Building Enterprise AI Agents

A comprehensive blueprint for developing scalable AI agents in regulated industries, including finance, healthcare, manufacturing, energy, and government.

cohere

SECURE AI FOR BUSINESS

COHERE.COM

Table of Contents

Why we created this guide: Preparing for the age of AI agents

The enterprise AI landscape is rapidly evolving from basic generative AI (GenAI) implementations to sophisticated AI agents that can transform how organizations operate. We created this guide to empower companies as they move from experimenting with foundation models to deploying autonomous AI agents that can truly rewrite strategic playbooks and reshape business operations.

This comprehensive resource will equip you with:

- An overview of the enterprise AI evolution
- A clear understanding of the value of AI agents
- Practical insights into core use cases that deliver immediate business value
- The most popular AI agents being used today
- Common build challenges and how to address them
- Confidence to deploy AI agents into production

Whether you're leading AI adoption in your organization or building solutions firsthand, this guide is your starting point to unlock the business value of private and secure AI agents for your organization.

2023

GenAI's breakout year: First enterprise experiments

2024

Scaled deployment of task-specific GenAI solutions

2025

Rise of autonomous GenAI agents in enterprise

What are enterprise AI agents?

AI agents are advanced software programs designed to reason, access information from various sources, and execute intricate tasks across the enterprise.

Equipped with specialized tools — connections to company-wide systems — they navigate complex workflows step by step, delivering efficiency and precision.

These applications redefine how businesses work, enabling scalable operations, smarter decision-making, and seamless integration across systems.

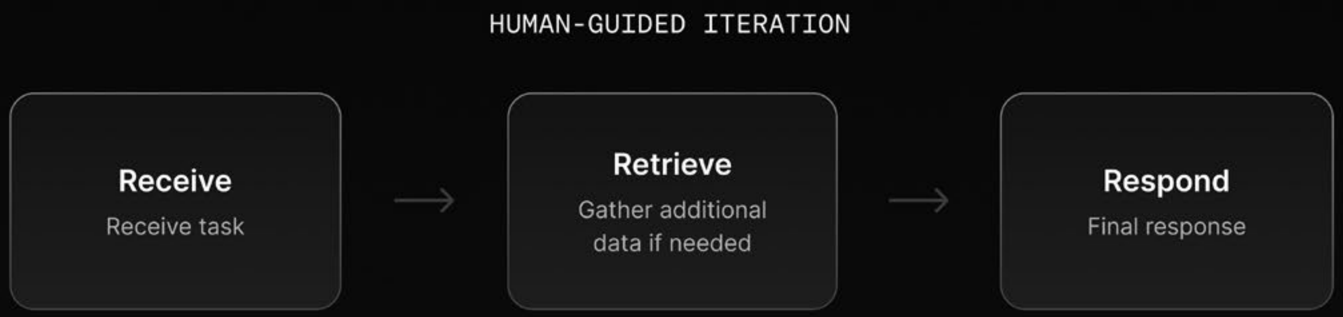
The promise of enterprise AI agents

2025 marks the rise of enterprise AI agents — advanced digital systems that are reshaping how work gets done. These AI agents don't just automate isolated tasks, they grasp context, formulate plans, and execute entire business processes. Unlike earlier AI tools that pieced together individual steps, today's AI agents drive end-to-end operations, marking a revolutionary shift in how work is accomplished.

Organizations in highly regulated and complex industries are using AI agents to reinvent workflows and create new automations. As a result, they can reap huge benefits — whether in financial services, healthcare, energy, manufacturing, or the public sector. Think about largely automating the following tasks: anticipating equipment failures, optimizing production quality, streamlining public services, monitoring financial integrity, and enhancing patient care. These industries operate under stringent regulatory scrutiny and handle vast amounts of sensitive information. This makes security a top priority, especially regarding customer data, transaction records, and proprietary insights. Private deployments, whether on-premises or in a virtual private cloud (VPC), are becoming the preferred choice for AI agents to ensure security, safety, and compliance. Within secured environments, companies are already deploying AI agents to handle everything from intricate human resources tasks to personalized customer interactions.

