The knowledge engineering imperative

January 2026

kpmg.us/ai

Image created with GenAI

Introduction

What does it take for companies to thrive in a world where humans and artificial intelligence (AI) agents have vastly different data needs? That's the question data leaders need to answer as they scale agentic AI from isolated pilots to enterprise-wide reality.

To make that vision real, companies must confront a fundamental mismatch between the way humans consume data and the way AI agents process it. This mismatch exposes the limits of traditional data tools like spreadsheets and dashboards that were built for human eyes, not machine intelligence. Created by humans for humans, they present data visually and in isolated rows and columns, assuming a person will interpret patterns and context.

Consider, for example, how an experienced human user intuitively understands the meaning of a column labeled "customer" in a spreadsheet, while a new hire needs time to learn internal nomenclature. Like a new hire, a new AI agent must go through a learning period to be production ready. Without semantic context, AI agents can't distinguish whether "customer" refers to a person or a company. And without this language-defining transformation, feeding traditional reports to an AI agent is like handing a human a pile of raw binary code: Technically it's data, but from a usability perspective, it's incomprehensible.

For agentic AI, the formats long relied upon by people become a barrier. AI agents are just learning to "see" charts and infer meaning from labels without additional structure. They still require data that is semantically rich, machine-readable, and interconnected. This richness gives data context, meaning, and relationships, enabling AI agents to reason, make decisions, and act autonomously.

> **For agentic AI, the formats long relied upon by people become a barrier. AI agents are just learning to "see" charts and infer meaning from labels without additional structure.**

If structured data is hard, then unstructured data like emails, videos, and chat logs is even harder for AI agents to interpret. Unstructured content, which now accounts for nearly 80 percent of all enterprise information,[1] lacks the consistent formatting and defined relationships that intelligent systems need. It also amplifies the call for a new approach that organizes and connects information in a way machines and humans can both understand. And, unless companies adapt, efforts to scale AI will stall, decision-making will drag, and competitive advantage will erode.

These challenges demand a new paradigm that bridges the gap between human intuition and a machine's intelligent capabilities. Knowledge engineering—the discipline of capturing and structuring an organization's information so machines can interpret and use it effectively—provides the answer. It ensures that AI systems don't just generate answers from patterns, but can also rely on trusted, contextual, business-specific knowledge. In the pages ahead, we'll demystify knowledge engineering and outline practical steps data leaders can take to accelerate the transition to a knowledge engineering framework, helping companies unlock faster, smarter, and more profitable ways to succeed in the era of agentic AI.

[1] IDC FutureScape: Worldwide Data and Analytics 2025 Predictions

The business case for knowledge engineering

Knowledge engineering is a strategic investment that directly impacts cost efficiency and return on investment (ROI). This strategic value becomes especially clear when evaluating how different AI architectures affect operational costs.

For example, although large, general-purpose language models can solve almost any problem, they are expensive for narrow use cases because they usually consume significant compute resources and tokens. Smaller, domain-specific models are more cost effective, but without robust knowledge engineering in place, even specialized models can produce incomplete or misleading outputs. The bottom line: As foundational models evolve and raw data dependency declines, context and curated knowledge remain critical for competitive advantage, relevance, and ROI.

With a well-designed knowledge engineering system, companies can enable transformative business outcomes that go far beyond incremental improvements. Consider a digital procurement assistant tasked with recommending the best supplier for a new product line. It can easily retrieve performance reports, contract terms, shipping histories, and risk assessments. But without a structured understanding of how your company defines “preferred supplier,” how exceptions are handled, or how risk tolerances vary by region, the assistant cannot reliably make decisions—no matter how powerful the underlying model is.

Knowledge engineering closes that gap by giving AI agents the semantic scaffolding they need to interpret information as your business intends. This shift unlocks a series of compounding benefits that reach far beyond data quality or model accuracy.

> Knowledge engineering closes that gap by giving AI agents the semantic scaffolding they need to interpret information as your business intends. This shift unlocks a series of compounding benefits that reach far beyond data quality or model accuracy.

