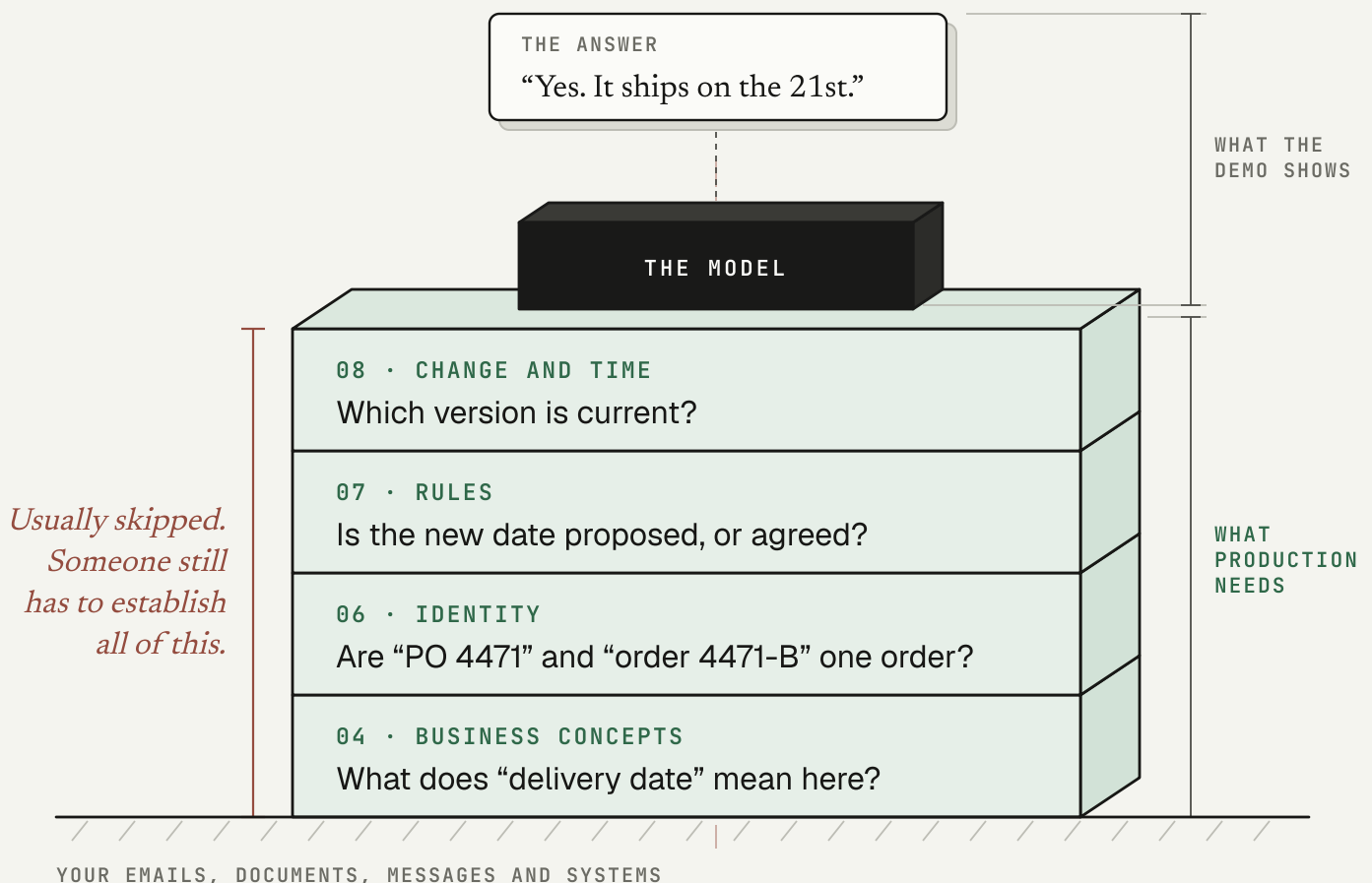


FOR THE PEOPLE WHO HAVE TO MAKE AI WORK ON REAL DATA

You connected your data to AI. *So why is it still giving you wrong answers?*

A practical guide to connecting enterprise data, understanding where retrieval breaks down, and building AI answers you can trust.



A FAMILIAR MOMENT • ILLUSTRATIVE

The answer looked right. *The order had changed.*

ASK YOUR DOCUMENTS

Can we still deliver order 4471 on the 21st?

Yes. Supplier A confirmed 9,000 L for the 21st.

SOURCE • EMAIL FROM SUPPLIER A, 6 NOV

SUPPLIER A • 6 NOV • PARAGRAPH 1

We can deliver 9,000 L by the 21st.

— *Found, and cited.*

SUPPLIER A • 6 NOV • PARAGRAPH 2

Provided the order is confirmed by the 10th.

