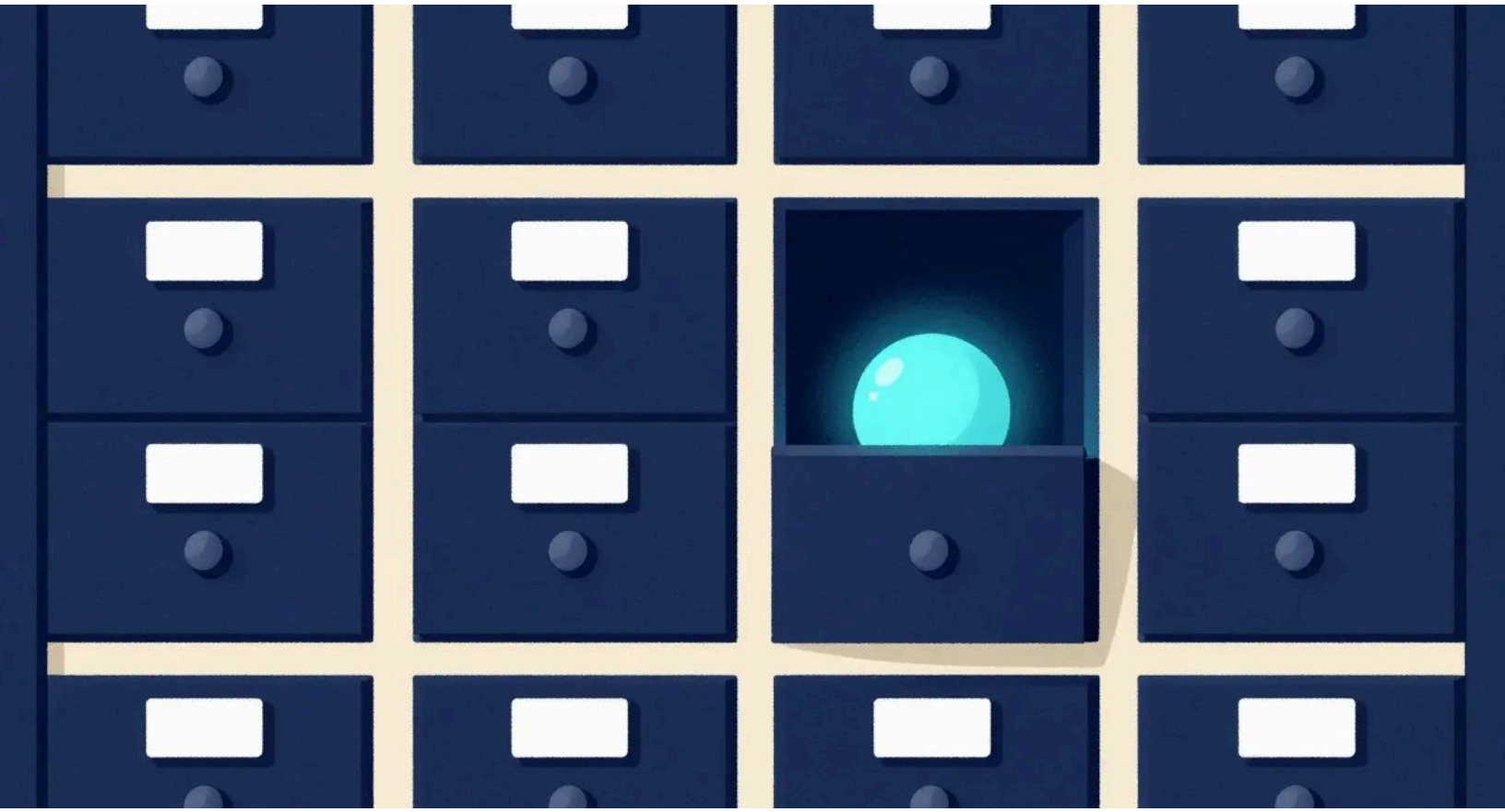


EXECUTIVE PAPER · AI

Define Your Data Before You Deploy an Agent

An AI agent reasons only as well as the company has defined its own data. Technology and data leaders who build a semantic model first get answers they can defend.

