KPMG

# The roadmap to AI-ready data for enterprise AI

A five-move guide for CDAOs to make enterprise data searchable, contextual, and trusted for AI agents, RAG, and autonomous workflows

KPMG.com

# Unprepared data stalls enterprise autonomous AI

Chief Data and AI Officers (CDAOs) face immense pressure to fuel agentic AI, yet enterprise efforts often stall because data isn't prepared for autonomous work. The problem isn't the large language models; it's that data cannot be globally searched, interpreted through business context, and trusted for governed action.

The initial wave of generative AI, which was characterized by conversational assistants and passive copilots—is rapidly giving way to a much more profound enterprise transformation: the rise of agentic AI. We are no longer building tools; we are onboarding a new class of reasoning agents and superagents that require AI-ready data. These virtual coworkers or digital employees are poised to redefine how work gets done, shifting from simply generating text to autonomously executing complex, multi-step workflows.

To bridge this gap, organizations should follow this roadmap to engineering AI-ready data for enterprise AI. The journey focuses on producing data that is searchable for grounding, contextual for reasoning, and trusted for action. By engineering AI-ready data products and encoding business meaning into machine-readable structures, autonomous agents can interpret, reason, and act independently. To be able to trust agents to make decisions, enterprises must fundamentally rewire their data foundations.

This doesn't require a major overhaul. Instead, it requires a sequenced roadmap that delivers early value while raising the autonomy ceiling with every move. The path goes from cataloging data and creating trust to engineering context, embedding governance, and upgrading the operating model. This is the roadmap to power a truly agentic enterprise.

# The value of searchability, context, and trust in agentic AI

This roadmap fuels the journey to agentic AI, which requires:

**Searchability**

AI can find and connect to the right data

**Context**

AI can interpret meaning and relationships

**Trust**

AI can act safely with traceable decisions

If these conditions are not met, AI remains limited to pilots.



# How this roadmap enables enterprise AI to scale safely

Many data programs begin with a technology stack and work outward. This roadmap begins with a priority AI outcome and works backward to the data, context, governance, and operating model capabilities required to support it.

The question is not, "How do we modernize all data?" It is, "What does this use case need in order for AI to search the right evidence, interpret business meaning correctly, and act within policy?" That shift matters because it keeps investment defensible, aligns the work to business outcomes, and avoids platform-first programs that never get close enough to execution to prove value.

It also matches how enterprise AI matures. Early copilots can survive on partial context because humans are constantly correcting them. Agents and autonomous workflows cannot. They need authoritative data sources, machine-readable metadata, consistent semantic meaning, clear permissions, and traceable decisions. In other words, they need a data foundation built for AI, not just analytics.

MOVE 1:

# Build enterprise searchability by tracking down and cataloging all data

## What this solves: AI cannot reason over data it cannot find. Enterprise searchability creates the evidence base agents need to ground responses, recommendations, and actions.

The first move is to solve the enterprise searchability problem. An AI model cannot ground responses, assemble quality recommendations, or trigger preferred actions based on data it cannot find or partially access.

This is why the first step is to evolve data products to be AI-ready with discoverability and searchability of the full data estate. It's easier to start with one or two AI wished-for outcomes and develop muscle memory. That means cataloging the known platforms, but also the dark assets (data generated business activities, but never used) along with policies, procedures, contracts, emails, historical files, ticketing records, PDFs, and local repositories. In many organizations, these sources contain the logic and evidence that AI needs. There is also the collective knowledge of employees that can be digitized through AI-driven interviews using a knowledge elicitation tool.

A usable enterprise AI data catalog must be machine interpretable. It should identify authoritative sources, track lineage, preserve sensitivity and access classifications, and make structured and unstructured sources discoverable through a common search path. For retrieval-augmented generation and agentic workflows, this is the prerequisite to better grounding. Search quality is only as good as the evidence base AI is allowed to search.



# What to catalog first for enterprise AI

Start with the sources likely to shape a high-value AI decision:

- Structured systems for customer, product, transactions, finance, and risk data
- Policy and procedure documents that govern what AI is allowed to do
- Contracts, communications, and historical records that provide business/legal context
- Dark assets are volumes of structured/unstructured data that are hidden, siloed, or completely untracked across the organization

The goal is not better visibility. It is to create a trusted search surface for AI, ensuring agents have the full picture of the business rather than a narrow slice of database records.

MOVE 2:

# Create AI-ready data products

## What this solves: Data prepared for human analysis prevents AI from scaling, as data assets lack clear ownership, embedded rules, and certified trust required for automated execution.