— *In the next chunk. Never retrieved.*

SUPPLIER A • 11 NOV • REPLY

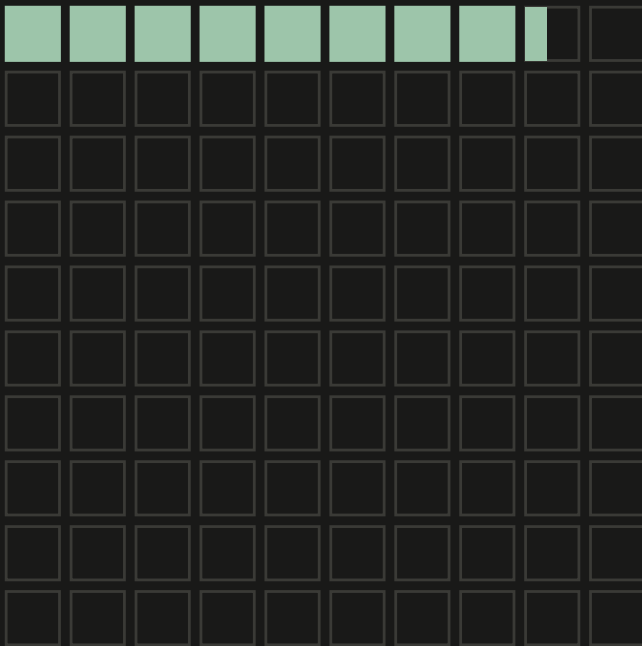
We didn't receive confirmation by the 10th. Earliest delivery is now the 24th.

— *Arrived later. Didn't rank in the top results.*

The AI did what it was built to do: it found a chunk that looked like the question, and cited it. The condition sat in the next paragraph and the change arrived five days later, so someone on your team still pieces the real answer together, every time anyone asks.

— Finding the chunk is not the same as knowing what was agreed.

The model is ready. *The data isn't.*



8.4%

of data leaders say the data feeding their AI is trustworthy enough for production.

17% → 42%

of companies abandoned most of their AI initiatives before production, in S&P Global's previous survey and then in 2025.

60%

of AI projects unsupported by AI-ready data will be abandoned through 2026, Gartner predicts.

21%

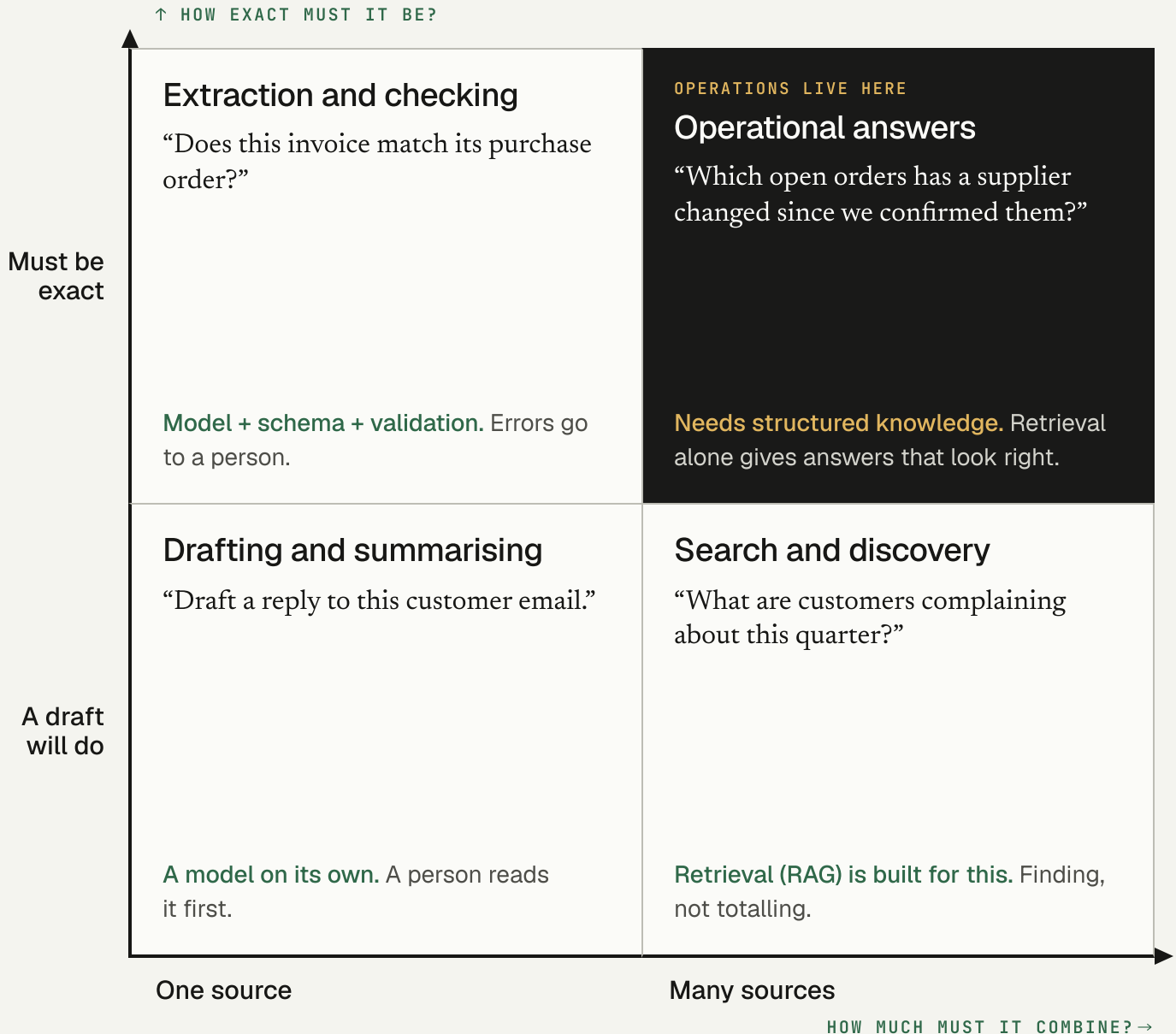
of executives rate their data curation well developed. They are over three times as likely to be effective at data and AI.