AUTHOR

Chris Watkins

SERVICE

AI

FOR

CTO, CDO · 12 min read

AT A GLANCE

The problem. AI agents answer from whatever data they can reach, and they answer with the same confidence whether they are right or wrong. When two systems define "active customer" differently, or a help page contradicts the policy it summarizes, the agent picks one and states it as fact. A Canadian tribunal held Air Canada liable in 2024 for exactly that kind of answer.

Why it persists. Business definitions live in people's heads, report code, and spreadsheet tabs. Retrieval tools index documents and tables without knowing which source is authoritative, how fields relate, or who may see them.

The recommendation. Before an agent goes live on a domain, build a semantic model for that domain: an ontology of business entities and how they relate, a classification of each source's authority and sensitivity, lineage from source to answer, and access rules enforced when the agent queries. Test the agent against that model before release, and require every answer to cite its source.

The Decision in Front of You

Once an organization decides to deploy an agent on its own data, the technical team faces a design choice. The agent can read source systems and document stores directly, with the meaning of the data explained in prompts. Or the company can put a semantic model between the agent and the data, a maintained, machine-readable description of what each piece of data means, where it came from, how trustworthy it is, and who may use it.

The first path produces a demo in days. Building the model takes weeks per domain before the first answer. This paper argues for the second path wherever a wrong answer carries a financial, legal, or customer cost.

The stakes are already visible in survey data. McKinsey's 2025 global survey found inaccuracy was the AI risk organizations most often experienced, with nearly a third of respondents reporting consequences from it [9]. Informatica's 2026 survey of 600 data leaders found half named data quality as the top challenge in deploying agentic AI [10]. Informatica sells data management software, so read that figure with its source in mind.

What's Happening

Standards treat confident wrong answers as a named risk

NIST's AI Risk Management Framework 1.0 (NIST AI 100-1), released in January 2023, organizes AI risk work into four functions: Govern, Map, Measure, and Manage [1]. NIST notes the framework is now being revised [1]. Its Generative AI Profile (NIST AI 600-1), published in July 2024, lists twelve risks specific to generative AI [2]. One is confabulation, which NIST defines as systems that "generate and confidently present erroneous or false content" [2].

The profile's suggested actions point at data definition. They include documenting the origin and history of data (GV-1.2-001), establishing practices for determining data origin and content lineage (MP-2.1-001), testing data flows from original sources through transformations (MP-2.1-002), and verifying that data used for retrieval-augmented generation is grounded (MS-2.5-005) [2]. Retrieval-augmented generation, or RAG, is the common pattern where an agent searches company content and answers from what it finds.

ISO/IEC 42001:2023, the management system standard for AI, carries a group of Annex A controls, A.7, on data for AI systems [3]. They cover data acquisition, data quality requirements for accuracy, completeness, and consistency, data provenance with a documented audit trail of sources and changes, and data preparation [3]. An organization seeking certification needs evidence of each.

Organizations are deploying agents ahead of definitions

IBM's 2025 survey of 1,700 chief data officers found 79% were still early in defining how to scale and govern AI agents [11]. Only 26% were confident they could use unstructured data, the documents and messages most agents read, to deliver business value [11]. IBM's 2026 study of technology executives recommended treating governance as an engineered system that defines what agents may do and keeps their decisions explainable [13].

A semantic layer measurably changes accuracy

The clearest public evidence comes from a 2023 benchmark by Juan Sequeda, Dean Allemang, and Bryon Jacob [7]. They asked GPT-4 43 business questions against an insurance database built from an industry-standard data model. Querying the SQL tables directly, the model answered 16.7% correctly. When the same data was described through an ontology and knowledge graph, accuracy rose to 54.2% [7]. On questions that required navigating a complex schema, direct SQL scored zero, while the knowledge graph scored between 35.7% and 38.5% [7].

A 2024 follow-up added a step that checks each generated query against the ontology's rules and repairs it [8]. Overall accuracy reached 72%, and another 8% of responses were "I don't know," leaving an error rate of 20% [8]. The authors worked for a company that sells data catalog and knowledge graph software, the benchmark is small, and models have improved since. The direction of the result is what matters: the same model gave far better answers when the meaning of the data was written down.

