

1 Introducing Azure AI Agent Service



[monalisawhalin](#)



Microsoft

Nov 19, 2024

2 Azure announces managed capabilities to empower developers to build secure, stateful autonomous AI agents that automate every business process

Today, more than 60,000 organizations turn to [Microsoft Azure AI](#) to accelerate their business results and empower their teams. While large language models have provided remarkable productivity gains, many organizations still encounter significant challenges when dealing with complex, manual, and time-consuming workflows with minimal process automation.

AI agents have proven to be a great fit to automate such workflows. However, deploying reliable AI agents in real-world environments has proven to be difficult as current frameworks lack secure, integrated tools to trigger actions and enable secure data grounding, and organizations continue to face issues with data privacy, challenges with monitoring agent cost and performance quality, poor interoperability, and scalability struggles.

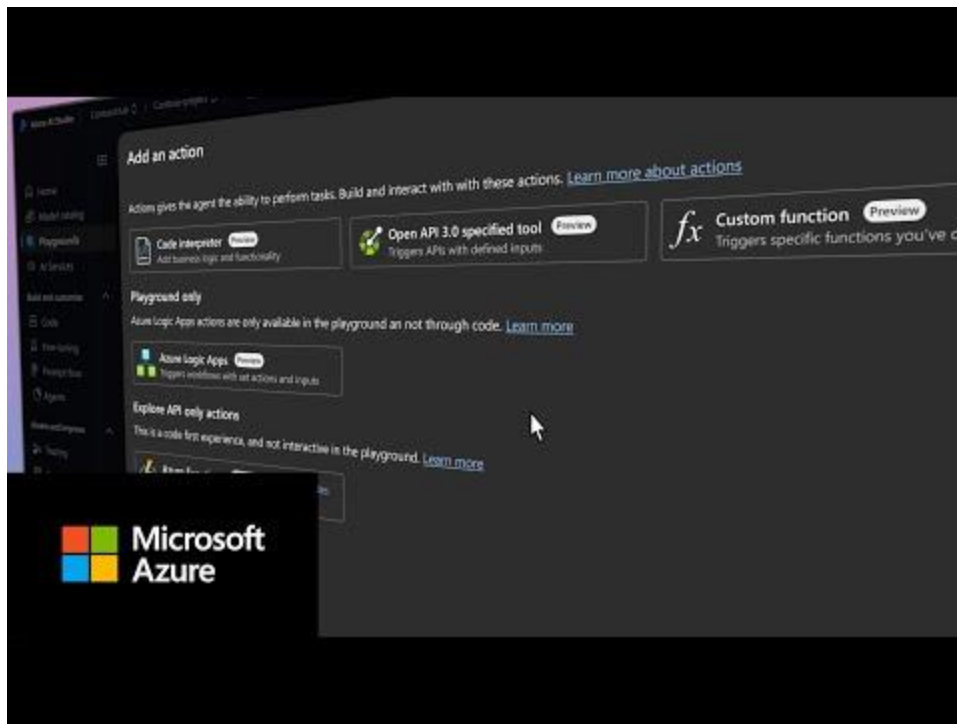
To fully realize the potential of autonomous AI agents, organizations need flexible, secure platforms for agent creation, deployment and monitoring.

3 Unlock autonomous agent capabilities with Azure AI Agent Service

Inspired by our customers' needs and the promise of autonomous AI agents, today at Ignite 2024, we are announcing the upcoming public preview of **Azure AI Agent Service, a set of feature-rich, managed capabilities that brings together all the models, data, tools, and services that enterprises need to automate business processes of any complexity.**

Azure AI Agent Service integrates the **latest models, tools and technology** from Microsoft, OpenAI and industry-leading partners such as Meta, Mistral and Cohere; seamlessly **extends your agents with knowledge** from Bing, SharePoint, Fabric, Azure AI Search, Azure Blob and licensed data, and **enable taking actions** across Microsoft and Third-Party applications with Azure Logic Apps, Azure Functions, OpenAPI 3.0 specified tools and Code Interpreter; an **intuitive agent building experience through [Azure AI Foundry](#)**; and **rich enterprise-ready features** including the ability to bring your own (BYO) storage, BYO-virtual private network, on-behalf-of authentication, and enhanced agent observability through OpenTelemetry based evaluation.

