The data leader's primer for agentic AI

How to build your first AI agent with fresh, centralized, governed data.

Table of contents

Executive summary

Agentic AI is undergoing the most aggressive technology adoption curve in a generation. According to the 2026 Gartner CIO and Technology Executive Survey, more than 60% of organizations plan to deploy AI agents in the next 2 years.[1] Only 17% have done so today. McKinsey estimates that agentic AI could add $2.6 trillion to $4.4 trillion in value annually across enterprise use cases.[2]

Early adopters have already successfully leveraged AI for incredible results. Retailer Wayfair built agents to support suppliers, fully automating 41,000 support tickets per month.[3] The company also used AI to automatically tag millions of products, a task that would be highly impractical for humans. Logistics giant C.H. Robinson built agents to read shipping request emails, connect information across messages and attachments, and fetch additional context, automatically creating over 5,500 shipping orders and saving 600 person-hours per day.[4]

However, there are cautionary tales, as well. Payment network Klarna became, in the words of one analyst, the "poster child for bad AI deployments" after deploying a customer service agent that was meant to perform the work of over 850 employees and save $60 million.[5] Later reporting suggested that the AI-first support push contributed to quality issues and declining customer satisfaction, turning Klarna into a cautionary tale about prematurely or excessively automating customer-facing workflows.

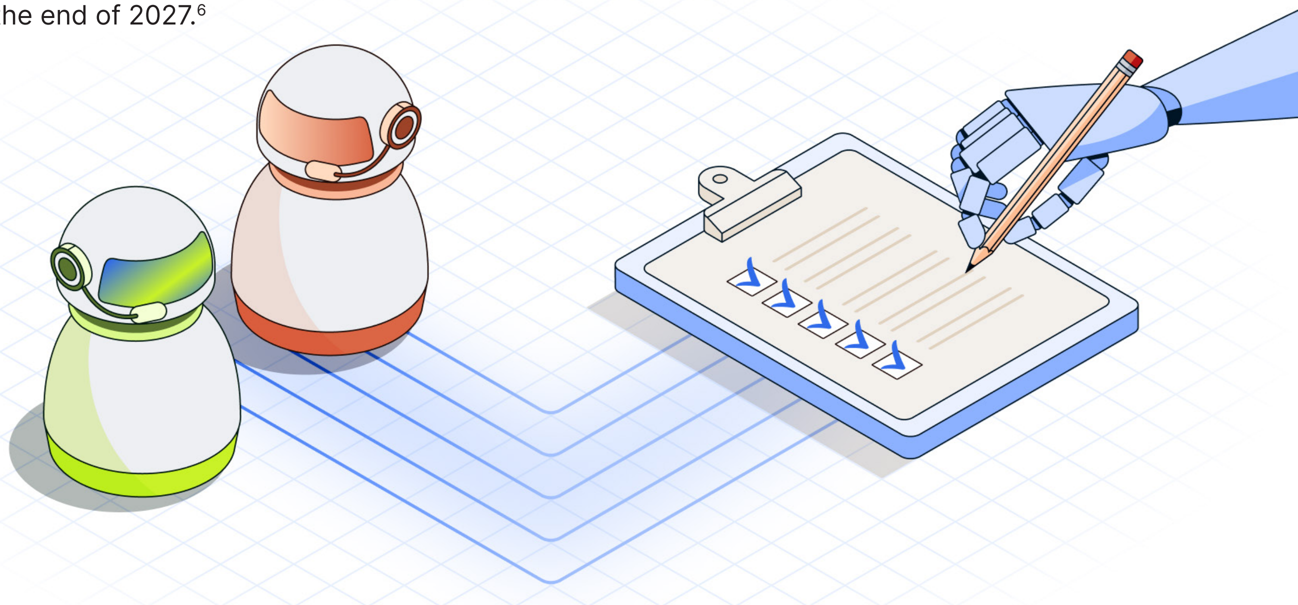
Klarna is not alone. Overall, Gartner projects that more than 40% of agentic AI projects will be canceled by the end of 2027.[6]

The key limiting factor for successful agentic AI implementation is generally not model quality, but poor data quality and governance. Frontier AI labs have released powerful foundational models. But if the data that feeds them is inaccurate, incomplete, or inconsistent, you can produce poor results. This is fundamentally an infrastructural issue that calls for:

1. **Automation and reliability** to free up technical resources for more impactful work
2. **Scalability** to enable expanded analytical and operational usage of data while limiting costs
3. **Flexibility** to handle multiple data formats
4. **Interoperability** to support multiple data use cases and complementary technologies
5. **Governance and reusable context,** including semantics, business definitions, end-to-end lineage, and metrics, to ensure trust in data

In addition, the best deployments of agentic AI are narrow, high-volume workflows where the agent can access authoritative data, produce reviewable outputs, and act within clear permission boundaries.