And this is just the beginning. By 2028, Gartner predicts that AI agents will be woven into one-third of enterprise software and shape 15% of daily business decisions. This isn't incremental progress — it's a fundamental transformation in which AI agents evolve from mere tools into trusted partners, augmenting human expertise to drive smarter decisions across the organization.

A human-led, GenAI approach to task completion



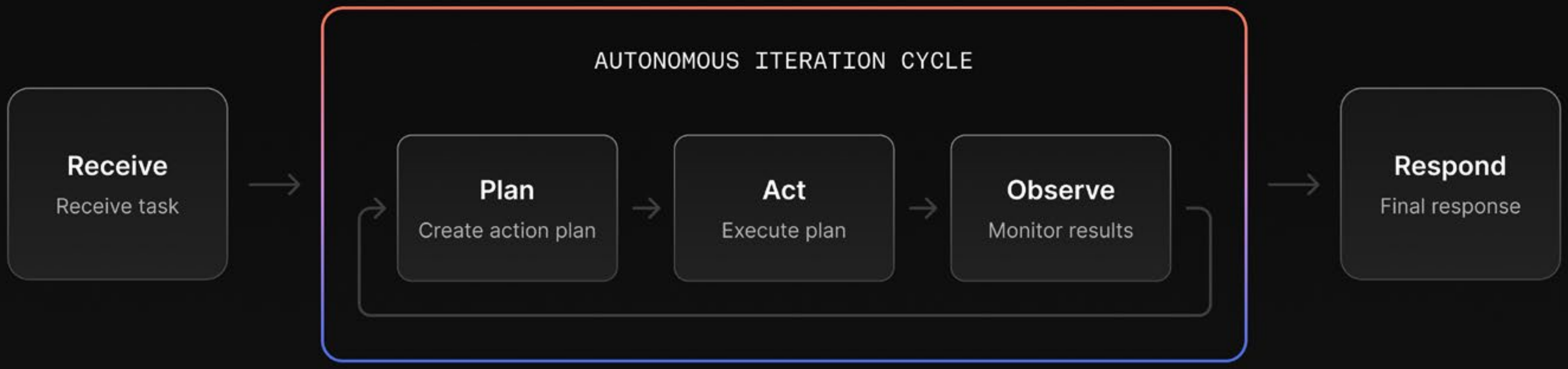
HUMAN-GUIDED ITERATION

Receive
Receive task

Retrieve
Gather additional data if needed

Respond
Final response

An agent-led, "human-like" GenAI approach to task completion



Receive
Receive task

AUTONOMOUS ITERATION CYCLE

Plan
Create action plan

Act
Execute plan

Observe
Monitor results

Respond
Final response

Image: Cambridge Centre For Alternative Finance

AI agents don't just follow predetermined paths — they reason and actively navigate your enterprise environment to accomplish goals, representing a new frontier in business automation.

What makes AI agents different from traditional automation tools?

Traditional automation tools like robotic process automation (RPA) excel at executing rigid, predefined, rules-based workflows. However, these systems require every step to be explicitly mapped out in advance, making them inflexible and unable to adapt to variations in a process. This leads to broken automations when interfaces or code (inevitably) changes. Current AI systems primarily serve as transactional partners for specific tasks like "generate a meeting brief" or "analyze this dataset."

AI agents represent a fundamental shift in this paradigm. Rather than following fixed paths or handling isolated tasks, they can navigate abstract goals and complex environments through iterative reasoning. When given a high-level objective like "provide a project update and share it with the team," an AI agent will:

- Determine the necessary steps to achieve the goal
- Adapt its approach based on available information and tools
- Navigate multiple systems while maintaining context
- Course-correct if initial attempts don't succeed
- Execute the complete workflow from start to finish

This autonomous reasoning capability transforms how enterprises approach complex work. While traditional automation and AI systems excel at clearly defined, transactional tasks, AI agents can handle abstract objectives that require understanding context, making decisions, and coordinating multiple tools and systems.