With data cataloged, the next move is to package priority domains into AI-ready data products. Traditional data products, often built for human analysis in dashboards, are insufficient for agentic AI. AI requires data it can compose, reason over, and act upon—not just analyze. This shift from analysis to enabling governed execution through AI-ready data products is the core of creating a data product that both machines and humans can use.

An AI-ready data product is a certified, reusable asset bundled with rich, machine-readable metadata. This includes a clear business definition, authoritative lineage, quality thresholds, and embedded permissions and purpose limitations. It must also define its interfaces for how it can be combined with other products, ensuring it can be used across various Retreival-Augmented Generation (RAG) pipelines, copilots, and agentic workflows without bespoke integration work.

This evolution from a human-centric asset to a machine-operable and human interpretable one involves several critical transformations:

- **From data catalogs to semantic marketplaces:** The classic data catalog, designed for humans, must evolve into a marketplace that includes a machine-interpretable semantic layer. This allows an AI agent to autonomously discover, understand the context of, and reason over the products it finds.
- **From custom-curated to agent-composable:** Instead of relying on engineers to create custom, one-off data assets for every new workflow, AI-ready products are designed with standard interfaces that allow an agent to compose them on the fly.
- **From human IAM to digital identities:** Governance must also become agent-native. The model of human-only Identity and Access Management (IAM) must be extended to grant digital employee identities and entitlements to AI agents, allowing their permissions to be auditable and their actions governed.
- **From retrospective DQ to runtime trust:** Trust can no longer be a backward-looking metric on a dashboard. AI-ready products must emit runtime trust signals, allowing an agent to verify data quality and lineage at the moment of execution.

The most effective approach is to identify five to ten domains that matter most to the business—such as customer, product, risk, or supplier—and assign owners to productize them for both human and machine use. This practice moves the enterprise toward building a scalable, trusted foundation. Transforming raw data into AI-ready data products enables organizations to establish fundamental components essential for ensuring the reliable and confident performance of AI systems.



MOVE 3:

# Engineer context with semantic layers, ontology, and knowledge graphs

## What this solves: AI must interpret meaning, not just retrieve data. With contextualized data, AI is equipped to reason and, most importantly, to act reliably.

With a trusted data foundation in place, the next challenge is to solve for meaning and action. This is the work of context engineering, which relies on two core, symbiotic components: an ontology that provides business context, and an agentic harness that governs action.

The ontology serves as the brain of the enterprise, defining the business in a machine-readable format. It starts with a semantic layer—a consistent business vocabulary that maps technical fields to the nouns of the business (e.g., customers, products). From there, the ontology defines the verbs—the rules, relationships, and logic that dictate how those entities interact (e.g., a 'premium customer' is the same as a 'strategic account'). It tells an agent not just that objects are related, but how and why, providing the rich context needed for sophisticated reasoning.

However, context without a mechanism for governed action is incomplete. This is the role of the agentic harness, which can be thought of as the decision and execution layer for the ontology. As a structured architectural layer, the harness controls the planning and execution of agentic behaviors, ensuring that the context provided by the ontology is used safely and reliably. Within the KPMG AI framework, the harness provides critical capabilities:

- Enforces explicit planning to help ensure agents remain aligned with high-level goals.
- Provides behavioral bounding to limit agent actions to defined safety and policy constraints.
- Incorporates persistent memory to maintain context during complex, multi-step tasks.
- Uses hierarchical control to separate high-level strategy from low-level execution.

Finally, a knowledge graph brings these layers to life. It takes the physical data from AI systems, interprets it through the semantic layer, and structures it according to the rules set by the ontology. This creates a dynamic and navigable map of the business that the agentic harness can traverse to execute complex tasks, which transform AI from a simple retrieval tool into a powerful, governed system for autonomous business execution.



# Why AI requires ontology, not just traditional data models

A data model tells systems how to store and organize data.

An ontology tells AI how to interpret meaning, elationships, and business rules.

That distinction is one of the clearest dividing lines between data that is merely usable for reporting and data that is ready for enterprise AI.

MOVE 4:

# Build trust through embedded governance

## What this solves: The fundamental conflict between slow, human-paced governance and the speed of agentic AI, which forces constant manual supervision and prevents autonomous action at sale.

To move beyond human supervised AI, trust must be built directly into your data foundation. Traditional governance, designed for human-paced decisions, is a structural barrier to the speed and scale of agentic AI. The solution is to create a decision layer by embedding governance directly into the runtime environment. To earn autonomy and embed governance, incorporate the types of human-AI interaction and define areas of use: human-in-the-loop for direct approval; human-in-the-loop for supervisory oversight, and human-in-the-loop feedback for continuous improvement.