Four independent sources, one finding: adoption is racing ahead of the foundations. As a data leader told MIT CISR, “largely all the GenAI use cases are using unstructured data,” and judging its quality is “human intensive, and it’s not scalable.”

— The bottleneck is the data, not the model.

SOURCES The Modern Data Company survey, interim findings, Aug 2026 (vendor), via IT Brief • S&P Global Market Intelligence, Oct 2025 • Gartner, 26 Feb 2025 • MIT CISR, May 2026.

Two questions place *any workflow.*

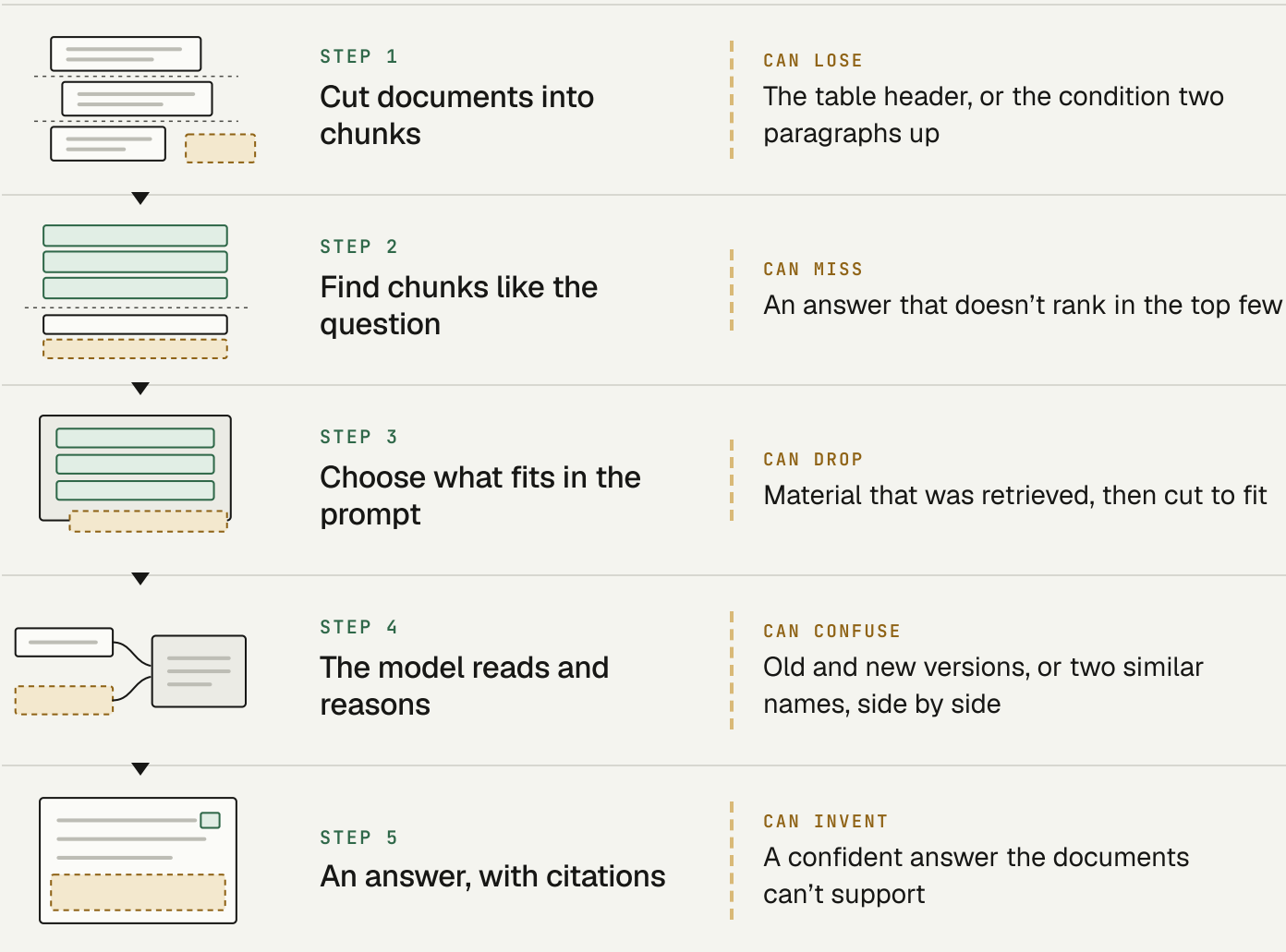


Operations are exact and cross-cutting by nature. Promising a delivery date combines a customer request, a supplier's latest email, a price list and a rule about what counts as confirmed. Each “yes” to these questions moves a workflow up and to the right: must it count every record, do names vary, can facts change, do your rules decide?

— Place the question before you choose the tool.

03 · HOW RETRIEVAL (RAG) ANSWERS A QUESTION

Retrieval finds chunks that *look like the question.*



GOOD AT

Finding relevant material fast, and drafting from it with citations a person can follow.