With a contextual foundation in place, leaders gain the ability to explore dynamic, interconnected scenarios rather than static projections. That same procurement assistant could simulate how a sudden tariff change might ripple across supplier networks, or how shifting demand in one market could affect lead times in another. Instead of reacting to disruptions, executives can see several moves ahead by testing “what if” situations that enable them to evaluate trade-offs in real time.

Customer and employee experiences transform as well. When data is unified and semantically enriched, AI-powered assistants can understand complex, multi system questions such as, “What’s the status of my order across all suppliers and regions?” or “What steps are needed to onboard this customer under our latest policy update?” Instead of routing issues or escalating tickets, these assistants produce fast, precise, personalized responses that improve customer satisfaction and reduce operational strain.

Decision support becomes dramatically more autonomous and more trustworthy. An AI agent reviewing a supplier contract, for example, can connect clauses to historical disputes, procurement policies, and compliance guidelines. It can identify risk patterns, recommend negotiation changes, or compare contract terms to similar agreements across the enterprise. The agent isn't just surfacing data—it's interpreting relationships, understanding business logic, and delivering judgment with explainability built in.

> Knowledge engineering determines whether a company's investment in AI becomes a sustainable competitive advantage or a series of expensive experiments.

Knowledge engineering also elevates automation from rote task execution to intelligent workflow orchestration. When AI understands the rules, dependencies, and exception paths inside complex processes such as regulatory reporting or cross border transactions, it can execute end-to-end flows, adapt when conditions shift, and deliver consistent results without constant human oversight.

And because the system captures how the business defines terms, decisions, and processes, it becomes a living map of enterprise expertise. New employees and AI agents can onboard quickly, drawing from shared definitions, prior decisions, and networks of subject matter experts. This accelerates ramp up time while also protecting institutional knowledge that otherwise risks leaving the organization as people change roles.

Ultimately, knowledge engineering determines whether a company's investment in AI becomes a sustainable competitive advantage or a series of expensive experiments. As foundation models become more efficient and raw data becomes less of a differentiator, the advantage shifts to the richness, clarity, and relevance that knowledge engineering can uniquely provide. But this differentiation is only durable if knowledge engineering continuously evolves to capture the creativity and contextual nuance that frontier models increasingly commoditize. In this way, knowledge engineering becomes the dynamic engine of ROI: reducing operating costs, increasing decision speed, and ensuring that every AI action reflects ever-advancing human insight.

Knowledge engineering and its data management elements

Knowledge engineering transforms structured and unstructured data into actionable intelligence. It does this through three key components: the semantic layer, ontology, and knowledge graph.

The semantic layer

The semantic layer connects data, relationships, and meaning, which enables interpretation by both humans and machines. It serves as the foundation for building more intelligent systems and is often reflected in how information is organized and presented. For example, AI agents can use the semantic layer to interpret the meaning and relationships of sales data from multiple sources. This understanding allows them to combine data points—such as revenue, costs, and product performance—and generate charts and dashboards. It also enables tailored responses to a prompt like, "Show me profit and loss across multiple divisions with a list of top-selling products."

As companies move toward agentic AI, the semantic layer becomes the essential bridge between raw data and intelligent action. The semantic layer is not a physical database or a new data warehouse but a conceptual framework that translates technical data structures into business-friendly language and relationships. This translation is what enables both humans and AI agents to interpret, reason, and act on information with confidence.

The process begins by identifying the core business concepts that matter most, such as "customer," "product," "revenue," or "region." These concepts are mapped to the underlying data sources, whether structured (like enterprise resource planning tables) or unstructured (like emails or contracts). The semantic layer then defines how these concepts relate to one another, establishing clear meanings and connections. For example, it clarifies whether "customer" refers to an individual, a business, or a channel partner, and how "revenue" is calculated across different divisions.

A well-prepared semantic layer enables AI agents to answer nuanced questions, such as "Show me quarterly revenue by region and product line," without manual data wrangling. It also reduces ambiguity and misinterpretation, ensuring that humans and machines speak the same business language. It also accelerates the onboarding of new data sources, as each can be mapped into the existing semantic framework rather than requiring bespoke integrations.

Building the semantic layer is an iterative, collaborative effort. It requires input from business leaders, data stewards, and technical teams to ensure that definitions are accurate, relevant, and aligned with real-world operations. The goal is to create a living framework that evolves as the business grows and new data sources emerge. Imagine a retailer that uses a point of sales system with ever-changing inputs like transactions, pricing, and store data, and you get the idea.