The costs of getting this wrong include not only costly disappointments but also degraded operations, reputational losses, and diminished competitiveness. According to a 2026 Fivetran report, only 15% of organizations are fully ready for agentic AI, even as the vast majority have already invested tens of millions.[7] A data foundation for AI is your ticket to breaking the curve.



CHAPTER 1

Why agentic AI projects fail without trusted data

Agentic AI extends the reasoning ability of generative AI to decisions and actions performed through software, not only producing information but also acting in the world.

Most practical agents combine foundation models with enterprise context through retrieval (cf. retrieval-augmented generation, or RAG), structured data access, semantic layers, and tool interfaces such as model context protocol (MCP), an emerging standard for connecting AI systems to external tools, data sources, and applications.

Agentic AI imposes far greater demands on an organization's data infrastructure than human-centric analytics workflows, especially reports and other forms of decision support.

Without adequate data and context, agentic AI can misanalyze a situation, choose the wrong response, perform the wrong action at scale, and cause cascading workflow errors. In addition, with poor security, governance, and accountability, including at the level of data assets, it can become difficult or impossible to trace the origin of a wrong decision and remediate it.

A chatbot deployed by Air Canada infamously gave a customer inaccurate information about bereavement pricing, failing to refer to and correctly cite the company's internal policies.[8] In another notable instance, Replit's software development copilot deleted a production database during a code freeze and created false data in the process — the kind of incident that hard production controls should prevent, whether the actor is human or AI.[9]

More generally, you can organize the risks posed by agentic AI into 3 major categories:

1. **Operational correctness risk:** the agent acts on stale, missing, or misunderstood data.
2. **Control-plane risk:** the agent has the wrong permissions, weak auditability, or poor security boundaries.
3. **Human-system risk:** humans overtrust, under-review, or cannot effectively supervise the agent.

Human-centric analytics	Agentic AI
Limited scalability; requires hiring and training employees	Extreme scalability; requires spawning new software instances
Intermittent data consumption	Continuous data consumption
Can learn continuously and self-teach	Can be augmented with context, but cannot truly "learn" without expensive retraining
Readily absorbs tacit, tribal knowledge through experience	Needs explicit access to context
Can intuit, remember, or investigate where a data asset came from	Requires explicit governance and access to data lineage

CHAPTER 2

Agentic AI needs infrastructure to provide data and context

The 3 failure modes for AI mean that the main limitation to effective AI isn't the power of major foundation models, but a lack of context. Specifically, your organization must be able to access all relevant data for its operations through a common data foundation. Relatedly, you must be able to scalably locate all relevant data, understand what it means and how it was modeled, and control how agents (or humans) can access it.

When agentic AI operates on stale, undocumented, or siloed data, the problem is not merely technical but an operational risk. Metrics diverge across teams, context fragments across systems, and agents act on incomplete versions of reality. As enterprises scale agents across more workflows and platforms, inefficient transformation logic can also drive compute costs that grow faster than the value those agents create.

The key challenge is the lack of a governed, consistent context across the enterprise data estate. AI agents need more than raw data — they need reliable data movement, shared business logic, semantic context, lineage, and access controls across every source and consumer.

Agentic AI depends on a solid architectural foundation for data infrastructure. The 2 key pillars are automation and centralization. You must be able to move, manage, and transform data at a much higher scale than for human consumption alone. Centralized, quality, up-to-date data is essential for ensuring that AI can access your organization's full context. Once the data is centralized, your team will need to systematically transform — i.e., model — it into a usable context layer for AI. Data transformation requires capabilities such as collaboration, version control, and testing.