BUILT-IN LIMIT

“RAG fails on global questions directed at an entire text corpus,” say Microsoft’s GraphRAG researchers.

— Keep retrieval for finding. Build the structure that deciding needs.

SOURCES Barnett et al., Seven Failure Points When Engineering a RAG System, 2024 · Edge et al., Microsoft Research, From Local to Global, 2024. Dashed amber: what each step can lose.

THE FAILURES THAT MATTER IN OPERATIONS

Three ways a good-looking answer *goes wrong*.

01

The condition gets cut off

A chunk cut by length keeps the promise and loses the clause that limits it.

SUPPLIER EMAIL

“We can deliver 9,000 L by the 21st” provided the order is confirmed by Tuesday.

The promise was retrieved. The condition wasn't.

02

Similar isn't the same as agreed

A proposal, a counter-offer and the final agreement all look like the question.

THREE CHUNKS, ONE TRUTH

~~Proposed: \$412/t~~ • ~~Counter: \$398/t~~

Agreed: \$405/t

03

It can't count what it didn't retrieve

“How many” needs every record. Retrieval returns the top few, so the answer is a sample presented as a total.

RETRIEVED • 10 CHUNKS



NEVER SEEN • 12 CHANGED ORDERS



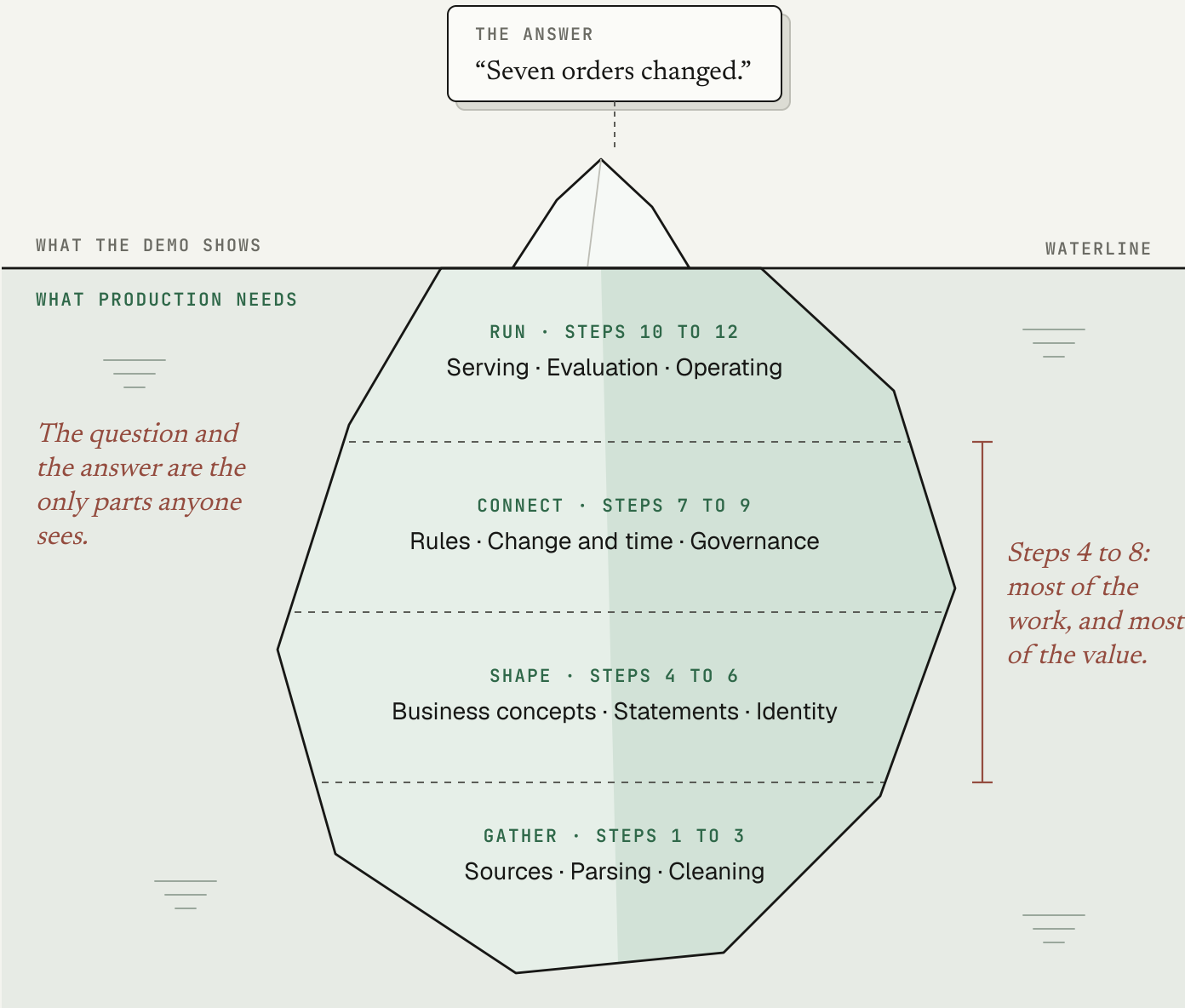
“Seven orders changed.” *There were nineteen.*

GREEN: CHANGED ORDERS IN THE CHUNKS IT SAW

Three more show up in operations: the latest message isn't always the agreed one, one thing goes by many names, and the rules that decide the answer aren't written in the documents at all.