How we got to enterprise AI agents

The path to AI agents reflects a natural progression in how enterprises have adopted and matured their AI capabilities. Each phase of large language model (LLM) evolution has built upon previous innovations, enabling increasingly sophisticated systems that can handle more complex tasks with greater autonomy.

The six phases of LLM evolution to date

PHASE 1

Base language models

AI began with fundamental language models designed for text completion and prediction. These models could continue text sequences and fill in gaps, but couldn't follow specific instructions or engage in meaningful dialogue. While powerful for tasks like code completion or text prediction, they weren't suited for complex business applications.

PHASE 2

Instruction following

The next breakthrough came with models that could understand and follow specific instructions. This advancement enabled models to perform specific tasks like summarization, translation, or analysis when given clear directions. However, these models were still limited to working with their training data and couldn't access or incorporate new information.

PHASE 3

Conversational AI

Language models then evolved to engage in natural dialogue, with systems like ChatGPT demonstrating the ability to maintain context through conversations and generate contextually appropriate responses. While revolutionary for user interaction, these implementations were still constrained by their inability to access current information or company-specific knowledge.

PHASE 4

Retrieval-augmented generation (RAG)

RAG marked a significant advance by connecting language models to external knowledge sources. This enabled AI systems to ground their responses in current, company-specific information, making them practical for business use. The process was straightforward but powerful: retrieve relevant information, incorporate it into the context, and generate an informed response.

The six phases of LLM evolution to date (continued)

PHASE 5

Tool use

The next major development was enabling models to interact with external tools and APIs. Models could now not only process information, but also take specific actions through well-defined interfaces. Even so, these implementations typically required explicit instructions about which tools to use and when, limiting their autonomy.

PHASE 6

AI agents — the current frontier

AI agents represent the culmination of these capabilities combined with sophisticated reasoning abilities. Today's agents can:

- Access and process information through RAG
- Use tools and APIs to take action
- Plan and execute multi-step processes
- Reason about their approach and adjust as needed
- Maintain context across complex workflows

For example, an AI agent handling customer support can analyze incoming tickets, search knowledge bases for relevant information, route tickets to appropriate departments, update tracking systems, and draft initial responses — all while adapting its approach based on the specific situation and available tools. This combination of capabilities makes AI agents particularly powerful for transforming enterprise operations.

Enterprise AI agents for regulated industries

In highly regulated and complex sectors, such as finance, healthcare, manufacturing, energy, and government, autonomous AI agents promise to reinvent and transform operations while enabling uncompromised security and compliance. These industries handle vast amounts of confidential and sensitive information while adapting to a shifting regulatory landscape. Secure enterprise AI agents provide additional operational advantages with more control and customization opportunities designed and tailored to the unique demands of regulated industries.

The most secure approach for production is with private deployments. A private deployment ensures that sensitive data remains within an organization's controlled environment, and significantly reduces risks associated with data transmission.

Key benefits of private deployment include:

- **Enhanced security and compliance:** Keep sensitive data on-premises or within a VPC to minimize exposure and data sharing.
- **Tailored performance:** Customize models for domain-specific workflows, ensuring reliable performance in high-stakes environments.
- **Competitive advantage:** Mitigate market risks with proprietary AI solutions that help deliver strategic, long-term wins.
- **Improved auditability and transparency:** Operate within a controlled environment to enable comprehensive logging and monitoring. This level of auditability not only supports compliance with strict regulatory standards, but it also facilitates transparent reviews of AI-powered processes.

- **Long-term cost efficiency:** Avoid potential costs associated with non-compliance penalties, data breach remediation, and operational disruptions, delivering long-term savings and strategic value.

By combining autonomous reasoning with robust security measures, private AI agents not only streamline complex workflows — like real-time market analysis in finance or sensitive data handling in healthcare — but they also boost human productivity.

Industry use cases for AI agents

AI agents offer regulated industries significant operational advantages and growth opportunities. Historically constrained by compliance requirements, legacy infrastructure, and risk-averse cultures, these industries are well-placed to drive transformational change with agentic AI — and overcome many long-established challenges in the process.