> **The semantic layer connects data, relationships, and meaning, which enables interpretation by both humans and machines. It serves as the foundation for building more intelligent systems and is often reflected in how information is organized and presented.**

The ontology

Companies run on data with applications that were designed for human workers. AI agents in a workflow process must interact with data in a way that is machine-readable and interpretable.

That's where the ontology comes into play. The ontology defines the concepts, relationships, and rules within a specific domain, acting as a living, shared business dictionary that defines and explains the relationships of key terms, such as "division," "profit," and "top-selling product." It creates a common language so both humans and AI systems interpret terms like "customer" or "order" consistently, enabling accurate reasoning and decision-making.

More capable AI agents demand the precise definitions provided by an ontology because they reduce the risk of hallucination. Imagine an AI agent tasked with answering, "What were last quarter's sales for our premium products?" Without an ontology, the agent might misinterpret "premium" as any high-priced item or even confuse it with unrelated marketing terms found in documents. This could lead to incorrect or fabricated answers.

But with an ontology in place, "premium product" is clearly defined as a specific category in the company's product hierarchy, linked to attributes like SKU, pricing tier, and division. The ontology also defines relationships, such as "premium product belongs to product line X" and "sales figures are measured in currency Y." This structured understanding ensures the agent retrieves accurate data from the right sources instead of guessing.

More capable AI agents demand the precise definitions provided by an ontology because they reduce the risk of hallucination.

As the foundation of knowledge engineering, the ontology organizes scattered data into a coherent framework that allows AI agents to understand context, communicate effectively, and act with precision. The process unfolds through a well-defined lifecycle, each phase building on the last to ensure clarity, consistency, and long-term adaptability.

Data products

With the ontology lifecycle in mind, following a modern approach to data products is a good place to start. That's because a data product—a collection of data assets taken from raw to curated status for a business purpose—enables prioritization and unification of data. And, because the data products journey mirrors the one taken in establishing an ontology, your company can skip the foundational stage and advance to enriching data through a knowledge graph, accelerating contextualization with knowledge engineering guiding the way.

The ontology creates a common language so both humans and AI systems interpret terms like "customer" or "order" consistently, enabling accurate reasoning and decision-making.

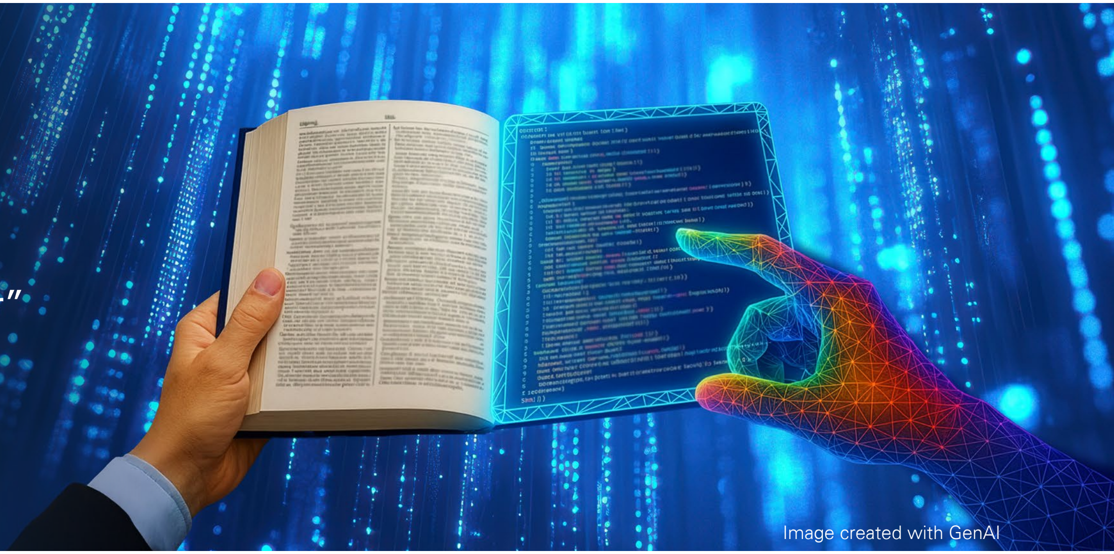


Image created with GenAI

Building an ontology starts with specification, where the organization defines its scope, purpose, and requirements. This phase sets the foundation by clarifying which business problems the ontology will solve and which domains are most critical. From there, the process moves to conceptualization, mapping key concepts and relationships into a coherent model that reflects how the enterprise truly operates. Here, subject matter experts and data strategists collaborate to capture the entities and connections that matter most, creating a blueprint for intelligent automation.

