

Guide

A Compact Guide to AI Agents



Contents

Introduction3

What Is an AI Agent?5

What Is an AI Agent System?7

What Are Use Cases for an AI Agent System?9

Building an AI Agent System 11

Conclusion and General Guidelines 15

What’s the Next Step If I Want to Start Building AI Agent Systems? 15

Introduction

Scaling GenAI apps into production using AI agent systems

Generative AI holds enormous promise, but for many organizations, transitioning from pilot projects to fully deployed applications remains a significant challenge. According to a recent [Economist report](#), “Unlocking Enterprise AI,” 85% of global enterprises are already using GenAI, a number expected to reach 99% by 2027. However, many organizations face challenges in scaling these projects effectively. The report also states that only 22% of enterprises are confident their infrastructure is ready for AI, and just 37% believe their GenAI models are truly production-ready. These gaps highlight the need for a robust infrastructure and advanced tools to ensure GenAI applications meet the high standards required for success.

Many GenAI projects fall short by failing to integrate with enterprise data, which can lead to inaccurate or irrelevant results. To fully unlock the potential of GenAI, businesses need more than stand-alone models — they need comprehensive AI agent systems that are tailored to their specific data and business needs.

Databricks Mosaic AI: Building high-quality, scalable agent systems

Databricks Mosaic AI empowers organizations to build and deploy high-quality AI agent systems. Built on lakehouse architecture, Mosaic AI allows secure customization with enterprise data, ensuring accurate, domain-specific outputs. It offers a secure way to connect to open source or commercial models, providing the flexibility to choose the best-fit solutions.

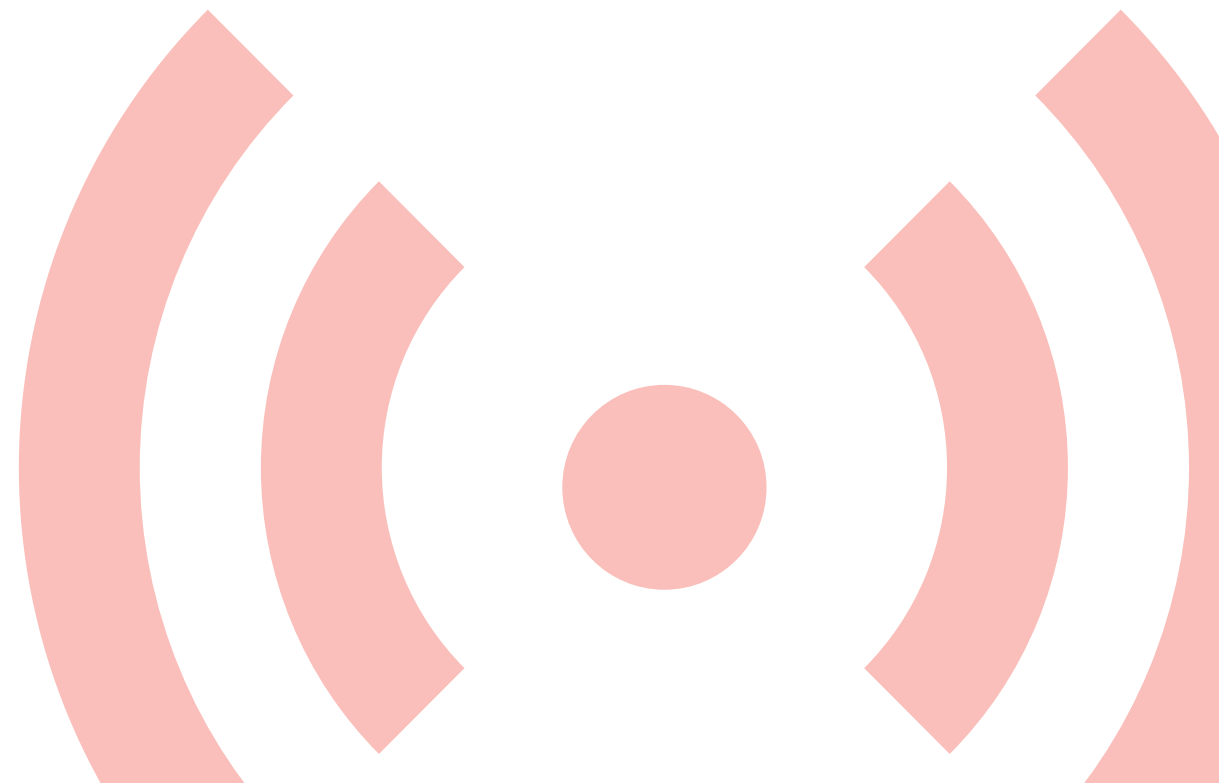
With automated tools to evaluate and improve agent systems quickly, Mosaic AI ensures agile development and robust governance across every component, from data to models, giving businesses full visibility and control.