SIDEBAR: WHAT A SEMANTIC MODEL LOOKS LIKE

Take the phrase "active customer." In the semantic model it becomes a defined entity: an account with at least one invoiced order in the last 12 months, per the finance definition. The system of record is the ERP, and the CRM's "status" field is marked non-authoritative. The owner is the finance controller. The data is classified confidential. Sales staff may read it for their own territory, and the support agent may read only the customer's own record. Each answer that uses it cites the ERP table and the date it was last refreshed. An agent reading the CRM and ERP without this model has to choose between two conflicting answers.

Recent Cases

Air Canada, 2024. A passenger asked the airline's website chatbot about bereavement fares. It told him he could buy a full-fare ticket and apply for the reduced rate within 90 days [4]. The airline's own bereavement policy page, which the chatbot linked to, said the policy did not apply after travel [4]. In *Moffatt v. Air Canada*, 2024 BCCRT 149, decided February 14, 2024, the British Columbia Civil Resolution Tribunal rejected the airline's argument that it was not liable for its chatbot's statements [4]. The tribunal held the company responsible for all information on its website and said customers should not have to cross-check one part of a site against another [4]. It ordered C$812.02 in damages, interest, and fees [4]. The award was small. The ruling placed the error with the company even though its policy page was correct, because two company sources disagreed and nothing told the customer which one governed.

Cursor, 2025. In April 2025, users of the AI coding tool Cursor reported being logged out unexpectedly [5]. When they emailed support, an AI agent named "Sam" replied that one device per subscription was a core security policy [6]. No such policy existed [5][6]. Users announced cancellations on Reddit, and the company later apologized [5][6]. Public reports do not describe what sources the bot consulted. Its answer fits NIST's definition of confabulation: an invented policy, stated with confidence [2].

Microsoft 365 Copilot rollouts, 2026. CoreView's 2026 survey of security leaders found roughly two-thirds of organizations had delayed or canceled Copilot deployments over concerns it could expose confidential data [12]. Among those, 73% cited the risk that AI would surface confidential internal information, and the concerns centered on SharePoint permissions and sharing links [12]. The survey's sample size was not published. The worry is that an assistant reading on a user's behalf will surface anything that user's permissions allow, including content shared too widely years ago [12].

Why the Usual Approach Fails

The usual deployment points an agent at a document store or database, adds a retrieval index, and writes business rules into the prompt. Each part has a gap.

Retrieval ranks content by similarity to the question. It has no concept of which policy version is current, which table is the system of record, or which draft was superseded. A 2019 memo and the 2026 policy that replaced it can look equally relevant.

Database schemas hide meaning. A column named CUST_STAT_CD with values 1 through 7 means something specific to the people who built the reports. The model sees only the code. The benchmark above found direct SQL querying scored zero on complex-schema questions [7].

Departments define the same term differently. Finance, sales, and operations often count customers, orders, and revenue by different rules, each correct for its own purpose. An agent without a declared definition merges them.

Prompt instructions do not scale. When each agent team writes its own definitions into its own prompt, the company ends up with as many versions of "active customer" as it has agents, and none is visible to the data owners.

Access rules sit outside the agent's path. A retrieval setup that indexes content with an account able to read everything must filter results afterward, and every gap in that filter becomes a leak. The CoreView findings show executives already distrust arrangements like this [12].

The Framework: Four Parts of a Semantic Model

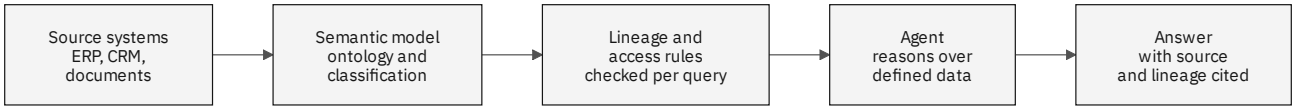
A semantic model for agents has four parts. Each answers a question the agent cannot answer for itself.