Once the conceptual model is established, it is formalized into a machine-readable language such as OWL or RDF, applying logical constraints and semantic clarity to enable precise reasoning. The next step, implementation, brings the ontology to life by integrating it with real data, knowledge graphs, and semantic layers. Rigorous evaluation follows, validating accuracy through reasoning tests and business checks, while performance assessments confirm that AI agents can deliver reliable answers. Finally, maintenance ensures the ontology remains current as business conditions evolve. Treated as a living asset, it is continuously refined to reflect new products, regulations, and market dynamics, safeguarding trust and sustaining the integrity of AI systems.

Bringing the ontology to life with enterprise digital twins

An enterprise digital twin is the dynamic representation of your organization's ontology in action. While the ontology defines relationships between data, processes, and knowledge, the digital twin visualizes how these elements interact in real business terms. It enables leaders to simulate scenarios, monitor performance, and assess productivity across functions.

For example, a CFO can see how finance processes powered by intelligent agents respond to changing conditions, revealing opportunities for efficiency and risk mitigation. By connecting semantic layers, knowledge graphs, and agent behaviors, the digital twin transforms abstract models into actionable insights—making the enterprise's knowledge framework tangible and measurable.

The knowledge graph

Once the semantic layer and ontology are in place, the knowledge graph brings them to life. A knowledge graph is a rendering of the ontology and allows agents to traverse relationships, infer new knowledge, and answer complex queries that go beyond the simplicity of data retrieval. It uses the business dictionary to connect real data points—for example, how products link to divisions, costs, and revenue. This connected view helps AI agents answer complex questions, such as comparing performance across regions or identifying trends, without manually pulling data from multiple systems.

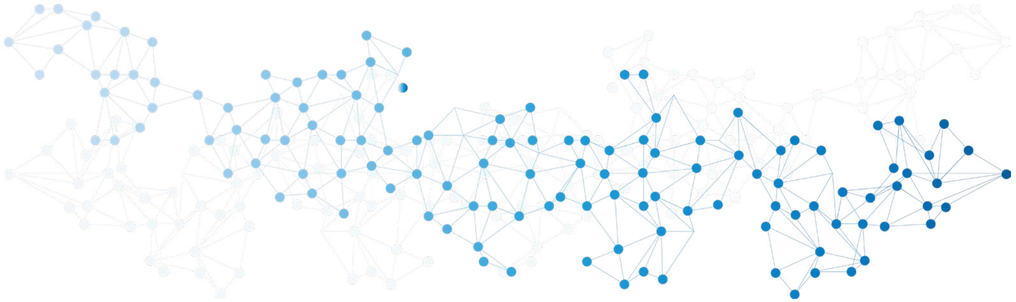
In practice, the knowledge graph ingests and organizes data from multiple sources according to the ontology's definitions and the semantic layer's mappings. Each node in the graph represents a business entity (like a customer, product, or transaction), while the edges capture relationships (such as "purchased," "belongs to," or "managed by"). This structure allows AI agents to traverse the graph, uncovering patterns and insights that would be hidden in siloed data.

For example, when an executive asks, "Which premium products drove the highest profit in the Northeast last quarter?" the knowledge graph enables the agent to:

- Identify all products classified as "premium" (as defined by the ontology).
- Link those products to sales transactions, regions, and profit figures.
- Aggregate and rank the results, even pulling in supporting visuals or documents if multimodal data is integrated.

The knowledge graph also supports advanced use cases such as:

- Detecting anomalies or emerging trends by analyzing relationships over time.
- Powering recommendation engines that suggest next-best actions for sales, service, or operations.
- Enabling explainable AI, where agents can show the reasoning path behind their answers.