— These are properties of the question, not bugs in a product.

The answer is the tip. *The work is underneath.*



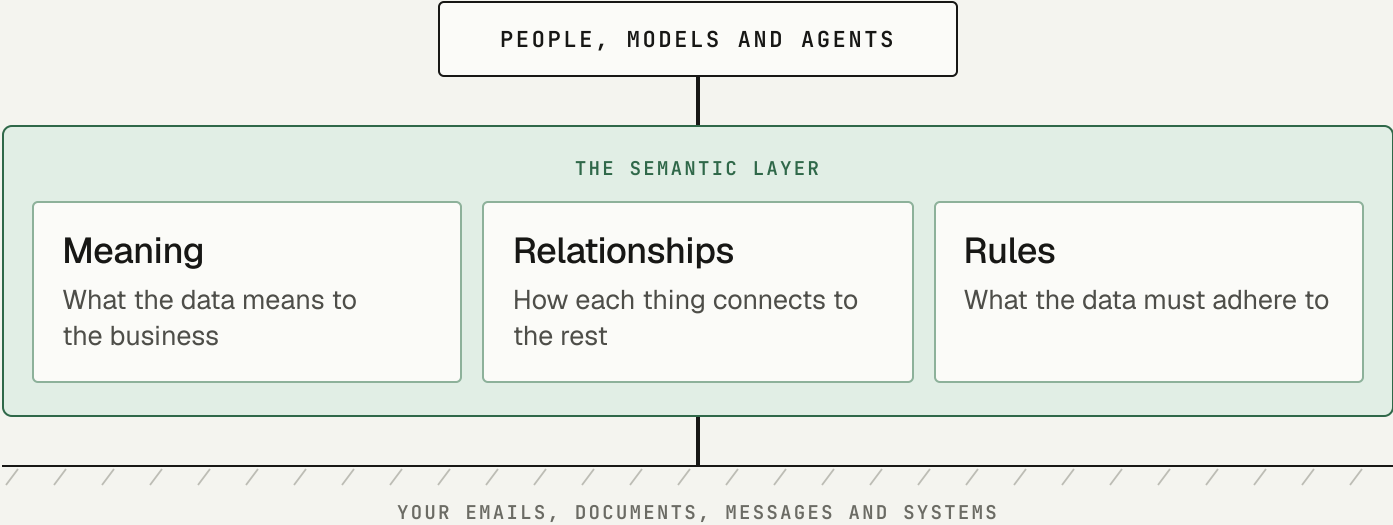
Underneath every answer you can trust sit twelve steps of work. The companies that published their engineering stories spent about a year on them: LinkedIn “over 1 year across a virtual team,” Uber “the past year”, and RAND advises committing a team to one problem “for at least a year.”

— Budget for what sits below the waterline before you buy the tip.

SOURCES LinkedIn Engineering, Practical text-to-SQL, 9 Dec 2024 · Uber Engineering, QueryGPT, 19 Sep 2024 · RAND, The Root Causes of Failure for AI Projects, 13 Aug 2024.

WHAT THE MOST ADVANCED ENTERPRISES DO

The leaders build *the layer underneath.*



MIT'S CASE • HEALTHCARE IQ • TWO HOSPITALS, TWO NAMES

Hospital A: "Sterile Gauze Pad, 4x4" = Hospital B: "SG-44-STR" = One product

Healthcare IQ supplies purchasing and pricing data to hospitals. It recognised products like these with ontologies, business rules and former procurement analysts and clinicians, over years.

80%

of the effort to onboard new hospitals, automated by its semantic technologies

MIT CISR calls it a semantic layer: “a consistent and unified representation of data from various sources that is interpretable by humans and machines.” It sits between your data and whoever uses it, and the organisations furthest ahead built it before their AI tools worked.

— Better AI starts before the model.

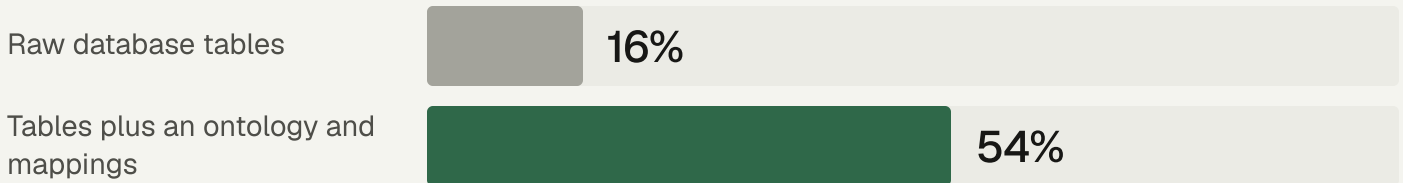
SOURCE MIT CISR, The Case for a Semantic Layer, Research Briefing XXVI-5, 21 May 2026. Healthcare IQ is MIT's case, not a Taskd customer.