This decision hierarchy layer ensures probabilistic systems operate within deterministic business logic. Instead of residing in static documents, policies, permissions, lineage, and purpose limitations are encoded as machine-readable rules that travel with the data itself. When governance is an inseparable part of the data pipeline, AI agents can act securely and decisively without requiring constant manual supervision.

This enables earned autonomy, shifting the enterprise from human for every decision to human reserved for strategic oversight. By making automated actions traceable, auditable, and compliant by design, the organization gains the confidence required to let AI execute complex, high-value work safely and at scale.

### The price of unready data

Pursuing AI without preparing data for AI comes with risks and potential consequences. Low trust in data traps AI in a cycle of manual supervision, throttling its potential from machine speed down to human speed.

Without AI-ready data, there is a greater risk of compliance penalties, brand and reputational damage, and legal liability. Autonomous outputs, actions, and decisions that are incorrect can violate policies and rules, as well as put your firm in a bad light. Also, beware of novel security threats in an agentic environment, including agent goal hijacking, ontology poisoning, and identity and privilege abuse for non-human identities.

MOVE 5:

# Establish the operating model required to sustain AI-ready data

## The final move calls for upgrading the operating model to turn AI initiatives from isolated projects into repeatable enterprise capabilities.

The core challenge is ownership. Tasks unique to AI—such as maintaining ontologies, establishing runtime controls, and certifying data for AI agents—often lack clear accountability. If these responsibilities remain informal, scaling AI becomes impossible.

An operating paradigm called "AgentOps" can help. Domain teams own the quality and business context of their AI-ready data products, while a central AI function dictates enterprise standards for governance, interoperability, and semantic patterns. To sustain this, the enterprise must shift to an AI product management mindset—treating AI as a continuous lifecycle requiring ongoing model evaluation and data pipeline monitoring.

Execution requires small, cross-functional pods combining domain SMEs, knowledge engineers, and AI platform engineers to seamlessly translate business logic into machine-usable assets.

Finally, mandate explicit value realization. Anchor AI use cases to concrete business metrics like revenue lift, risk reduction, or cycle time. Start with high-visibility projects to prove a central thesis, like "contextual, trusted data accelerates enterprise AI performance."

With data that's searchable, contextual, and trusted, along with an AI-first mentality, the organization will also need to evolve with new roles and responsibilities like forward deployment ontologist curators as well as environmental ontologists charged with adapting to the specific context and nuances of the business environment. A Fusion Team Lead may oversee hybrid teams of human experts and AI virtual coworkers.

It will require a transformation of the workplace, not to mention a change-management initiative to help every employee embrace a new way of working.



# What makes a data product AI-ready

A traditional data product supports analysis. An AI-ready data product goes further: it is certified for retrieval, reasoning, composition, and governed execution.

**Features include:**

- Machine-readable business definitions and semantic labels
- Explicit lineage, embedded permissions, certification status, purpose limitations
- Defined interfaces for retrieval, composition, and agent access

**Good data products support analysis. AI data products support retrieval, reasoning, and governed execution.**

# What enterprise AI looks like when data is AI-ready

When this roadmap is realized, the enterprise goes from passive chatbots to true agentic AI. Agents search across certified structured and unstructured sources, interpreting business meaning consistently through a shared ontology.

Because permissions, lineage, and decision logic are embedded at runtime, these agents can act securely and decisively within defined guardrails.

This transforms the business into a cohesive, reasoning organism where agentic AI orchestrates complex outcomes. For example, during a supply chain disruption, an AI agent not only flags a delay, it also autonomously reroutes logistics by cross-referencing real-time warehouse inventory, supplier ESG certifications, and contract penalty logic. It executes a complete recovery plan in minutes. This supply chain snafu once took cross-functional teams weeks of manual coordination.

The result is a reusable, trusted data foundation that allows the organization to scale agentic AI safely and defend automated actions. It raises the confidence level required to let AI do real work, shifting a portion of the human workforce from manual execution to strategic oversight. Business agility has evolved to a state of earned autonomy where data serves as a living, machine-readable strategy powering intelligent, enterprise agents.

## The shift from AI experiments to enterprise AI

**Most organizations are still here:**
- AI pilots, copilots, isolated wins

**Few have reached here:**
- AI embedded in processes, decisions, and operations

The difference is not models. It is whether data is **searchable, contextual, and trusted enough to support execution.**

# How KPMG helps CDAOs build and scale AI-ready data for enterprise AI

KPMG LLP (KPMG) helps CDAOs turn AI-ready data from a strategic priority into an execution model. The work begins with the business outcomes leaders are already trying to achieve: safer AI adoption, faster time to value, trusted outputs, stronger governance, and agentic workflows that can scale beyond isolated pilots.