Some prominent use cases include:

- **Predictive maintenance and analysis:** Asset-intensive industries like manufacturing and oil and gas can leverage AI agents to help reduce downtime and prevent disruptions by identifying potential equipment failures before they occur. AI agents are especially well-suited to analyze unstructured data, such as maintenance logs, operator notes, and design documents, and to identify potential equipment failures, prioritize maintenance tasks based on field observations, and conduct root cause analysis. By leveraging insights, AI agents can reduce downtime, improve safety, and enhance resource allocation.

- **Improved services:** Efficient, accurate services are critical in highly regulated sectors like healthcare and government. AI agents can help by improving accessibility, reducing administration load, and personalizing experiences. For example, in healthcare, AI agents can streamline patient interactions, help patients stick to treatment plans, and automate coding and billing. Similarly, government agencies can reduce administrative burden with automated document processing and tailored citizen support.
- **Safety and risk management:** Agentic AI risk monitoring and management is poised to generate tremendous benefits, particularly for financial institutions and insurance companies. AI agents can analyze and identify anomalies, leading to better and faster fraud detection. By processing multimodal financial data, underwriting agents can enhance risk protection and lending practices.
- **Enhanced decision-making:** AI agents can aid leaders and decision makers at all levels of an organization with more accessible, data-driven insights. Accessing company-wide knowledge and datastores, AI agents can process vast amounts of structured and unstructured data in multiple languages, and deliver the latest information in just one click.

AI agents aren't just an efficiency tool. They represent a transformational shift in how regulated industries can operate.

AI agent examples by industry

For inspiration, here are some AI agents — and the cookbooks to build them — that leading companies are already using.

Financial services

ADVANCED FINANCIAL DECISION-MAKING

Using agents to analyze various factors and generate reports, enterprises can automate financial, operational, and tabular data analysis. As an example, using a knowledge agent built with Cohere Command, Accenture empowers its finance and treasury teams to detect financial variance and make better decisions.

→ Get started: build a financial AI agent

Public sector

AUTOMATED DOCUMENT PROCESSING

To reduce manual efforts and accelerate document workflows, governments and public sector organizations can use AI agents to automatically create, review, and approve documents, such as contracts, reports, and compliance paperwork. AI agents can extract information from — and ask questions within — documents.

→ Get started: build a multi-step PDF extractor

Healthcare

AUTOMATED CARE SUPPORT

By facilitating seamless communication between APIs and integrating with existing tools, AI agents can enhance customer care support. Imagine automatically booking and responding to patient questions about upcoming appointments.

→ Get started: build an automated workflow

Manufacturing and energy

REAL-TIME TRACKING AGENT

AI agents can now interact with SQL databases using natural language, enabling users to query complex data without SQL expertise. This cookbook provides guidance on building an analytics agent that can retrieve production metrics, analyze performance, and track product progress.

→ Get started: build a tracking agent

Addressing five common build and implementation challenges

From managing tool integration and ensuring structured reasoning to maintaining context and preventing hallucinations, building a robust AI agent requires careful planning and implementation. As AI agents become increasingly autonomous, capable of interacting with their environments and making decisions that carry real legal and societal consequences, managing these systems requires thoughtful strategies and cross-functional collaboration across the enterprise to ensure transparency, accountability, and reliability — especially as their ability to learn, adapt, and operate in unstructured or unpredictable situations grows.

Below we explore five of the most common hurdles developers face when creating AI agents — along with practical solutions to overcome them. This is not an exhaustive list. To build secure enterprise AI agents, developers will also need to ensure they adhere to the principles of the software development lifecycle (SDLC) and secure coding best practices. This includes proper testing, code reviews, version control, and security measures to protect against potential vulnerabilities. For guidance on secure deployment of large language models (LLMs), please refer to our AI Security Guide on Deploying LLMs.

