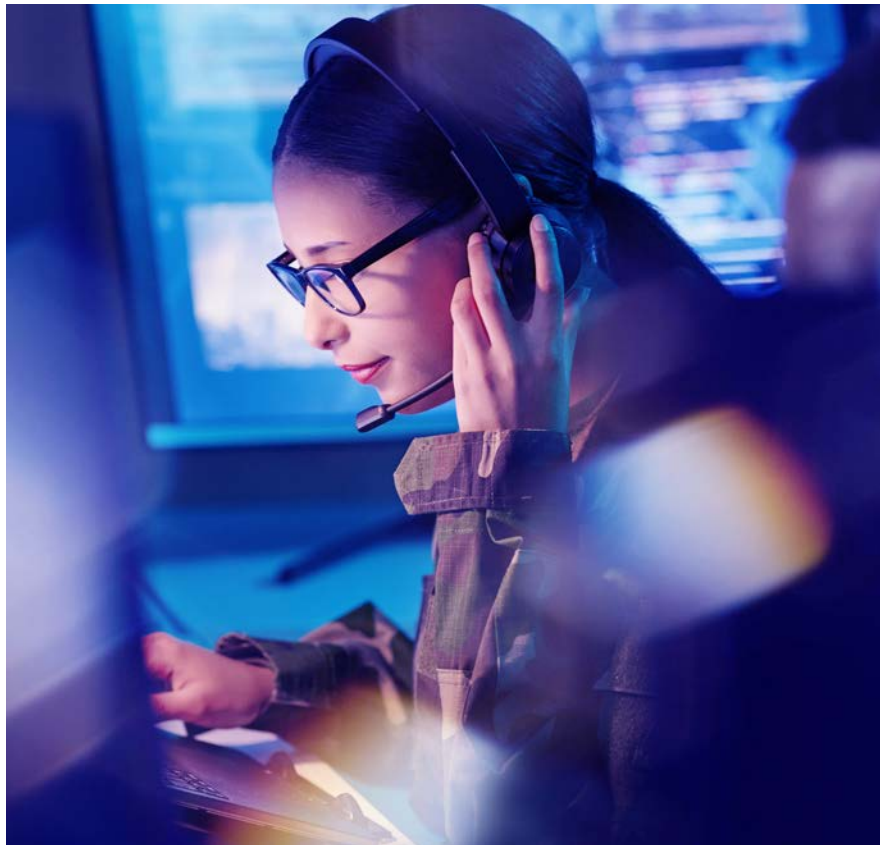
AI-ready data is the prerequisite for moving from human-in-the-loop to fully autonomous agentic systems.

KPMG LLP helps CDAOs identify and close the AI data readiness gaps that can prevent enterprise AI from scaling: searchability, context, trust, governance, and AI-specific ownership.

How KPMG helps CDAOs identify and close AI data readiness gaps

Our work typically begins with a priority AI use case, then diagnoses where searchability, context, and trust break down.

From there, KPMG helps organizations connect the underlying data estate, including dark assets; engineer semantic layers, ontology, and knowledge structures; design AI-ready data products; and embed lineage, permissions, and decision controls so AI outcomes are explainable, defensible, and scalable. The result is a business-first path to trusted, governed, contextual data that supports AI safely, proves value, and scales with the enterprise.



Contact us

Talk to us about how we can help you achieve your enterprise AI objectives.

Matteo Colombo

Principal, Global and U.S. Advisory
KPMG LLP

MatteoColombo@kpmg.com

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

Learn about us: **kpmg.com**