Multimodal data

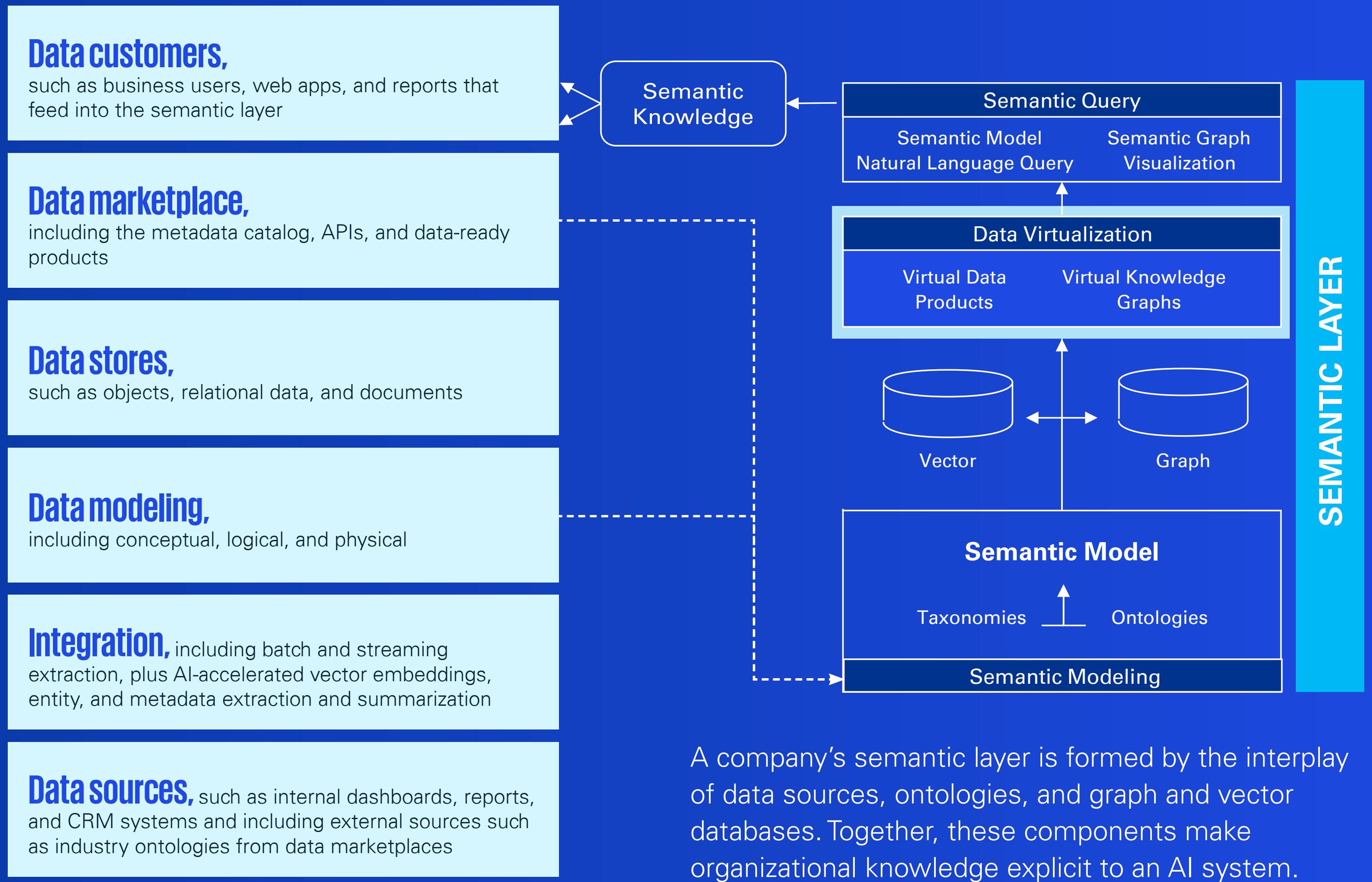
Adding multimodal data such as text, images, videos, and other unstructured formats takes knowledge engineering to the next level. When these diverse sources are integrated into the same framework, AI agents gain richer context and deeper insights. For example, instead of just reporting sales numbers, an agent could also include product images or marketing visuals to provide a more complete picture. Or in healthcare, multimodal integrates data from sources like medical imaging and clinical notes to enable a more accurate diagnosis. This capability enables richer insights, leading to faster decisions, fewer errors, and more intelligent automation.