Watch a quick overview:



Video provider requires cookies to play the video. Accept to continue or [watch](#) it directly on the provider's site.

Accept

Azure AI Agent Service is **flexible and use-case agnostic**. This represents endless possibilities to automate routine tasks and unlock new possibilities for knowledge work - whether it is **personal productivity agents** that send emails and schedule meetings, **research agents** that continuously monitor market trends and automate report creation, **sales agents** that can research leads and automatically qualify leads, **customer service agents** that proactively follow-up with personalized messages, or **developer agents** that can upgrade your code base or evolve a code repository interactively.

Customers are already building agents with this new offering to automate exciting use cases and unlock new business value. Here's what our preview customers are saying:



"Azure AI Agent Service gives us a robust set of tools that accelerate our enterprise-wide generative AI journey. They help us quickly deploy impactful agents that provide scalable actions for Q&A, analysis, and tasks. By leveraging the service, we're able to shift our engineering time away from custom development and support the differentiators that matter to us." - Ethan Sena, Executive Director, AI & Cloud Engineering and Enablement, Bristol Myers Squibb



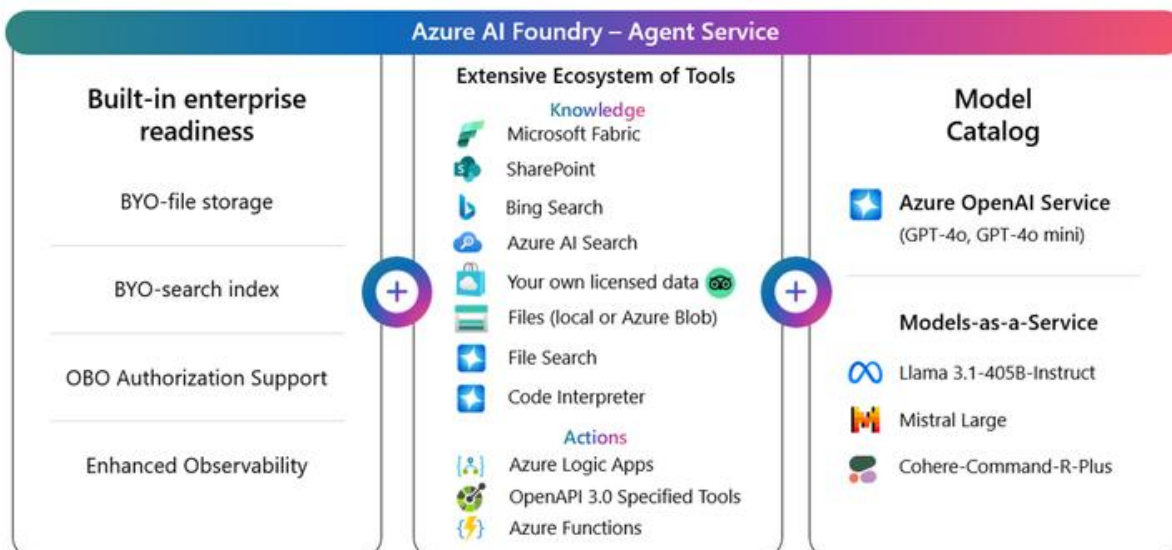
"Core42 plans to integrate Azure AI Agent Service APIs into Compass, its managed AI platform, enabling customers to make AI more actionable across their enterprise use cases through plug-and-play integration with various data sources and enterprise applications." – Raghu Chakravarthi, EVP, Engineering & GM US, Core42

4 What sets Azure AI Agent Service apart?

In our experience talking to hundreds of organizations, we have learned that developing secure, reliable agents rapidly requires four primary ingredients:

1. **Rapidly develop and automate processes:** Agents need to seamlessly integrate with the right tools, systems and APIs to perform deterministic or non-deterministic actions.
2. **Integrate with extensive memory and knowledge connectors:** Agents need to manage conversation state and connect with internal and external knowledge sources to have the right context to complete a process.
3. **Flexible model choice:** Agents built with the appropriate model for its task can enable better integration of information from multiple data types, yield better results for task-specific scenarios, and improve cost efficiencies in scaled agent deployments.
4. **Built-in enterprise readiness:** Agents need to be able to support an organization's unique data privacy and compliance needs, scale with an organization's needs, and complete tasks reliably and with high quality.