It also makes lineage, semantic layers, and governance critical. Lineage shows users, including agents, where data came from and whether it can be trusted. Semantic layers apply shared business meaning to tables, granting humans and agents alike a shared, consistent understanding of how data maps to real-world business concepts. Governance controls what users and agents can access, decide, and change. Together, they turn agentic AI into an auditable, reliable enterprise workflow.

For agents, stale data is not merely an analytics problem; it can become an operational action taken on the wrong version of reality.

Centralizing data and ensuring that it's inventoried and defined once is essential for scaling access to trusted data, controlling infrastructure costs, and managing compliance risks. The use of automation at scale raises the stakes considerably. Agentic AI increases the cost of a closed, fragmented data architecture. If each AI use case requires copying data into a proprietary silo, organizations lose governance, portability, and control.

Agentic AI also depends on interoperability. As AI tooling continues to evolve, the best model, compute engine, orchestration layer, or activation channel for one workflow may not be the best for another. Interoperability gives teams the freedom to connect systems without duplicating data, rebuilding pipelines, or locking agent workflows into a single vendor. Interoperability is key to making data usable across the full range of tools, systems, and agents that need it.

The requirements necessitate a broader paradigm shift toward the **Open Data Infrastructure (ODI).**

Learn more about the Open Data Infrastructure at https://www.fivetran.com/odi

What is the Open Data Infrastructure?

The centerpiece of the ODI is an open, governed data foundation — a modern data lake — that combines scalable storage with the structure, governance, and reliability required for analytics and AI.

Modern data lakes use open table formats such as Apache Iceberg and Delta Lake, which provide data with a shared structure, metadata layer, and transaction model. Open table formats not only make data lakes behave like reliable, queryable databases but also enable multiple engines to read and write the same data reliably. This means that data lakes decouple storage from compute: organizations can store data once in inexpensive, scalable object storage and choose the best compute engine for any workload. By leveraging open standards, organizations ensure seamless interoperability across diverse tools, clouds, and AI services, effectively mitigating vendor lock-in and preserving the flexibility to select optimal systems for any given workload.

Open table formats, such as Apache Iceberg, Delta Lake, and Apache Hudi, make data in a data lake usable for analytics and AI by making it reliable, queryable across any compute engine, and less locked into a single vendor.

Your Open Data Infrastructure requires the following essential capabilities:

A central data repository — notably, a modern data lake that leverages open table formats — that can serve as a scalable, single source of truth that is interoperable with a wide range of query or compute engines, tools, or any layer of abstraction that sits between users and data

A data integration solution that reliably and automatically centralizes structured and unstructured data from all sources at scale and features:

- Fast, timely updates
- The ability to deliver data in open table formats to a data lake
- Automatic schema migration
- Reliability and the ability to quickly recover from failures

The capability to model and transform data in a collaborative, version-controlled manner without being tied to a particular vendor or proprietary format

Data governance capabilities such as:

- The ability to block and hash sensitive data before it arrives in a central repository
- Access control
- Lineage tracking, allowing teams to determine the provenance of data assets
- Data cataloging, combined with a unified semantic layer to account for all data assets and their real-world mappings
- Observability

Automation and open standards are central to efficient, reliable, and scalable data integration. Much like AI itself, automated data integration is a formidable force multiplier, freeing up engineering resources for higher-value analytical and operational activities while ensuring data is fresh and reliable for all downstream analytical uses, such as agentic AI.

On the organizational side, your team will need to be prepared with the following practices and structures:

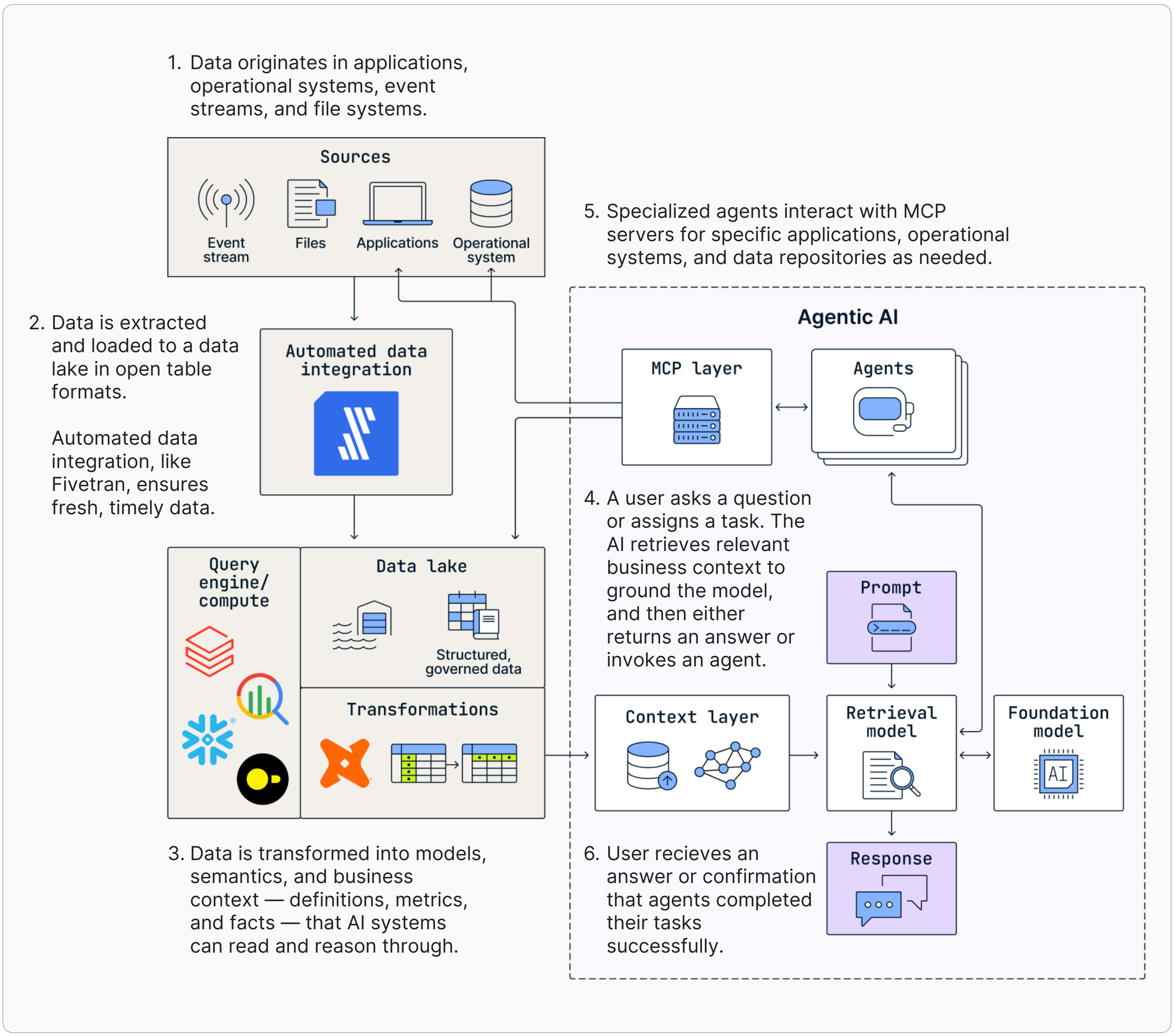
- Routine use of automation wherever possible to manage labor and infrastructure costs and efficiently produce data assets
- Data literacy to ensure that data assets are of high quality and AI-ready
- Visibility, governance, and stewardship to ensure compliance with data-related legal, regulatory, and ethical considerations
- Product development mindset and the ability to align data to AI use cases and tailor data products to the needs of stakeholders

These practices and structures demonstrate that your team effectively and responsibly handles data and is ready to pursue agentic AI.

How to build your data platform architecture for AI

Building agentic AI models from scratch is a complex undertaking that can cost many millions of dollars and months of development time. A more practical and less risky option is to augment a foundation model with your organization's unique, proprietary data using an RAG architecture, creating specialized agents to perform specific tasks by interacting with your operations through MCP servers.

The key pattern is simple: operational data flows into an open, governed foundation; transformations turn raw data into business context; agents retrieve that context and act through controlled interfaces such as MCP; outputs are logged, reviewed, and activated back into operational systems.