Mapping it out

The ontology is your formal map of business meaning that defines entities, attributes, relationships, and rules (e.g., Customer → places → Order). The semantic layer applies that map to real data, translating raw fields into consistent business concepts and metrics so humans and AI agents share the same interpretation.

The vector database complements these by storing embeddings of unstructured content such as emails, chats, and documents to enable similarity search (“find decisions like this,” “surface complaints similar to these”). In operations, an AI agent grounds a query in ontology and semantic layer definitions for accuracy and compliance, then uses the vector database to pull contextually relevant, semantically similar knowledge.

> Together, the ontology provides explicit meaning, the semantic layer enforces consistent interpretation, and the vector database adds implicit meaning, building a foundation for trustworthy RAG and agentic workflows.



A company’s semantic layer is formed by the interplay of data sources, ontologies, and graph and vector databases. Together, these components make organizational knowledge explicit to an AI system.

Key recommendations

Transitioning to knowledge engineering is a strategic imperative for organizations seeking to unlock the full value of AI and data-driven decision-making. To make the shift manageable, KPMG recommends starting with a single data product or business unit; however, it's important to recognize that successful implementation demands an enterprise-wide shift in culture, governance, and operational processes. By embracing cross-disciplinary collaboration, unified governance, flexible data integration, context-driven decision-making, purposeful change management, hybrid workforce empowerment, and robust data governance, leaders can position their organizations to thrive in an agent-driven future.

1 Educate your team on the value of knowledge engineering

Knowledge engineering transforms data from a passive asset into an active driver of business outcomes, making it a strategic imperative. Educate colleagues that transitioning to knowledge engineering is a strategic change, not just a technical upgrade. Develop organizational "muscle memory" by adapting processes, repeating new workflows, and building pathways for agent integration. Leadership should communicate the vision, secure buy-in, and support teams through the transition.

Also, knowledge engineering calls for self-assessments. Ask: Does your company have the foundational expertise and the right people and tools for managing ontologies and knowledge graphs? You'll also need management buy-in and strategies for risk mitigation. Companies often make the mistake of trying to do too much. Prioritization is key with a focus on achievable, high-impact goals.

When knowledge engineering isn't needed

Agents don't always require all the elements of knowledge engineering. A chatbot that retrieves data from a database, for example, doesn't require an ontology that promotes a common language. Likewise, retrieval-augmented generation (RAG) is often enough for straightforward, fact-based queries.

RAG works by retrieving relevant documents or snippets from a knowledge source and feeding them into a language model to generate an answer. This approach is effective when the question is narrow and the answer exists in the retrieved content, such as searching for total sales in an annual report, because it avoids the complexity of reasoning across multiple relationships or domains.

Priority actions:

- Communicate the strategic vision and benefits of knowledge engineering across the organization.
- Provide training and resources to help teams adapt to new processes and technologies.
- Set measurable milestones and celebrate progress to build momentum.
- Address resistance proactively by involving stakeholders in the change journey.

2 Create measurable business impact with cross-disciplinary collaboration

To ensure your knowledge engineering investment delivers on ROI and scalability, organizations must break down silos and foster collaboration across information technology (IT), data science, business analysis, and domain expertise. IT silos no longer suffice; although IT is essential for infrastructure and security, leadership must actively involve data scientists, business analysts, and domain experts. This collaborative approach ensures that knowledge engineering models reflect true business meaning, not just technical schemas, and cultivate a corporate culture of data sharing.

Priority actions:

- Define and publish key success metrics around decision-making speed, stakeholder satisfaction, cost takeout, and ontology adoption.
- Establish cross-functional teams for knowledge engineering initiatives.
- Create regular forums for business and technical stakeholders to align on goals and requirements.
- Incentivize knowledge sharing and joint problem-solving across departments.
- Appoint collaboration champions to drive engagement and accountability.