Leveraging the intuitive interface and comprehensive toolset in [Azure AI Foundry SDK and portal](#), Azure AI Agent Service provides these ingredients for end-to-end agent development through a unified product surface.



Now let's take a closer look at Azure AI Agent Service capabilities.

5 1. Rapid agent development and automation through powerful integrations

Building on the powerful yet flexible Assistants API, Azure AI Agent Service has built-in memory management and a sophisticated interface to seamlessly integrate with popular compute platforms, bridging LLM capabilities with general purpose, programmatic actions.

- **Enable your agent to take actions with 1400+ [Azure Logic Apps](#) connectors:** Leverage a wide ecosystem of connectors in Logic Apps to enable your agent to complete tasks and take actions on behalf of your users. With Logic apps, you simply need to define the business logic for your workflow in Azure Portal to connect your agent to external systems, tools and APIs. Examples of connectors include Microsoft products such as Azure App Service, Dynamics365 Customer Voice, Microsoft Teams, M365 Excel, and leading enterprise services such as MongoDB, Dropbox, Jira, Gmail, Twilio, SAP, Stripe, ServiceNow and many more.
- **Think beyond chat mode by implement stateless or stateful code-based actions with [Azure Functions](#):** Enable your agent to perform external interactions with the world, such as calling APIs or asynchronously sending and waiting for events. Azure Functions and Azure Durable Actions enable you to execute serverless code for **synchronous, asynchronous, long-running, and event-driven actions** such as approving invoices with **human-in-the-loop**, monitor an end-to-end product supply chain over long periods of time, and many more.
- **Perform advanced data analysis with Code Interpreter:** Enable your agent to write and execute Python code in a secure environment, handle diverse data formats and generate files with data and visuals. Unlike the Assistants API, you can integrate data in your storage to use with this tool.
- **Build standardized library of tools with OpenAPI specifications:** Connect your AI agent to an external API using an OpenAPI 3.0 specified tool, allowing for scalable interoperability with various applications. Enable your custom tools to authenticate access and connections with managed identities (Microsoft Entra ID) for added security, making it ideal for integrating with existing infrastructure or web services.
- **Extend [Llama Stack](#) agents with cloud-hosted tools:** Azure AI Agent Service supports the agent protocol for developers that are using Llama Stack SDKs. We will natively offer scalable, cloud-hosted, enterprise grade tools, while being wireline compatible with Llama Stack.

6 2. Securely ground your agent outputs with a rich ecosystem of knowledge sources

Easily configure a rich ecosystem of [enterprise knowledge sources](#) to enable agent to access and process data from multiple sources, improving accuracy of responses to user queries. Designed to work seamlessly with your data, these data connectors operate within your network parameter. These built-in data sources include:

- **Real-time, web data grounding with Bing:** Grounding with web data enables your agent to provide the most up-to-date information to users. This meets the shortcomings of LLMs today of not being able to provide factually responses to prompts about current events (for example, asking for today's top news headlines).
- **Grounding with private data in Microsoft SharePoint:** Connect your agent with your organization's internal documents in SharePoint for grounded responses. Agents simplify secure data access to SharePoint through on-behalf-of (OBO) authentication, so your agent can only access the SharePoint for which the end user has permissions.

- **Chat with your structured data in [Microsoft Fabric](#):** Power data-driven decision making in your organization without user knowledge of query languages like SQL or the data context. Enable your agent with the built-in Fabric AI Skills to create your own conversational Q&A systems on Fabric data using generative AI. Fabric also supports the OBO authentication mechanism for secure data connection.
- **Enrich agent outputs with private data in Azure AI Search, Azure Blob and your local files:** We have re-invented the File Search tool in Assistants API to enable you to bring your Azure AI Search index or create a new one using files in your Blob Storage or files in your local storage that leverages an inbuilt data ingestion pipeline. This new file search provides complete control over your private data by supporting file storage in your Azure storage account and search indexes in your Azure Search Resource.
- **Gain a competitive advantage with licensed data:** Enhance your agent responses with licensed data from proprietary data providers, such as Tripadvisor, to provide your agent with most recent, high-quality data tailored for your use case. We will be bringing you more licensed data sources in the future from a variety of industry verticals and specialties.