What you'll learn

In this guide, you'll gain a comprehensive understanding of AI agents and AI agent systems, their applications and how to build them effectively. Specifically, you'll learn:

- **The fundamentals of AI agents and agent systems:** Understand the key components, capabilities and benefits of these systems in solving complex business challenges
- **Real-world use cases:** Explore practical examples of AI agents in action, such as customer service automation and multi-agent collaboration for advanced problem-solving
- **Building and scaling AI agent systems:** Learn how to design scalable, modular agent systems that integrate seamlessly with enterprise data, incorporate generative and classical AI, and adapt to your business needs

By following the insights, use cases and practical strategies outlined in this guide, you'll be equipped to deploy AI agent systems that are high-performing, governed and aligned with your organization's goals.



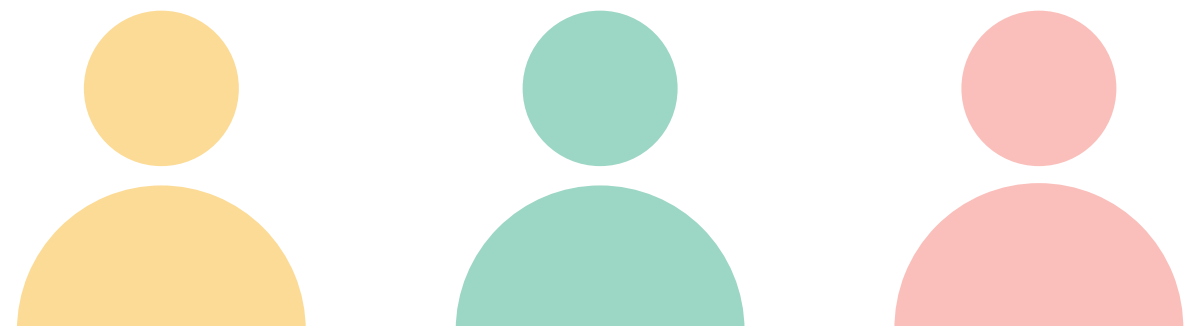
What Is an AI Agent?

AI agents are intelligent applications designed to automate tasks and enhance human productivity. They can analyze information, make decisions and take actions to achieve specific goals, freeing up time and resources for more strategic work.