THE EVIDENCE

Business context *moves accuracy.*

16% → 54%

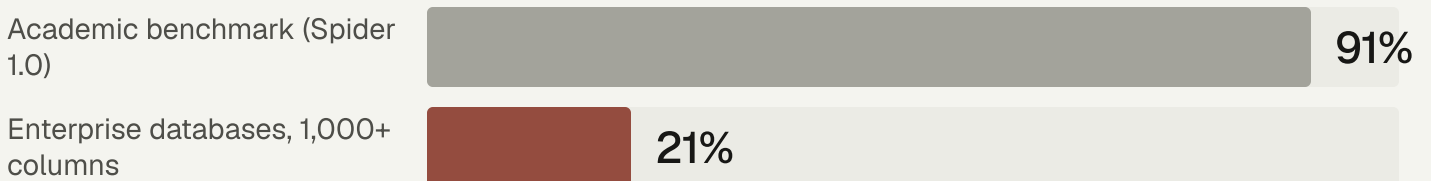
GPT-4 accuracy on enterprise questions, without and with a business model of what the data means



Same questions, same insurance database. Adding the context more than tripled the share answered correctly. Models have improved since 2023; the lesson about context hasn't.

91% → 21%

tasks solved by the best agent at publication, on an academic test and then on real enterprise databases



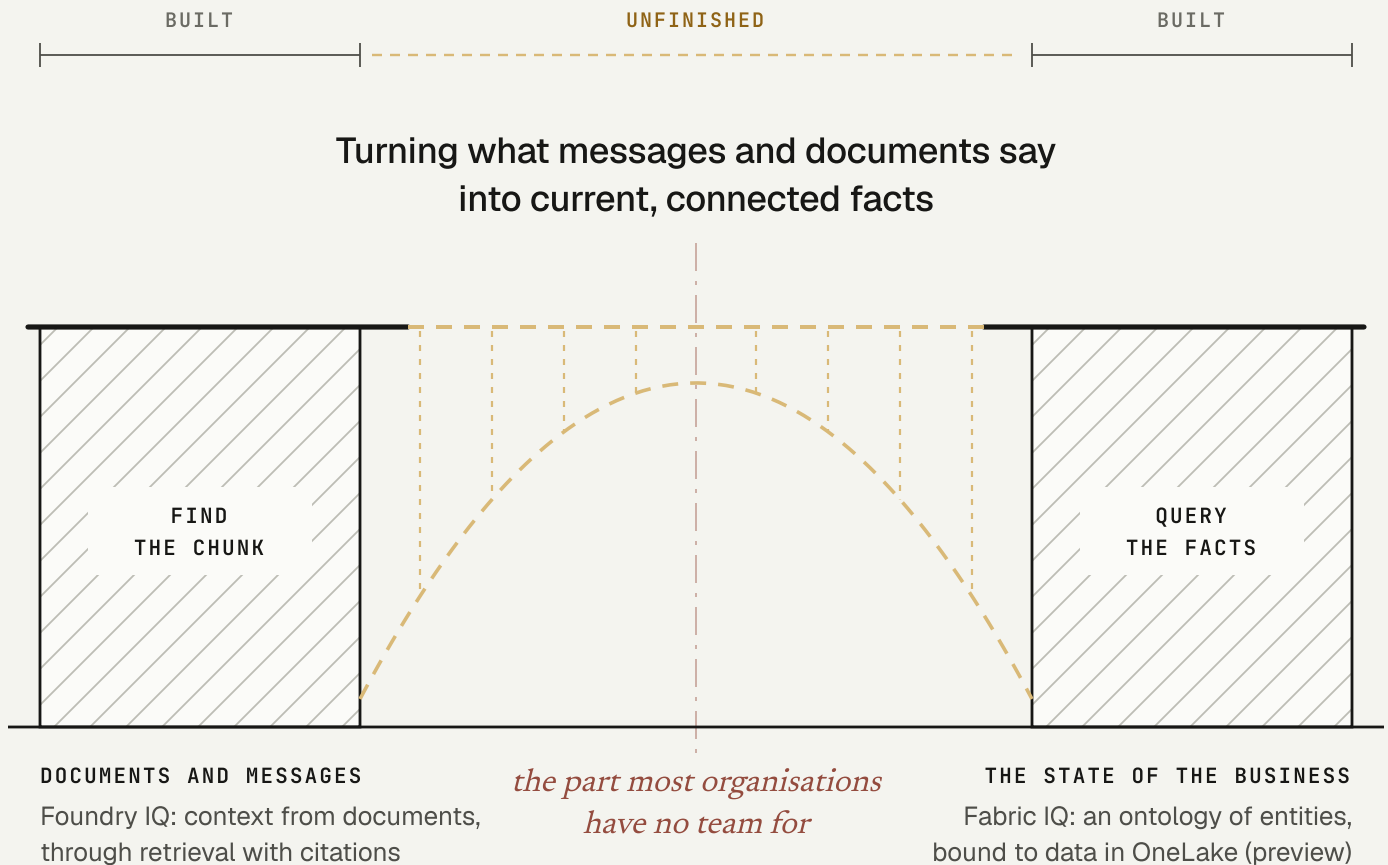
Real schemas need the model to understand what the data means, which the researchers say “frequently requires understanding and searching through database metadata.”

— Context is what moves accuracy, and somebody has to write it down.

SOURCES Sequeda, Allemang and Jacob, data.world, arXiv 2311.07509, 2023 · Lei et al., Spider 2.0, ICLR 2025, results at publication. Each bar is a share of 100%.