*"We're excited to partner with Microsoft as the first data and intelligence provider for its Azure AI Agent Service," said **Rahul Todkar, Vice President, Head of Data and AI at Tripadvisor**. "At Tripadvisor, we are focused on leveraging the power of Data and Generative AI to benefit all travelers and partners across the globe. This partnership makes it simpler than ever for broader businesses and developers to seamlessly connect with Tripadvisor's proprietary data in a compliant, privacy preserving way. Azure customers now can create advanced AI Agents powered by Tripadvisor's unique travel data and insights, unlocking new value and opportunities."*

- **Built-In conversation state management:** In addition to enterprise knowledge, managing threads or conversation state is essential for AI agents as it allows them to retain context, deliver personalized interactions, and enhance their performance over time. Azure AI Agent Service takes the pain out of building thread management systems by managing and retrieving conversation history from each end-user's interactions, providing consistent context within their sessions for richer interactions. This also helps you work around potential context window constraints of the model powering the AI agent.

7 3. Use the best model for the job, whether GPT-4o, Llama 3, or others

Developers have loved building AI assistants with the latest OpenAI's GPT models in Azure OpenAI Service Assistants API. Now, we are expanding model selection to include state-of-the-art models from industry-leading model providers so you can build task-specific agents, optimize your total cost of ownership, and more.

- **Leverage Models-as-a-Service:** Azure AI Agent Service will support a variety of models from Azure AI Foundry model catalog, while reusing cross-model compatible, cloud-hosted tools that perform code execution, retrieval-augmented generation, and much more. In addition to OpenAI's models from Azure OpenAI, developers can now create agents with Meta Llama 3.1, Mistral Large, and Cohere Command R+, supported via the Azure Models-as-a-Service API.
- **Multi-Modal support:** Unlock new scenarios with multi-modal support, enabling AI agents to process and respond to diverse data formats beyond text, expanding the potential use cases. We will be launching support for GPT-4o's image and audio modalities so that you can analyze

and combine data from various formats to deliver comprehensive insights, make decisions, and provide relevant outputs tailored to specific user needs.

8 4. Designed for building secure, enterprise-ready agents from the ground up

Business leaders are eager to build and innovate with AI-powered agents. Yet, enterprise customers fear their critical enterprise data getting into the wrong hands, customer satisfaction impacted by slow agent response time, and inability to quickly diagnose what has gone wrong to ensure reliably performing agents. Azure AI Agent Service packages necessary enterprise features to help organizations protect their sensitive data while meeting their regulatory requirements.

- **Bring your own storage:** Unlike Assistants API, you can now connect your enterprise data sources to enable your agent to securely retrieve enterprise data.
- **Bring your own virtual network:** Design agent applications with data privacy and compliance in mind with strict no public egress for your data traffic, keeping interactions secure within your network.
- **Keyless setup and OBO authentication:** Easily configure and authenticate your agents with features like keyless setup and on-behalf-of authentication, simplifying resource management and deployment.
- **Limitless scale:** Take advantage of limitless performance and scaling when running Azure AI Agent Service on your provisioned deployments. Now you can build agent-powered apps with predictable latency and high throughput, while maintaining flexibility.
- **Monitor agent performance with OpenTelemetry-based tracing:** Gain insights into your AI agent's performance and reliability. You can instrument OpenTelemetry-compatible metrics into your monitoring dashboard for offline and online evaluation of agent outputs through the Azure AI Foundry SDK.
- **Build responsibly with content filters and XPIA mitigation:** Azure AI Agent Service supports prebuilt and custom content filters that detect harmful content at varying severity levels. Prompt shields protect agents against cross-prompt injection attacks from malicious actors. As with Azure OpenAI Service, prompts and completions processed by the Azure AI Agent Service are not used to train, retrain, or improve Microsoft or 3rd party products or services without your permission. Customers can delete their stored data when they see fit.