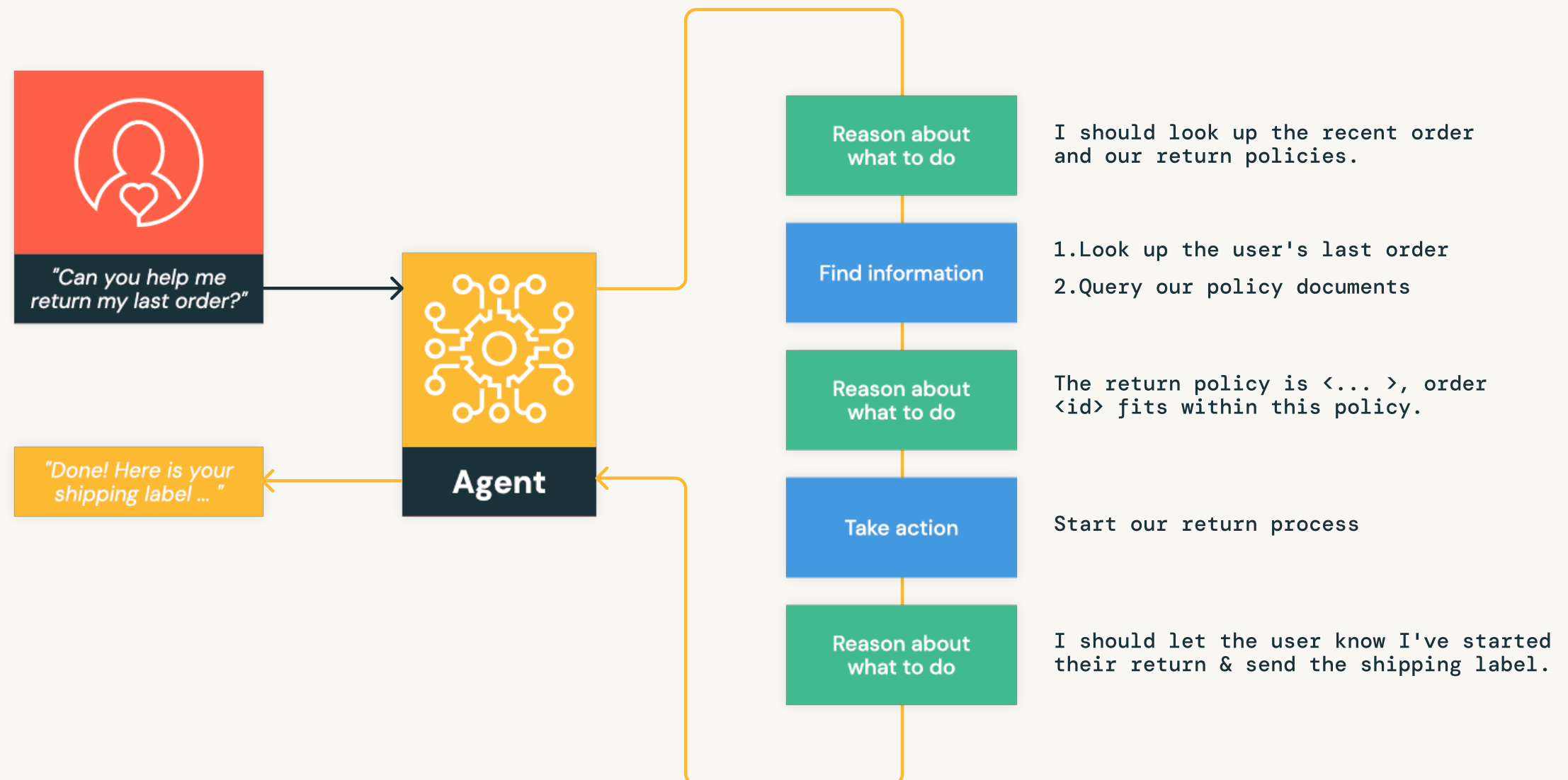
For example:

- **Customer service agent:** These agents interact with customers, understand their queries and provide timely and relevant responses, aiming to improve customer satisfaction and resolve issues efficiently
- **Campaign generation agent:** These agents analyze data, identify target audiences and generate personalized marketing campaigns to meet specific objectives, such as increasing brand awareness or driving sales
- **Code generation agent:** These agents assist developers by writing, debugging and optimizing code based on given requirements, streamlining the software development process and improving productivity

To build and deploy an effective AI agent, you need an AI agent system, regardless of whether it involves a single agent or multiple interacting agents.



An agent uses the LLM as its brain for reasoning



What Is an AI Agent System?

An AI agent system enables enterprises to build and operationalize an agent or set of agents that can perform complex tasks by combining multiple interacting components. An agent system goes beyond using a single model to integrate a variety of components, such as large language models (LLMs), classical machine learning (ML) models, enterprise data and tools to achieve specific goals efficiently. Additionally, an agent system also has built-in evaluation techniques and governance to ensure that the system delivers at high quality against the set goals and in a fully governed manner across all components of the system.