Initially, this architecture resembles that of other analytics use cases. You need high-performance pipelines to extract and load structured and unstructured data from a wide range of sources to a destination, namely a data lake. Teams should favor open table formats such as Apache Iceberg, Delta Lake, or Apache Hudi. Teams should perform transformations at the destination rather than in-flight, maximizing the power and flexibility of the team's compute and query engine of choice. These choices form the backbone of an Open Data Infrastructure that keeps data accessible across multiple engines, clouds, and AI services.

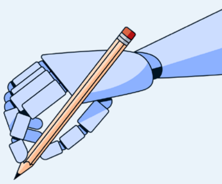
The wide range of AI use cases and rapid evolution of tooling make interoperability critical: the best systems for one particular use case may not be the best for another, and may be obsolete in a year. Building on open formats reduces the risk of vendor lock-in, preserves optionality, and makes it easier to move workloads between different tools as requirements change.

Modern tools such as Databricks Mosaic AI, Snowflake Cortex, Google Vertex AI, AWS Bedrock, and other offerings from leading cloud providers can abstract away complexity, including vector databases and other systems used to store business context in formats that AI models can read and reason over. These tools can accelerate implementation, but they should sit atop a data foundation that remains open, interoperable, and portable.

Such tools also don't absolve your team of the responsibility of context engineering. Context engineering is the practice of deliberately supplying an AI system with the data, metadata, definitions, tools, permissions, and instructions it needs to perform a task reliably. This includes not only transforming data into formats that an AI can read but also supplying it with defined metrics and rules. Context determines whether the agent understands the business, uses the right source of truth, and acts within appropriate boundaries.

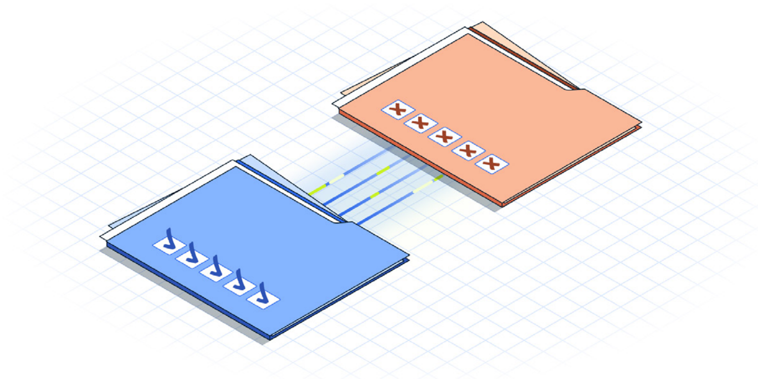
From there, most of your engineering effort is likely to go toward managing production infrastructure, context engineering, quality controlling your AI's outputs, and tuning prompts, skills, and instructions (more on that later).

AI is not conscious and lacks human-like continuity of experience, goals, feelings, or values. Under the hood, agentic AI is linear algebra, statistics, computation, and software integration on a huge scale.



CHAPTER 3

Where agentic AI works best



AI has a very "jagged" ability profile due to its design, excelling at some tasks while deficient at others. The following table lists some notable strengths and weaknesses of AI.

AI strengths	AI weaknesses
Pattern recognition and completion for text and code, including: • Drafting • Editing • Summarizing • Translation • Coding	Discerning truth from plausibility, especially when facts are obscure or highly specific, such as: • Citations • Recent events • Legal, medical, and other sensitive details
Ideation and brainstorming, especially when breadth is required, and the cost of a bad suggestion is low	Knowing when not to answer (e.g., when the premise of a request may be wrong or inappropriate, or there is no valid solution)
Reasoning through problems with clear, specified premises and constraints	Open-ended problems that require discernment, causal reasoning from limited evidence, and long-horizon planning and (especially) execution
Explaining and tutoring	Adversarial interactions

Keep in mind that agentic AI's ability to integrate with external tools — such as calculators, predictive models, databases, knowledge bases, and others — can mitigate some of these weaknesses.

In short, the best use cases for agentic AI share the following characteristics:

- High volume
- Repeatable structure
- Text/code-heavy inputs
- Clear success criteria
- Low-cost human review
- Reversible or low-risk actions
- Available authoritative data

By contrast, poor use cases for agentic AI have:

- Ambiguous accountability
- High legal or safety stakes
- Sparse data
- Adversarial users
- Long-horizon planning
- Irreversible actions

Take a close look at the workflows in your organization. Where are they bottlenecked? Which tasks are repetitive, would be impractical for a human, and involve text or code?