3 Establish unified governance and accountability

Unified governance streamlines data flows, reduces inconsistencies, and ensures transparency and security for both structured and unstructured data. However, traditional data lifecycles, from creation and storage to reporting, have operated in isolation across different teams. Unifying disparate processes under a single governance framework reduces delays and inconsistencies. Aligning policies, tools, and accountability allows both structured and unstructured data to flow seamlessly into knowledge engineering systems.

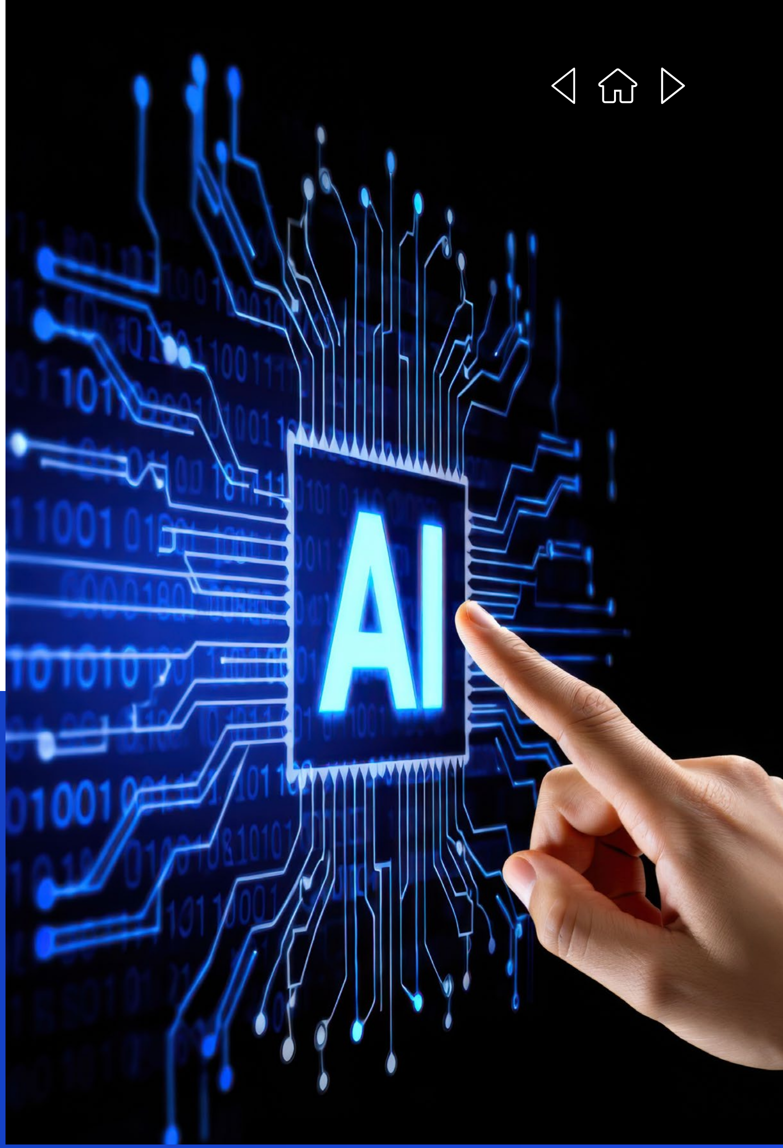
Priority actions:

- Develop a single governance framework covering all data sources and processes.
- Assign clear ownership for data stewardship and compliance (e.g., designate a chief data officer).
- Standardize policies and tools for data management across the enterprise.
- Implement regular audits to monitor adherence and effectiveness.

Prioritize context and relevance over perfection

Delivering information enriched with business context is ultimately more valuable than achieving perfect data quality in every field. Traditional approaches emphasize eliminating errors, but knowledge engineering prioritizes relevance—in other words, providing the right information, with the right context, at the right moment, just as understanding a customer's purchase history or regional trends often matters more than flawless spelling.

To achieve this, organizations should concentrate on enriching attributes that influence business decisions, establish clear guidelines for contextualizing data based on specific use cases and audiences, and equip teams to recognize when context outweighs technical precision. Continuous refinement is essential, using feedback from AI agents and business users to ensure that relevance criteria evolve alongside business needs. In this way, knowledge engineering creates the nervous system that drives trust and actionable insight across the enterprise.