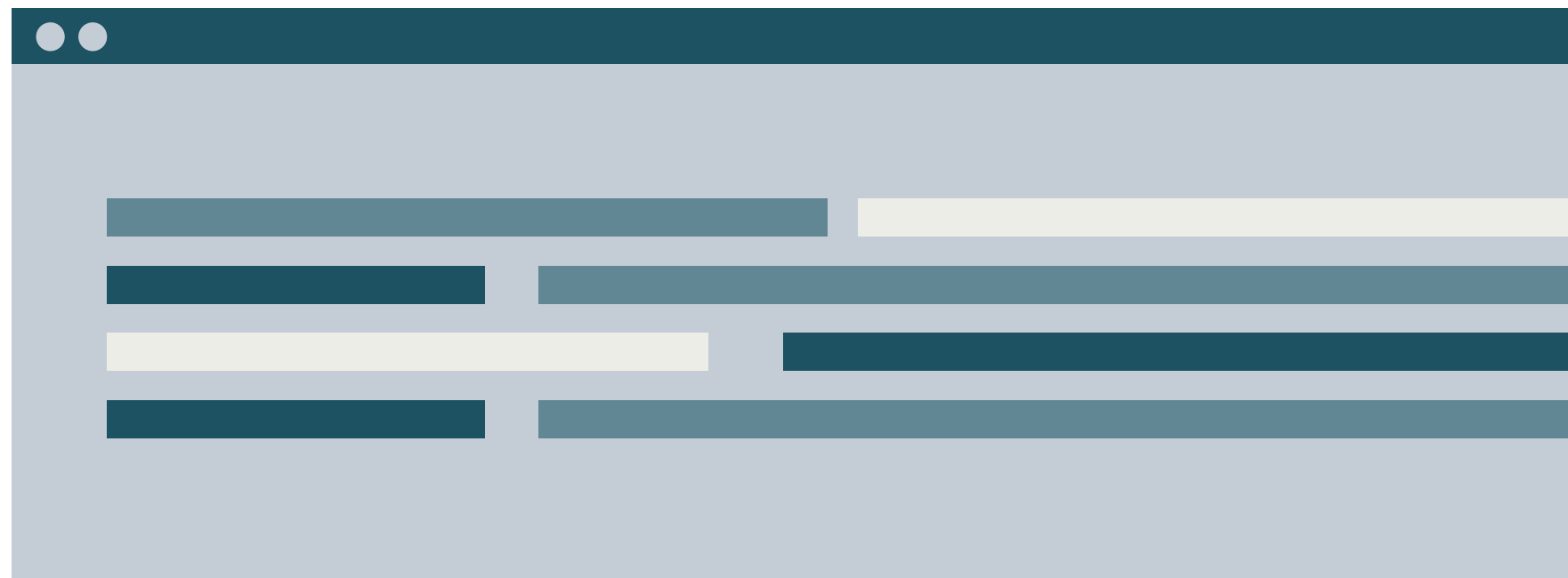
The key stages of developing and managing an AI agent system are the following:

- **Prepare data:** Organize and preprocess data to ensure it's accessible and relevant for agent decision-making and interaction
- **Build agents:** Combine generative AI models, classical ML models and tools to perform specific tasks
- **Deploy agents:** Implement the agent in real-world settings to interact with users and systems, ensuring secure and efficient operations
- **Evaluate performance:** Measure and assess the agent's output and progress to ensure it meets objectives and improves through feedback
- **Govern operations:** Maintain security, compliance and ethical standards while tracking agent activities and ensuring accountability

An example of an AI agent system is a retrieval augmented generation (RAG) application, which combines LLMs and data retrieval systems to provide accurate, context-aware responses. These systems excel at unstructured data tasks, such as answering customer queries by retrieving relevant documents from a database.

Classical machine learning models also play a key role in agent systems, particularly when working with structured data like tables, logs or time series. ML models act as specialized tools for tasks such as forecasting, classification or anomaly detection. For instance, an ML model within the system might predict inventory needs or flag high-risk transactions, complementing LLMs by focusing on precise, data-driven insights.

The systematic approach of an AI agent system ensures the integration of diverse AI components, allowing agents to perform complex, autonomous tasks effectively. This modular framework makes AI agents versatile and reusable across use cases like customer support, data analytics and workflow automation, transforming them from simple responders to decision-makers and action-takers.



What Are Use Cases for an AI Agent System?

AI agent systems can be applied across various business functions, delivering significant efficiency, accuracy and productivity gains.