1. Managing tool integration

As AI agents become more sophisticated, managing their access to, and their use of, various tools becomes increasingly complex. Each additional tool introduces new potential points of failure, security considerations, and performance implications. Ensuring that agents use tools appropriately and handle tool failures gracefully is crucial for reliable operations.

To address this challenge, create precise definitions for each tool in your agent's toolkit. Include clear examples of when to use the tool, valid parameter ranges, and expected outputs. Build validation layers that enforce these specifications, and start with a small set of well-defined tools rather than many loosely defined ones. Regular monitoring will help you identify which tools are most effective and where definitions need refinement.

2. Managing model reasoning and decision-making

A fundamental challenge in building AI agents is ensuring consistent and reliable decision-making. Unlike traditional software systems that follow explicit rules, AI agents must interpret user intent, reason about complex problems, and ultimately make decisions based on probability distributions. This non-deterministic nature makes it difficult to predict and control how agents will respond across different scenarios, especially in complex business environments.

To address this challenge, your organization can implement structured prompting approaches like ReAct, which provides a framework for systematic reasoning. Combining this with clear guardrails and validation checkpoints helps ensure reliable outputs.

LLM temperature settings play a key role in shaping a model's reasoning and creativity. These parameters control the randomness of the text that LLMs generate. Lower settings (close to 0) ensure precise, reliable outputs, while higher values (up to 1) introduce more variability and creativity. increasing the number between 0 and up to 1 allows for increased randomness in the next word selected (and, in turn, creativity). Experimenting with different settings can help you strike the right balance between creative problem-solving and predictable results. Based on our experience, a temperature between 0 and 0.3 works best for AI agent model calls.

3. Handling multi-step processes and context

Complex enterprise workflows often require agents to maintain context across multiple steps and interactions. As these processes become more intricate, it's increasingly challenging to manage state, handle errors, and maintain coherent context. Agents must track progress, understand dependencies between steps, and gracefully handle interruptions or failures at any point in the process.

The solution is to implement robust state management systems and clear validation checkpoints throughout multi-step processes. Build comprehensive error handling for each step of complex workflows, and design fallback mechanisms for when agents encounter unexpected situations.

For example, say a predictive maintenance agent is pulling equipment failure reports, and it tries three failure log databases in sequence. If all fail, it searches for recent equipment failure reports from the last 30 days before finally routing to an operator for manual review. If at any point the agent encounters results in an unexpected format, it would immediately route those results to the operator for review.

In addition to ensuring error handling and fallbacks, be sure to clearly document process flows and implement logging systems to track the progression of multi-step tasks. This structured approach ensures that agents can maintain context and recover from interruptions effectively.

4. Controlling hallucinations and accuracy

AI agents can sometimes generate plausible but incorrect information, particularly when dealing with complex queries or incomplete data. These hallucinations pose a significant risk in enterprise or public sector environments where accuracy is crucial. The challenge is particularly acute when agents are making decisions that impact business operations, customer interactions, or citizen services.

Combat this by implementing rigorous validation systems, leveraging grounding and citations, and using structured data formats like JSON to constrain responses. Embed human review processes for critical decisions, and create comprehensive testing suites to catch potential hallucinations. Regular monitoring and logging of agent outputs can help identify patterns of inaccuracy and suggest improvements to the system. Consider implementing confidence scores and establishing thresholds for when to escalate to human review.

5. Ensuring performance at scale

Running complex AI agents in high-traffic production environments introduces a new class of engineering and operational challenges that aren't apparent during development or initial deployment. Cascading failures from tool timeouts and failures, incorrect responses, and resource bottlenecks from model serving and inference can quickly degrade system performance as request volumes increase.

Address these challenges by:

- Implementing robust error handling at every tool integration point, with circuit breakers to prevent cascading failures.
- Building retry mechanisms with exponential backoff for failed tool calls, and maintaining a response cache to reduce duplicate model calls.
- Implementing a queue management system that controls the rate of model calls and tool usage for handling concurrent requests.
- Setting up LLMOps and other monitoring tools specifically focused on catching common failure patterns by tracking tool timeout rates, model response accuracy at scale, and system latency under load. This data will help you identify bottlenecks before they impact users, and adjust your rate limits and scaling policies accordingly.