AI-ready data requires coordination across teams that often move separately: data, risk, technology, governance, business domains, and AI engineering. KPMG helps bring those capabilities together around the AI use cases that matter most, so the organization can build data that is searchable, contextual, and trusted enough for enterprise AI execution.

Our approach starts with a priority AI use case and works backward. We identify what data the AI system needs to find, what context it needs to interpret, what rules it must follow, and what evidence the business needs to trust the outcome. From there, we help CDAOs build the data foundation, governance model, and operating discipline required to scale.

The work typically progresses through six connected moves:

## 1 Diagnose AI readiness priority use cases

We start where the business already feels urgency: a use case tied to revenue, risk, customer experience, compliance, efficiency, or operational resilience.

**Rather than assessing data maturity in the abstract, we evaluate where AI data readiness breaks down:**

- Can the AI system search the full evidence base, including dark assets?
- Is business context missing, inconsistent, or trapped in human judgment?
- Are lineage, permissions, and purpose constraints available at runtime?
- Is ownership clear for AI-ready data products, semantic standards, ontology, and decision logic?

This gives CDAOs a practical starting point and a defensible business case for investment.

## 2 Build the searchable data foundation

Enterprise AI cannot reason over data it cannot find.

KPMG helps organizations create a searchable data foundation across structured, unstructured, and dark assets. That includes cataloging relevant sources, identifying authoritative data, classifying sensitivity, and establishing lineage so AI systems can ground responses, recommendations, and actions in the right evidence.

The goal is not visibility for its own sake. The goal is to give AI systems access to the data they need—and only the data they are allowed to use.

## 3 Create AI-ready data products for agents and RAG

Traditional data products were built for people. AI-ready data products must also work for machines.

KPMG helps organizations define and build AI-ready data products with machine-readable definitions, lineage, permissions, purpose limitations, quality thresholds, and interfaces for retrieval, composition, and agent access.

This helps CDAOs move away from one-off data preparation for every AI use case and toward reusable assets that can support copilots, RAG pipelines, and agentic workflows.

## 4 Engineer semantic context for AI reasoning

AI cannot scale on access alone. It needs business meaning.

KPMG helps design the context layer enterprise AI depends on: semantic layers, ontology, knowledge graphs, and metadata patterns that allow AI systems to interpret meaning consistently across domains.

This is where data moves from being available to being usable for reasoning. AI can connect customer, product, policy, pricing, risk, and compliance context instead of returning disconnected fragments.

## 5 Embed trust, governance, and decision logic into execution

AI governance cannot remain separate from data execution.

KPMG helps organizations embed lineage, permissions, purpose constraints, runtime controls, and decision logic into the data foundation itself. That allows AI systems to act within defined boundaries and gives the business a way to explain, audit, and defend outcomes.

This is the shift from manual governance review to governed execution.

## 6 Establish the operating model to sustain AI-ready data

AI-ready data cannot stay current without ownership.

KPMG helps CDAOs define the operating model required to sustain AI-ready data across use cases. That includes ownership for data products, semantic standards, ontology, decision logic, runtime controls, and federated governance.

Domain teams own meaning and business context. The central data and AI function sets standards, patterns, and guardrails. Together, they make AI-ready data repeatable rather than bespoke.

# What makes KPMG different

## Many data approaches stop at platforms, pipelines, or isolated AI pilots. KPMG focuses on the layers enterprise AI depends on most:

- **Meaning:** semantic context, ontology, knowledge graphs, and metadata
- **Permission:** governance, access, purpose limitations, and runtime controls
- **Traceability:** lineage, certification, decision logic, and auditability
- **Execution:** AI-ready data products, operating model design, and value realization

KPMG also brings industry context to this work. The ontology, evidence model, policy logic, and governance controls required in financial services, healthcare, consumer, industrial, or public sector environments are not interchangeable. What counts as trusted, contextual, and usable data depends on the business process, the regulatory environment, and the consequences of AI action.

That is why our approach starts with the use case and the industry context—not the platform.

With KPMG, CDAOs can move from fragmented data conditions to an AI-ready foundation where systems can search, reason, and act reliably.

**The result is not just better data. It is an operating foundation that helps AI:**

- Ground outputs in the right evidence
- Interpret business meaning consistently
- Act within policy and permission boundaries
- Scale across use cases without rebuilding from scratch
- Produce outcomes the business can explain, defend, and measure

The question for CDAOs is no longer whether AI needs better data. It is whether the enterprise can make data searchable, contextual, and trusted quickly enough to support the AI outcomes the business now expects.

**KPMG helps make that work executable.**

# Contact us

Talk to us about how we can map and help you target your next AI-ready enterprise 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**