Here are some practical examples of how AI agent systems can be leveraged:

AI agent system applications across industries					
 Financial Services	 Healthcare & Life Sciences	 Comm, Media & Entertainment	 Retail & Consumer Goods	 Manufacturing	 Public Sector
Fraud monitoring and predictions	Biomedical literature summarization & discovery	Hyper-personalization for customer experience (CX)	Try before you buy with virtual fitting rooms	Streamlined, personalized customer experiences	Analysis of open source Intelligence
Automating compliance data gathering	Clinical trial optimization	Enhancing customer support and self-service	Optimizing demand prediction and inventory	Increasing productivity and efficiency in operations	Modernizing legacy code bases
Accelerate underwriting and claims processing in insurance	Health insurance claims processing	Intelligent content creation and curation	Generate innovative product designs	Prescriptive and proactive field service	Regulatory compliance assistance

Customer support automation

AI agent systems can automate customer interactions through chatbots or virtual assistants, providing fast and accurate responses to routine inquiries. This reduces the burden on human agents, enabling them to focus on more complex issues. For example, an AI agent system could be used to handle common service queries while human agents manage escalated or specialized support needs.

Sales and marketing assistance

AI agent systems help sales and marketing teams by automating repetitive tasks like lead qualification, follow-up emails and data entry. Additionally, they can analyze customer data to offer insights and recommendations for targeted marketing strategies. For instance, an AI agent system might identify the best leads for follow-up or suggest personalized promotions based on customer preferences.

Data analysis and reporting

AI agent systems excel at automating data collection, analysis and reporting. This is particularly useful for businesses that need to process large volumes of data quickly and accurately. An AI agent system could pull data from multiple sources to generate weekly sales reports or performance dashboards, providing teams with timely and actionable insights without manual effort.

Personalized recommendations

AI agent systems can offer personalized recommendations by analyzing user behavior and preferences. This is commonly used in e-commerce, entertainment and other industries to suggest products or content tailored to individual users. For example, an AI agent system might recommend products based on browsing history or provide content suggestions in a streaming service based on past viewing patterns.

Task automation in operations

AI agent systems can take over routine operational tasks such as scheduling, inventory management and order processing. This reduces the potential for human error and increases operational efficiency. In manufacturing, for instance, an AI agent system might automate inventory updates and trigger reorder actions when stock levels run low.