In 2016, Geoffrey Hinton, who would later win the 2024 Nobel Prize in Physics for his work on artificial neural networks, famously predicted that radiologists would be extinct as a profession by 2021 due to AI image recognition.[10]

By 2025, radiologists' pay, employment, and workloads had never been higher. Radiology includes many tasks beyond interpreting scans, such as advising other medical professionals, directly providing patient care, writing reports, and analyzing records.

AI is far likelier to augment complex workflows than eliminate roles.

Some practical categories of agentic AI use cases include automating:

Category	Practical example	Agentic before-and-after
Triage and alerting	Customer support escalations	From manually scanning queues to automatically flagging urgent tickets, routing them, and drafting escalation notes.
Order entry	Email-to-order processing	From rekeying PDFs and emails into ERP systems to extracting, validating, and drafting orders automatically.
Document review	Security questionnaires	From manually searching policies and prior answers to generating evidence-backed draft responses with exceptions flagged.
Unstructured data analysis	Customer feedback synthesis	From reading thousands of tickets and transcripts to clustering themes, extracting quotes, and ranking product asks.
Data enrichment	GTM prospecting	From manual account research to enriched, scored, personalized prospecting recommendations.
Conversational analytics	Executive KPI Q&A	From analyst ticket requests to natural-language answers over governed business data.

We have implemented several of these use cases at Fivetran and dbt Labs. Fivetran's Chief Product Officer uses agentic AI to perform conversational analytics on Jira data.[11] Directly querying Jira's MCP server was untenable at scale, so we moved Jira data into BigQuery via Fivetran, used a Claude Skill to query it, and produced product-ops insights in hours rather than multiple analyst sprints.

In another instance, our support team embedded a custom AI app in Zendesk to answer questions, draft responses, summarize handovers, and find similar tickets.[12] Built using Fivetran and dbt, it centralizes knowledge from Zendesk, Slab, Jira, GitHub, Google Drive, Gong, Salesforce, and docs.

Read more about this story at https://www.fivetran.com/blog/how-i-used-ai-agents-to-optimize-product-operations

CHAPTER 4

Making sure your agents are reliable

Your best chance of using agentic AI effectively is to build agents to solve specific, discrete problems rather than treating AI as a monolithic, do-everything robot.

The minimum viable workflow for usable agentic AI is something like:

1. Time- or action-based trigger
2. Gather context
3. Reason
4. Produce output
5. Ask a human-in-the-loop for approval (if applicable)
6. Write back into the system

How to systematize AI workflows — prompts, instructions, and skills

Users interface with agentic AI through natural language prompts. The key to crafting good prompts is to communicate clearly, minimize ambiguity, and provide as much useful context as possible.

Sample Prompt

THE FOLLOWING IS AN EXAMPLE OF A NAIVE PROMPT:

Which customers are at risk?

A MORE HELPFUL PROMPT PROVIDES AS MUCH CONTEXT AND DETAIL AS POSSIBLE:

Identify enterprise accounts with renewal dates in the next 90 days, declining product usage, unresolved P1/P2 support tickets, and negative sentiment in recent calls. Return account owner, evidence, likely risk driver, and recommended next action.

When generative AI first became popular, prompt engineering — the art of writing requests with as much relevant detail as necessary, and iterating to find the optimal construction — was the essential skill for AI power users. Prompt engineering has since evolved into **instructions** and **skills**, with **tools** enabling agentic AI to act through software.

Instructions contain persistent guidance about how a model should behave across conversations or tasks. They contain stable preferences, constraints, and parameters intended for broad application, such as company style guides, organizational structures, and best practices.

Skills are specific to task or workflow types and include instructions, reference files, templates, assets, and scripts. They are essentially reusable playbooks and miniature applications in their own right, specifying how a model should execute a task with a degree of exactitude.

Check out Fivetran's Skills library at https://github.com/fivetran/skills-library/tree/main

Instructions and skills can be quite detailed, sometimes running into the hundreds or thousands of words. Treat them as miniature applications that modularize agentic AI for specific tasks.