How KPMG can help

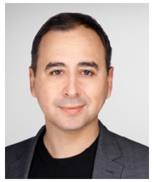
Imagine the art of the possible

In our complex business world, moving to an agentified business model has become imperative. Business must evolve from data management to knowledge engineering that can unify all data types and pair with data management tools like semantic layers, ontologies, and knowledge graphs to make agents more accurate and high-output productivity. Collaborate with KPMG to help lead this initiative. We have the skills and technology tailored to your business that can lead your teams forward with learnings and success markers.

You can win with AI

- KPMG is ranked #1 for quality AI advice and implementation in the US.
- KPMG achieved the 2025–2026 Microsoft AI Business Solutions Inner Circle award, demonstrating our sales achievement and innovation in AI business solutions.
- KPMG has been acknowledged by HFS Research for our capabilities in applying analytics, AI, data platforms, and automation to deliver value for enterprise clients.

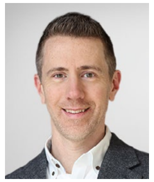
Authors



Matteo Colombo

Principal, Advisory
KPMG US
Matteo Colombo@kpmg.com

Matteo serves as the global leader for Digital Technologies at KPMG. In this pivotal role, he spearheads the growth, transformation, and innovation across the KPMG Digital Foundation and the global Centers of Excellence for Cloud, Data, and AI. Matteo plays an instrumental role in shaping the use of technology in the firm's collective strategy and helping ensure its effective implementation in local markets. Additionally, he manages relationships with key strategic technology partners. Matteo's profound experience encompasses both strategy and technology, with a particular emphasis on emerging trends that drive growth and innovation for clients and partners. He is a staunch advocate for the effective and ethical use of AI. With over 20 years of experience, Matteo has adeptly guided large enterprises in harnessing leading technologies to achieve significant, large-scale transformations.



Pete Irwin

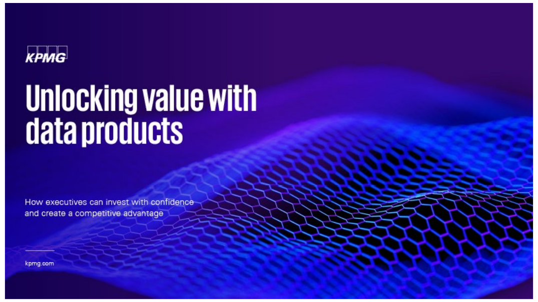
Principal, Advisory
KPMG US
pirwin@kpmg.com

Pete is the leader of the Technology & AI Engineering offering with KPMG Lighthouse, focused on helping clients modernize their application and AI capabilities and supporting software/cloud infrastructure. With over 20 years of analytics, AI, and technology consulting experience, Pete leads teams that leverage data and technology to develop and implement solutions embedding the latest AI, GenAI, and tech capabilities. Pete previously led the Data for AI offering, and helped develop the firm's AI-native data architecture approach. Other recent accomplishments include the creation of the KPMG AI Landing Zone offering, Intelligent Forecasting solution, and delivering new capabilities for multiple clients leveraging the Signals Repository and Ignite AI platform. Pete was also the leader for the KPMG Insights Center—an immersive experience designed to introduce clients to cutting-edge technology and innovations.

Contributors

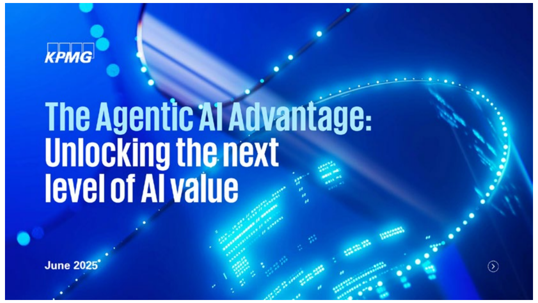
The authors thank Lisa Bigelow, Mary Lewis, Michael Ogden, Rich Schulte, and Kathy Wheeler for their contributions to this paper.

Related thought leadership



[Unlocking value with data products]

[Data governance in the age of AI]



[The agentic AI advantage: Unlocking the next level of AI value]

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

Please visit us: **kpmg.com**