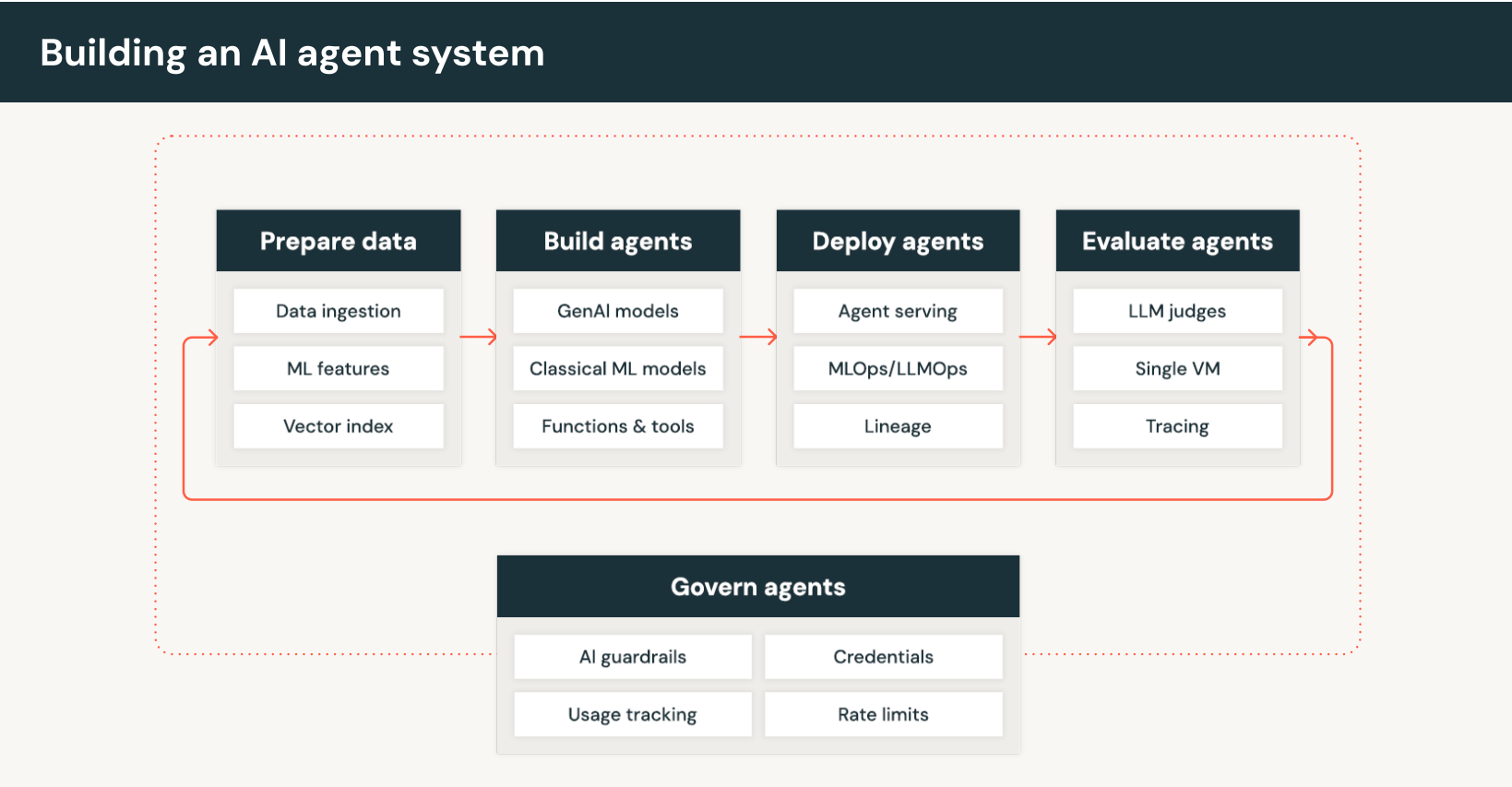
Fraud detection and prevention

AI agent systems can monitor transactions and activities in real time to detect and prevent fraud. By analyzing patterns and identifying anomalies, they can help businesses take proactive steps to mitigate risks. A financial institution, for example, could deploy an AI agent system to identify suspicious transactions and flag them for further investigation before any damage is done.

These examples illustrate how AI agent systems can be integrated into various business processes to drive automation, enhance decision-making and increase productivity, making them an invaluable asset for any organization.

Building an AI Agent System

Creating an effective AI agent system requires a structured approach to ensure it performs reliably and aligns with business goals. By breaking the process into key stages, organizations can design, deploy and manage agents that are accurate, scalable and well-governed. Below, we explore the core components of building an AI agent system:



1 Prepare data

The first step in building an AI agent system is ensuring your data is well organized and accessible. Agents rely on high-quality, relevant data to make informed decisions. Data preparation includes tasks like cleaning, processing and indexing data for easy access. With the integration of tools like vector **databases**, you can store and retrieve data in a format that allows the agent to interact with and learn from it effectively.

Additionally, agents can be customized to work with specific enterprise data. Instead of duplicating data, AI agent systems can generate vector indexes and ML features directly from production data, providing real-time insights and ensuring seamless integration of your data into the agent’s decision-making process.

2 Build agents

At the heart of the AI agent system is a central agent — the decision engine that processes inputs and drives actions based on predefined goals. Typically, this is a pretrained large language model that can understand and respond to natural language prompts. However, the agent can be customized to handle specific tasks by defining its persona and expertise.

Once the agent is established, memory is added to enable the agent to retain contextual information over time. This memory can be short-term (working memory) for ongoing tasks or long-term (episodic memory) to store past interactions, enabling the agent to recall previous decisions and learn from its experiences. The combination of short- and long-term memory enables the agent to offer more relevant, personalized responses as it continues to interact with users.

Planning is another essential component of building agents. This feature allows the agent to break down complex tasks into manageable steps, applying reasoning techniques like **chain-of-thought** or **hierarchical decision-making** to determine the best course of action. The planning function ensures the agent can tackle multifaceted problems effectively and adjust its approach as needed.

Finally, **tools** enable the agent to perform specific tasks, such as making API calls, executing code or retrieving information. These tools extend the agent's capabilities, allowing it to work autonomously while interacting with external systems to gather the necessary data or perform actions required to complete its objectives.