You should deploy agents in tiers of progressively growing autonomy and sophistication. Start with read-only agents that retrieve and summarize information. Then, build drafting agents that prepare outputs for human review and final implementation. Then, build bounded write-back agents that act within strict, narrow limits. Potentially risky actions by agents should require approval, while sensitive, irreversible, regulated, or safety-critical actions should remain prohibited from autonomous execution.

CHAPTER 5

Your checklist for starting with AI

To summarize everything we have covered so far, you can think of AI readiness as having the following checklist:

- ☐ **Pick the workflow:** Define the task, user, decision, and success metric.
- ☐ **Unify the data:** Use automated pipelines to centralize the sources the agent needs.
- ☐ **Govern the context:** Add lineage, quality checks, permissions, semantic definitions, and metadata.
- ☐ **Expose the right interface:** Give the agent the semantic layer, APIs, MCPs, and tools required for the job.
- ☐ **Build the agent:** Prompt, instruct, and orchestrate it against governed data and approved tools.
- ☐ **Validate and guardrail:** Test outputs, set confidence thresholds, require human review where needed, and log decisions. Validate agents against historical cases, require citations or source links, monitor false positives and false negatives, log tool calls, and review performance by workflow, not just by model benchmark.
- ☐ **Activate and reuse:** Push results into operational systems, capture feedback, and package the pattern as Skills, Instructions, or other templates.

Your AI success depends on reliable data infrastructure

AI promises a rich, innovative future in every industry, but it depends on a data foundation built on centralized, fresh, and governed data.

Every AI project starts with data. Fivetran centralizes data across a business, while dbt turns it into tested, documented, semantically meaningful context. Together, Fivetran and dbt Labs enable the data infrastructure you need to build reliable agentic AI.

Before you deploy agents, make sure they have the data foundation to act correctly.

Fivetran and dbt Labs help centralize, govern, and prepare the enterprise data agents need to work reliably.

Get started for free

Try dbt now

[1] Kandaswamy, Rajesh. (2026). 2026 Hype Cycle for Agentic AI. Gartner. https://www.gartner.com/en/articles/hype-cycle-for-agentic-ai

[2] Klein, B., Lewis, C., & Isenberg, R. (2025, October 16). Deploying agentic AI with safety and security: A playbook for technology leaders. McKinsey Quarterly. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/deploying-agentic-ai-with-safety-and-security-a-playbook-for-technology-leaders

[3] OpenAI. (2026, March 11). Wayfair boosts catalog accuracy and support speed with OpenAI. OpenAI. https://openai.com/index/wayfair/

[4] Langchain Team. (2026, March 10). How C.H. Robinson is transforming the logistics industry with LangChain. LangChain. https://www.langchain.com/blog/customers-chrobinson

[5] Doere, K. (2025, November 20). Klarna says its AI agent is doing the work of 853 employees. CX Dive. https://www.customerexperiencedive.com/news/klarna-says-ai-agent-work-853-employees/805987/

[6] Keen, Emma. (2025, June 25). Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027. Gartner. https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027

[7] Fivetran, Redpoint Insights. (2026, May 4). The 2026 agentic AI readiness index. Fivetran. https://www.fivetran.com/resources/reports/the-2026-agentic-ai-readiness-index

[8] Cecco, L. (2024, February 16). Air Canada ordered to pay customer who was misled by airline's chatbot. The Guardian. https://www.theguardian.com/world/2024/feb/16/air-canada-chatbot-lawsuit

[9] Nolan, B. (2025, July 23). An AI-powered coding tool wiped out a software company's database, then apologized for a 'catastrophic failure on my part'. Fortune. https://fortune.com/2025/07/23/ai-coding-tool-replit-wiped-database-called-it-a-catastrophic-failure/

[10] Lohr, S. (2025, May 14). Your A.I. Radiologist Will Not Be With You Soon. The New York Times. https://www.nytimes.com/2025/05/14/technology/ai-jobs-radiologists-mayo-clinic.html

[11] Kundavaram, A. (2026, May 8). How I used AI agents to optimize product operations. Fivetran. https://www.fivetran.com/blog/how-i-used-ai-agents-to-optimize-product-operations

[12] James, A. (2026, May 21). How we transformed customer support with Fivetran-enabled AI. Fivetran. https://www.fivetran.com/blog/how-we-transformed-customer-support-with-fivetran-enabled-ai