NOTICE WHAT'S MISSING: DOCUMENTS

From documents to facts *is the missing span.*



Today's semantic layers mostly describe tables: the Open Semantic Interchange, launched by Snowflake, Salesforce, dbt Labs and others, covers metrics and definitions over structured data. Microsoft draws the same line in its own products. The order that changed in a supplier's email sits between the two: written in a document, needed as a fact.

— That span is the semantic layer for unstructured data.

SOURCES Open Semantic Interchange announcement, Business Wire, 23 Sep 2025 · Microsoft Learn, What is Fabric IQ? and What is Foundry IQ?, read 22 Oct 2026. Fabric IQ's ontology was in preview and can be built on; the point is the job, not impossibility.

Turn every message into *facts you can check.*

SUPPLIER A • 6 NOV

“We can deliver 9,000 L by the 21st. Provided the order is confirmed by the 10th.”



WHO DID WHAT WHEN

Supplier A can deliver 9,000 L of order 4471 by 21 Nov,

CONDITION

if confirmed by 10 Nov.

SOURCE EMAIL, 6 NOV, PARAGRAPHS 1 AND 2 STATUS CONDITIONAL

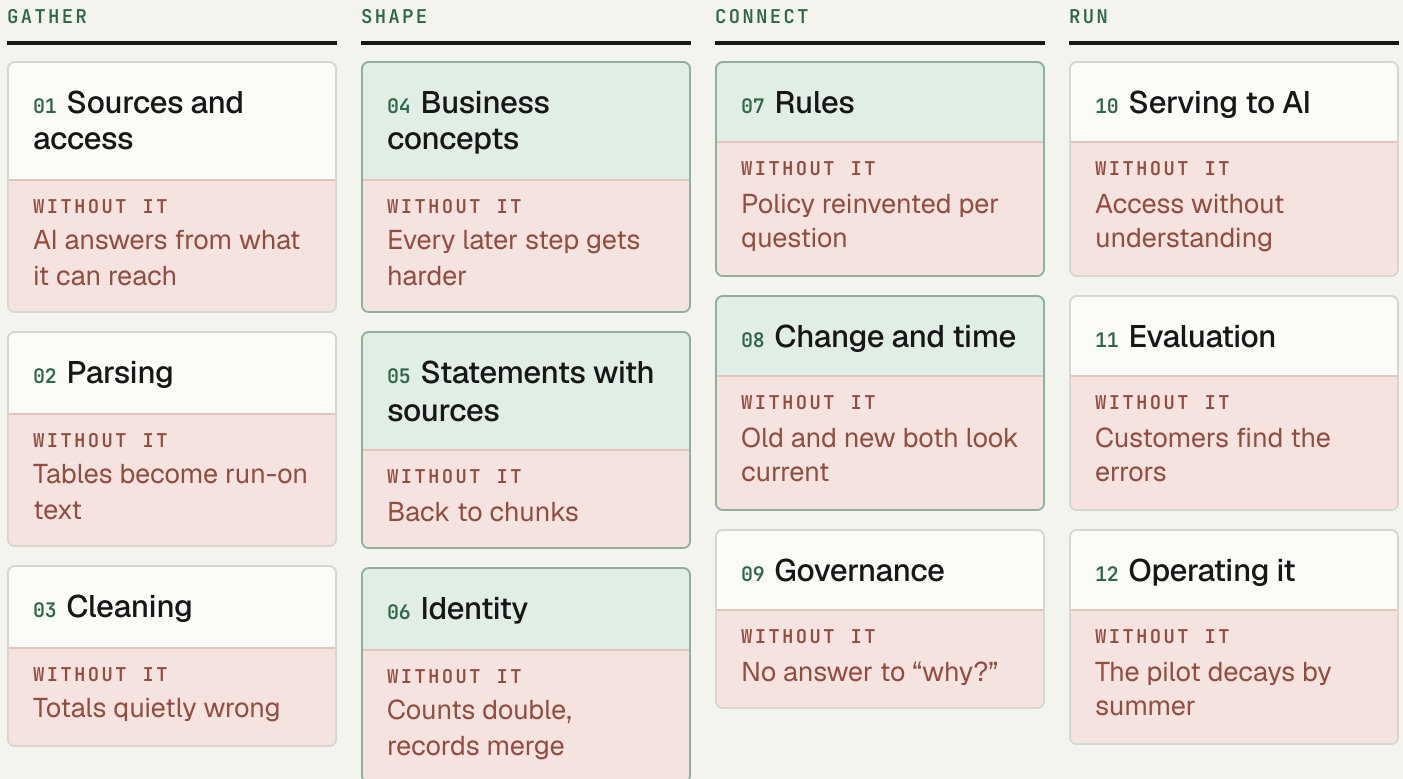
ORDER 4471 = PO 4471 = ORDER 4471-B



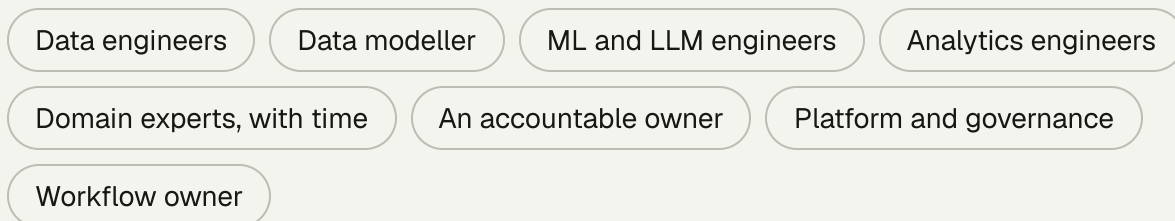
Each fact keeps who said it, about which order, when, on what condition, and the sentence it came from. Names resolve to one order. Changes keep what they replaced. The next person, or the next agent, checks the answer instead of rebuilding it.