Using platforms like **Databricks Notebooks** or **LangChain**, you can design, test and deploy these tools seamlessly:

- In **Databricks Notebooks**, developers can define tools as custom functions, register them with Unity Catalog and integrate them into workflows, leveraging enterprise data directly
- With **LangChain**, you can build dynamic task pipelines where tools are chained together to process complex queries or interact with external systems

3 Deploy agents

Once your agent is built, the next step is deployment. This is where the agent takes on real-world tasks, interacting with users, systems and data to achieve its goals. The deployment process involves ensuring that the agent operates securely and efficiently, with proper access controls and safeguards in place.

As agents interact with enterprise systems, they can continuously adjust their behavior based on new inputs and ongoing tasks. By using planning and feedback mechanisms like **ReAct** and **Reflexion**, agents can improve their performance over time, refining their decision-making based on iterative learning.

4 Evaluate agents

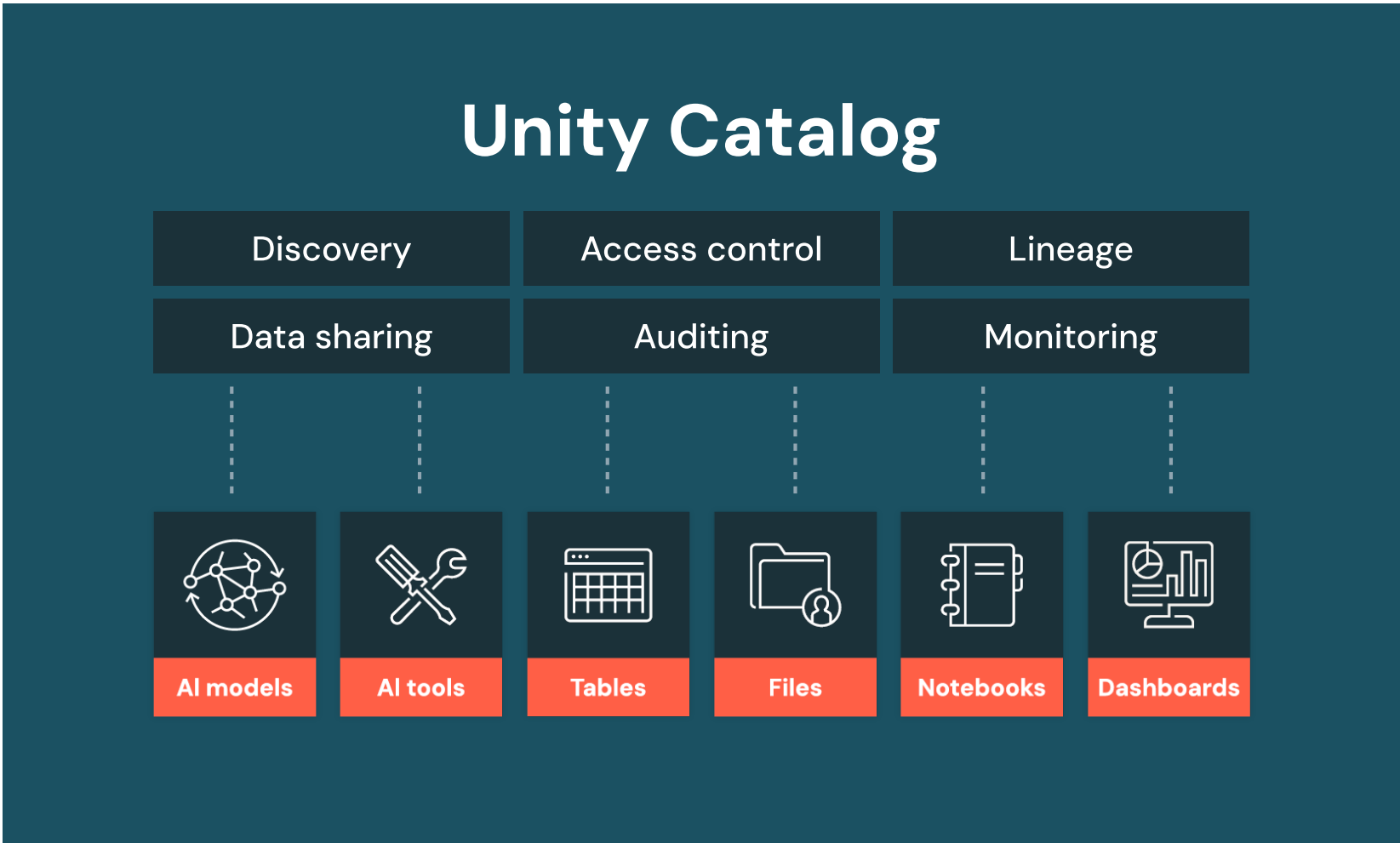
Effective evaluation is key to ensuring your AI agent system is delivering high-quality results. Evaluating agent performance involves measuring output quality and ensuring the agent meets your objectives. Evaluation tools help you track and assess the agent's performance against benchmarks, pinpointing areas for improvement.