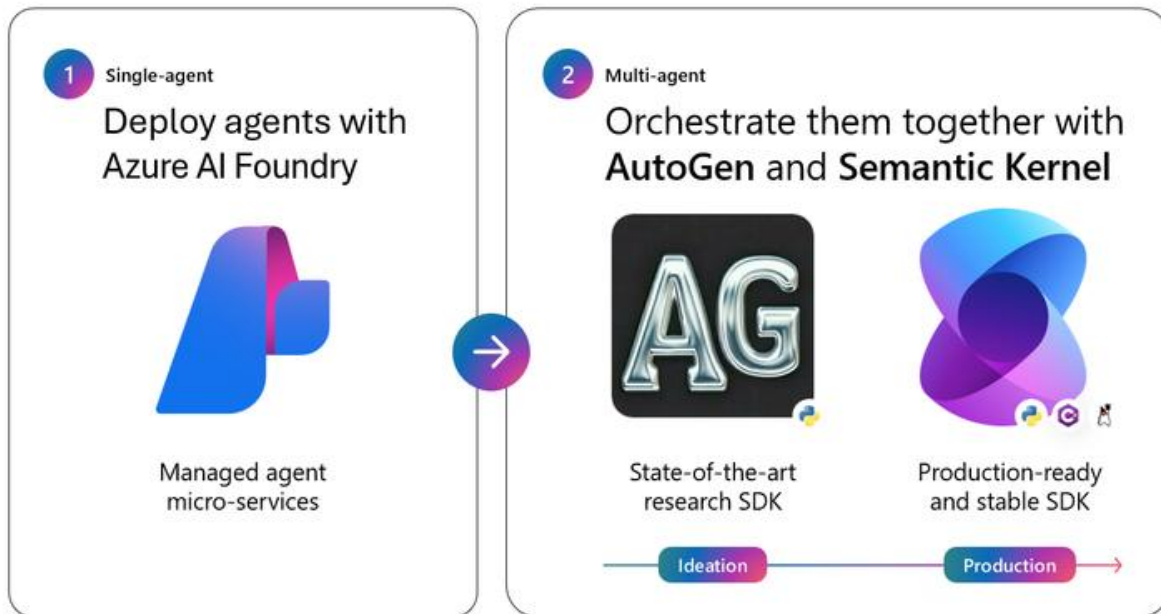
Orchestrate performant multi-agent systems with Azure AI Agent Service

Customers have been experimenting with multi-agent systems as a strategy for achieving performant autonomous workflows. In multi-agent systems, multiple context-aware autonomous agents, whether human or AI, interact or work together to achieve individual or collective goals.

Azure AI Agent Service works out-of-the-box with multi-agent orchestration frameworks that are wireline compatible with the Assistants API, such as [AutoGen](#), a state-of-the-art research SDK for Python created by Microsoft Research, and [Semantic Kernel](#), an enterprise AI SDK for Python, .NET, and Java.

When building a new multi-agent solution, start with building singleton agents with Azure AI Agent Service to get the most reliable, scalable, and secure agents. You can then orchestrate these agents together AutoGen is constantly evolving to find the best collaboration patterns for agents (and

humans) to work together. Features that show production value with AutoGen can then be moved into Semantic Kernel if you're looking for production support and non-breaking changes.



9 Get started with Azure AI Agent Service

- Start building with the SDK [documentation](#)
- Dive deep into [enterprise knowledge extensibility](#) in Azure AI Agent Service
- Learn how to design, customize and manage AI applications with [Azure AI Foundry](#)
- Learn how to empower data-driven decision making with [Microsoft Fabric](#)
- Watch this recorded [breakout session](#) from Ignite 2024 to learn more about how companies are automating key business processes with Azure AI Agent Service
- Watch this live [demo](#) at Ignite 2024 to learn how customers are going beyond chat-based interactions

We can't wait to see what you build with Azure AI Agent Service!