Exhibit 1. Each part of the semantic model prevents a specific kind of wrong answer

Part	Question it answers	What it contains	Failure without it	Standard reference
Ontology	What does this mean, and how does it relate to other things?	Business entities, definitions, relationships, units, valid values	Agent merges conflicting definitions or misreads coded fields	NIST AI 600-1 MP-2.1 [2]
Classification	How authoritative and how sensitive is this source?	System-of-record flags, version status, sensitivity labels, owner	Agent cites a superseded or non-authoritative source, as in Air Canada [4]	ISO/IEC 42001 A.7.4, data quality [3]
Lineage	Where did this value come from, and what changed it?	Source system, transformations, refresh time	No one can trace or defend an answer after the fact	NIST AI 600-1 GV-1.2-001, MP-2.1-002 [2]; ISO/IEC 42001 A.7.5, provenance [3]
Access rules	Who may see this, through which agent?	Policies tied to user identity and data classification, enforced at query time	Agent exposes confidential content to the wrong user [12]	NIST AI RMF 1.0 Govern function [1]

Source: author's framework, mapped to NIST AI 600-1 [2], ISO/IEC 42001:2023 [3], and NIST AI RMF 1.0 [1].

Exhibit 2. Agents query the semantic model, which resolves meaning, authority, lineage, and access before any data reaches them



Source: author's framework.

Three design rules make the model useful in practice.

The model is owned by the business. Data owners sign off definitions and authority flags for their domain. Engineers maintain the machinery, and the definitions belong to the people accountable for the numbers.

Access is enforced with the user's identity. The agent queries through the semantic layer on behalf of a specific person, and the layer applies that person's rights. Filtering after retrieval is too late.

Answers carry their evidence. Every response cites the source, its classification, and its refresh time. An agent that cannot find an authoritative source says so. In the 2024 benchmark, an explicit "I don't know" path converted errors into honest non-answers [8].

Costs, Risks, and Tradeoffs

Exhibit 3. Tradeoffs of building a semantic model before deploying agents

Area	Cost or risk	Mitigation
Modeling effort	Weeks of analyst and owner time per domain before the first agent answer	Model only the terms the first agent needs; start with 30 to 50 business terms
Scarce skills	Ontology and data modeling skills are hard to hire; IBM found 77% of data leaders struggling to fill key data roles [11]	Pair a small modeling team with business owners; train existing data analysts
Model drift	The business changes faster than the model, and stale definitions produce stale answers	Give each definition an owner and a review date; alert when a source schema changes
Added latency and complexity	A semantic layer adds a hop and a component to operate	Cache resolved definitions; monitor the layer like any production service
Over-modeling	Teams try to model the whole enterprise before shipping anything	Tie every modeled entity to a named agent or report that uses it
When this is wrong	Low-stakes uses, such as drafting internal text or searching public documentation, gain little; a wrong answer costs a reread	Use plain retrieval with a visible disclaimer and no authority to act

Source: author's analysis.

How to Act

Exhibit 4. A one-year roadmap

Window	Actions	Owner	Milestone
First 30 days	Choose the first agent's domain. List the business terms it will use. Name an owner for each source system it will read.	CDO with the business process owner	Term list of 30 to 50 entries; owners confirmed in writing
First 30 days	Write 100 test questions with correct answers, signed off by the process owner.	Process owner with data lead	Signed test set
First 90 days	Build the ontology for the domain. Mark system-of-record and sensitivity on each source. Capture lineage for every field the agent uses.	Data architect	Every in-scope field has a definition, owner, classification, and lineage
First 90 days	Enforce access through the semantic layer using the requesting user's identity. Remove all-access service accounts from retrieval.	CTO with security lead	No agent query runs under a shared all-access account
First 90 days	Run the agent against the test set with and without the semantic model. Require citations on every answer.	AI engineering lead	Measured accuracy and "I don't know" rate on the signed test set
First 365 days	Extend the model to two more domains. Set a release threshold for accuracy. Map controls to NIST AI 600-1 and ISO/IEC 42001 Annex A data controls. Review definitions quarterly.	CDO with CTO and compliance	Three domains modeled; no agent released below threshold; audit evidence on file

Source: author's recommendation.