Custom evaluation enables fine-tuning of the agent's capabilities, using AI-assisted judges and human feedback to grade responses. If any quality issues are detected, you can quickly trace the root causes, evaluate potential fixes and redeploy the agent, ensuring continuous improvement in real time.

5 Govern agents

Governance is critical in AI agent systems to ensure security, compliance and ethical operation. It spans data, models and tools, enforcing access controls, managing costs, preventing harmful content and tracking data lineage. Robust governance mitigates risks, maintains transparency and ensures agents operate responsibly.

For instance, consider a customer service agent handling sensitive user queries and accessing external APIs for tasks like order tracking or payment processing. With a governance tool like Mosaic AI Gateway, you can audit payloads. So every request and response is logged with detailed inference records, creating a complete audit trail. For example, if a customer disputes a response, you can trace the exact data and decision path that led to the agent's output.



By following this lifecycle — preparing data, building agents, deploying them, evaluating their performance and governing their operations — you can create a robust AI agent system that drives autonomous decision-making and continuous improvement, all while ensuring that human oversight remains integral to maintaining control, transparency and ethical standards.

Conclusion and General Guidelines

AI agent systems are reshaping how businesses automate, adapt and innovate. By combining generative AI, classical machine learning and specialized tools, these systems enable smarter, more scalable solutions. To maximize their potential, focus on these three key principles:

1. **Agents that reason over your data:** Directly connect agents to enterprise data for accurate, context-aware decision-making. Avoid data duplication and leverage tools to generate insights and features efficiently.
2. **Custom evaluation for your use case:** Regularly assess agent performance with tailored evaluation methods, including human-in-the-loop feedback. Continuously refine agents to align with specific business goals and ensure quality.
3. **Governance across data, models and tools:** Maintain transparency and security by implementing governance at every level. Enforce access controls, track lineage and monitor compliance to ensure reliable and ethical operations.

By adhering to these guidelines, businesses can harness the full power of AI agent systems to solve complex challenges and drive meaningful results.

What's the Next Step If I Want to Start Building AI Agent Systems?

Ready to start building AI agent systems? Explore these resources for step-by-step guidance, practical tips and advanced strategies to design modular, reliable and enterprise-ready AI applications.

- [Building Compound AI Systems With Agent Tools and Function Calls](#)
- [Create Your LLM Agents Leveraging Tools With Unity Catalog Functions | Databricks](#)
- [AI Agent Systems: Modular Engineering for Reliable Enterprise AI Applications | Databricks Blog](#)

Build high-quality AI agent systems — see how

Discover how to design and deploy AI agent systems that are autonomous, adaptable and tailored to your business needs. From customer support automation to advanced analytics, learn how to integrate data, models and tools into a governed, scalable framework. See why organizations worldwide trust Databricks Mosaic AI to streamline operations and drive innovation.

TRY DATABRICKS FREE

About Databricks

Databricks is the data and AI company. More than 10,000 organizations worldwide — including Block, Comcast, Condé Nast, Rivian, Shell and over 60% of the Fortune 500 — rely on the Databricks Data Intelligence Platform to take control of their data and put it to work with AI. Databricks is headquartered in San Francisco, with offices around the globe, and was founded by the original creators of Lakehouse, Apache Spark™, Delta Lake and MLflow. To learn more, follow Databricks on [LinkedIn](#), [X](#) and [Facebook](#).