Conclusion

Get started with agentic AI

AI agents represent a fundamental shift in how organizations can leverage artificial intelligence. By combining process automation with the ability to adapt to context and take action, secure enterprise AI agents can transform business operations across regulated and complex industries that demand more control and customization.

As you begin your journey with AI agents, keep in mind these essential takeaways from this guide:

- Success with AI agents starts with clear purpose and scope. Begin with well-defined use cases that deliver immediate business value.
- Tool engineering is as crucial as prompt engineering. Invest time in crafting precise tool definitions that help your agent make better decisions.
- Security and compliance cannot be afterthoughts. Build them into your agent architecture from the beginning and consider private deployment for sensitive operations, so you gain complete control over data and model behavior.
- Start simple and scale gradually. Test thoroughly in controlled environments before expanding to more complex workflows.
- Focus on robust error handling and monitoring. Agents need clear fallback procedures and comprehensive logging for production reliability.

We hope this guide has equipped you with a practical framework and actionable recommendations to confidently build secure, production-ready AI agents. The path to success lies in thoughtful implementation, robust security measures, and a clear understanding of both the potential and limitations of AI agents.

With these tools in hand, you're ready to spark transformative innovation and lead your industry into an exciting new era of possibility. The future is yours to shape — let's build it together!

Ready to get started?

Contact our team to learn how Cohere can help bring enterprise-grade AI agents to your organization.

Additional resources

To take a deeper dive into agents and RAG, and learn how to put them to use in your organization, check out these free resources:

- **Tutorial:** Agentic RAG, offering a six-step guide to building an agentic RAG system
- **Cookbook:** Agentic Multi-Step RAG, showing how to build simple agentic RAG using Cohere's native API
- **Cookbook:** Agentic RAG for PDFs, walking through best practices for setting up a RAG pipeline to process documents that contain both tables and text
- **LLM University module:** Automate tasks and workflows, leveraging the tool use capabilities of Command R+
- **Blog post:** Connect enterprise datastores to Command with build-your-own connectors
- **Blog post:** Seven essential resources and skills companies need to build AI agents
- **Blog post:** Understand agentic AI, including the broader market context
- **Webinar:** Learn how to build an HR agent with Cohere expert David Stewart in conversation with Borderless AI engineering leadership
- **Video:** Explore the latest thinking on evaluations for AI agents with Cohere expert Jay Alammar in conversation with Graham Neubig, an associate professor at Carnegie Mellon studying natural language processing and machine learning
- **eBook:** Learn how to build secure AI solutions with this guide on best practices

About the authors

Matt Koscak

A Solutions Architect at Cohere, focusing on helping companies across diverse industries implement AI agents using Cohere's advanced retrieval and generative models. Beyond his technical work, Matt enjoys being active and socializing with friends and family.

Johnny Nguyen

An Applied Technologist with more than twenty-five years of experience spanning software development, enterprise architecture, blockchain, machine learning, and generative AI. When not building jupyter notebooks for his customers, he is a lifelong learner and enjoys cooking, golfing, fishing, and practicing jiujitsu in his spare time.

Mitchell Wong

A Solutions Architect at Cohere who focuses on deeply understanding customer goals and tech stacks. Previously, he worked in research, data analytics, and ML roles across finance, manufacturing, and e-commerce. When not helping customers analyze test results, he enjoys travelling and hiking.

Acknowledgements

A special thanks to our contributors David Stewart and Maxime Voisin for providing additional context and expertise on the subject of AI agents.

About Cohere

Cohere is the all-in-one platform for private and secure AI. Cohere brings you cutting-edge multilingual models, advanced retrieval, and an AI workspace tailored for the modern enterprise — all within a single, secure platform.

For more, visit us at cohere.com

Getting in touch today!

Ready to put AI to work? Request a demo and see how Cohere's secure and private AI platform can unlock productivity for your business.