— A shared source of truth for what was agreed, and a source of reasoning for why.

Twelve steps from raw sources *to answers you can check.*



THE TEAM IT TAKES • AT LEAST SIX KINDS OF SPECIALIST



Green marks steps 4 to 8, where most of the effort and most of the value sit, and the steps the tools on the market leave to you. Even the plumbing is a job: in a 2021 survey for Fivetran, teams reported a median of 12 data engineers spending 44% of their time on pipelines, for structured data alone.

— Connectors move data. Models read it. Neither decides what your business means by “confirmed”.

SOURCE Wakefield Research for Fivetran, The State of Data Management Report, 2021 (vendor-commissioned, 300 data leaders).

07 · SCORE ONE WORKFLOW, 0 TO 3 EACH

How ready *are you, honestly?*

01	You know every source that holds the facts it needs	0	1	2	3
02	Your AI can reach them today, with the right permissions	0	1	2	3
03	PDFs and emails parse with tables and threads intact	0	1	2	3
04	Its core things have one written definition every team accepts	0	1	2	3
05	You can tell when two names are the same thing	0	1	2	3
06	Any fact your AI uses traces back to its sentence	0	1	2	3
07	The rules that decide the answer are written where a system can apply them	0	1	2	3
08	A change records the new value, what it replaced and when	0	1	2	3
09	A named person is accountable for the quality	0	1	2	3
10	You have real questions with known right answers to test against	0	1	2	3
11	Any answer shows what it used, and who was allowed to see it	0	1	2	3
12	You have the people to build and maintain this for more than a year	0	1	2	3

0 TO 12

Pilot territory. Demos impress; production stalls. Start with steps 4 to 8 on one workflow.

13 TO 24

Tables are covered; documents, identity, rules and change aren't.

25 TO 36

A real capability. Extend it to emails and documents, and measure every answer.

— 0 = no · 1 = partly, by hand · 2 = mostly, for some sources · 3 = yes, and someone maintains it.

AI on your data is core and you can fund a standing team

Your data is mostly tables and
your questions mostly search

You have the questions and experts, not the team, and your knowledge lives in documents

☒ Steps you carry
 ☒ You decide or review
 ☐ Handled by tools or a team

SIX QUESTIONS TO ASK ANY VENDOR, INCLUDING US

- 1 Does it keep what each document says, the facts and their conditions?
 - 2 Does it know when two names mean the same thing, and when they don't?
 - 3 Are the rules written as data your experts can read and correct?
 - 4 Can every answer show its sentence, and say "no supported answer"?
 - 5 If you change model or vendor next year, what do you keep?
 - 6 What does the hundredth question cost compared with the first?

— Whoever builds it, make sure the meaning stays yours.

SOURCES A year or more per capability: LinkedIn Engineering, Dec 2024; Uber Engineering, Sep 2024.

WHERE TASKD COMES IN

You have the questions and the experts. *We bring the structure.*

Taskd builds the semantic layer for unstructured data. Steps 4 to 7 are our daily work: agreeing the business concepts and rules with your experts, extracting source-linked statements from your emails and documents, reconciling identities, and connecting the conditions and rules a question needs.

Step 8 is what we're building now: order records that update as each customer, supplier and carrier message arrives. Our engine builds connected statements with the source passages behind them, and rule execution in our logic engine is deterministic.

70,000+

emails processed in our deployed UK maritime chemicals workflow

159,000+

entities structured and 8,000+ quote requests extracted there

\$861M

in high-risk related-party funding flows flagged in 48 hours across 8.1M public-record facts

Your experts explain what matters once. Every person and every agent works from it after that.

WHERE WE'RE GOING

The maritime workflow connects incoming requests with product, pricing, transport and delivery information to prepare quotes for human approval; the operator is not named. The \$861M figure comes from the Government of Alberta AI hackathon on public records, run with deterministic rules and no LLM at runtime: a demonstration, not a deployment, and flags are prompts for expert review, not findings of wrongdoing.

THE NEXT STEP

Bring us a question your team still has to *piece together*.

Bring us a question your team still has to piece together across documents. Let's work through the answer and the evidence it needs.

1 • ONE QUESTION

One question your team still pieces together by hand, and the person accountable for its answer.

2 • YOUR CONCEPTS AND RULES

We agree the business concepts and rules with your experts, and structure the sources it needs.

3 • THE ANSWER AND ITS EVIDENCE

We work through the answer, the evidence behind it, and what it would take to run it every day.

A GOOD FIRST QUESTION

- Sits in the top right of the map: exact, and spread across many sources
- Has someone accountable who feels the cost when it is wrong
- Depends on emails, documents and records that don't agree
- Comes up every week, not once a year

Ryan Hanley
Co-founder and CEO

Alex Dollery
Co-founder and CTO

ryan@taskd.ai
taskd.ai

Better AI starts before the model.

Every figure in this edition is sourced on its page. The full 28-page guide has each source with its link, the twelve steps in detail and the schema habits that work.