Questions to Ask Your Team

1. For the first agent we plan to release, which source is authoritative for each fact it will state?
2. Where do our two most-used definitions of "customer" or "revenue" disagree, and who decides?
3. If the agent gives a wrong answer next month, can we trace it to the source record and the date it was loaded?
4. Does the agent query data as the requesting user, or under an account that can read everything?
5. What happens when the agent cannot find an authoritative answer?
6. What accuracy does the agent achieve on a test set our business owners signed off?
7. Which ISO/IEC 42001 or NIST AI 600-1 data controls could we evidence today?

Closing

An agent cannot know what the company means by its own data unless the company writes it down. Air Canada's tribunal held the company responsible for what its chatbot said, and the same reasoning applies to any agent speaking for a business. Build the semantic model for each domain before an agent reads it: define the terms, mark which sources govern, record lineage, and enforce access by user. Agents built on that model give answers the company can trace and defend.

Sources

[1] National Institute of Standards and Technology. "AI Risk Management Framework." Program page for NIST AI 100-1, AI RMF 1.0 (January 2023), and NIST AI 600-1. 2026. https://www.nist.gov/itl/ai-risk-management-framework.

[2] National Institute of Standards and Technology. "Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile." NIST AI 600-1. 2024. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf.

[3] ISMS.online. "Understanding ISO 42001 Annex A Control A.7: Data for AI Systems." Summary of ISO/IEC 42001:2023 Annex A controls. 2025. https://isms.online/iso-42001/annex-a-controls/a-7-data-for-ai-systems.

[4] Civil Resolution Tribunal of British Columbia. *Moffatt v. Air Canada*, 2024 BCCRT 149. February 14, 2024. https://decisions.civilresolutionbc.ca/crt/crtd/en/525448/1/document.do.

[5] AI Incident Database. "Incident 1039: Cursor support bot invents login policy." 2025. https://incidentdatabase.ai/cite/1039/.

[6] CC Startup. "Company apologizes after AI support agent invents policy that causes user uproar." April 17, 2025. https://ccstartup.com/blog/2025/04/17/company-apologizes-after-ai-support-agent-invents-policy-that-causes-user-uproar.

[7] Sequeda, J. F., Allemang, D., and Jacob, B. "A Benchmark to Understand the Role of Knowledge Graphs on Large Language Model's Accuracy for Question Answering on Enterprise SQL Databases." arXiv:2311.07509. 2023. https://arxiv.org/pdf/2311.07509v1.pdf.

[8] Allemang, D., and Sequeda, J. "Increasing the LLM Accuracy for Question Answering: Ontologies to the Rescue!" arXiv:2405.11706. 2024. https://arxiv.org/abs/2405.11706.

[9] McKinsey & Company. "The state of AI in 2025: Agents, innovation, and transformation." 2025. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2025.

[10] Informatica. "New Global CDO Report Reveals Data Governance and AI Literacy as Key Accelerators in AI Adoption." Press release via Business Wire, CDO Insights 2026. January 27, 2026. https://www.businesswire.com/news/home/20260127455879/en/New-Global-CDO-Report-Reveals-Data-Governance-and-AI-Literacy-as-Key-Accelerators-in-AI-Adoption.

[11] IBM. "IBM Study: Chief Data Officers Redefine Strategies as AI Ambitions Outpace Readiness." Press release, IBM Institute for Business Value CDO Study. November 13, 2025. https://newsroom.ibm.com/2025-11-13-ibm-study-chief-data-officers-redefine-strategies-as-ai-ambitions-outpace-readiness.

[12] Infosecurity Magazine. "Microsoft Copilot Deployments Delayed Over Security Concerns." Coverage of CoreView, "State of Microsoft 365 Security and Governance 2026." July 2026. https://infosecurity-magazine.com/news/microsoft-copilot-delayed-over.

[13] CIO Dive. Coverage of IBM Institute for Business Value, "2026 Tech Leader Study." June 2026. https://ciodive.com/news/tech-leaders-ai-deployment-underprepared/822295.

The views in this paper are the author's own and do not represent any employer. Scenarios are hypothetical. This paper is not legal, financial, or compliance advice.