

Azure Ai

**Building Agentic Ai
Applications on the
Microsoft Cloud**



Coding AI Apps on Microsoft Azure

Welcome to Coding AI Apps on Azure, your guide to harnessing the power of artificial intelligence through Microsoft's cutting-edge Azure AI Foundry platform.

In an era where AI is transforming industries, redefining innovation, and reshaping how we solve problems, developers stand at the forefront of this revolution.

Whether you're a seasoned coder or just beginning your journey, this book equips you with the knowledge, tools, and practical skills to build intelligent, scalable, and responsible AI applications using Azure's robust ecosystem.

Azure AI Foundry is more than just a platform—it's a developer's gateway to creating AI solutions that are powerful, accessible, and tailored to real-world needs.

From accessing a vast catalog of over 1,900 frontier and open-source models to seamlessly integrating data, building agentic workflows, and ensuring responsible AI practices, Azure empowers developers to turn ideas into impactful applications. This book takes you step-by-step through the process, blending hands-on coding tutorials with insights into the platform's capabilities, all while keeping the focus on practical, enterprise-ready solutions.

In these pages, you'll explore how to prototype, customize, and deploy AI applications with ease, leveraging tools like Visual Studio, GitHub, and Azure's SDKs. You'll learn to harness Retrieval Augmented Generation (RAG), build autonomous AI agents, and implement safety measures to ensure your applications are both innovative and trustworthy. Through real-world examples and exercises, inspired by companies like BMW and Fujitsu, you'll see how Azure AI Foundry drives efficiency and creativity across industries.

Coding AI Apps on Microsoft Azure

Whether you aim to automate processes, enhance user experiences, or solve complex challenges, Coding AI Apps on Azure is your roadmap to mastering AI development. Let's dive in and start building the future, one line of code at a time.

2025: The Age of AI agents and Building the Open Agentic Web

As [Microsoft writes here](#), they define 2025 as the year adventing the era of the 'Agentic Web'.

This describes a transformative vision where AI agents act not merely as tools, but as autonomous partners capable of reasoning, decision-making, and task execution on behalf of individuals and organizations.

In a world reshaped by AI, this new era is driven by ground-breaking advancements in AI reasoning and memory, empowering developers and businesses to redefine workflows, accelerate innovation, and solve complex problems in unprecedented ways.

In this best practice series we explore how Azure AI is catalyzing the evolution of the Agentic Web—a paradigm where AI-driven agents, powered by advanced machine learning, natural language processing, and cognitive services, redefine human-digital interaction.

Azure AI

At the heart of this transformation lies Azure AI, Microsoft's powerful suite of artificial intelligence tools and services, which is empowering developers, businesses, and visionaries to build a web that is proactive, personalized, and profoundly intelligent.

From the evolution of GitHub Copilot into a full-fledged coding agent to the introduction of Azure AI Foundry and Microsoft Discovery, which streamline AI application development and scientific research, Microsoft is laying the foundation for a future where AI agents seamlessly integrate into every facet of work and life.

2025: The Age of AI agents and Building the Open Agentic Web

This book explores the technologies, platforms, and open standards—such as the Model Context Protocol (MCP) and NLWeb—that are powering this shift, offering a glimpse into how developers and enterprises are building a collaborative, secure, and intelligent digital ecosystem that promises to redefine the internet and fuel the next wave of global innovation.

Azure AI Foundry enables developers to build and manage AI applications, integrating models like xAI's Grok 3 and Grok 3 mini. It now supports the Agent2Agent (A2A) protocol for interoperability and includes observability tools for performance and security metrics.

Empowering Developers

AI models are now more capable due to improved reasoning and memory. GitHub Copilot, used by 15 million developers, features agent mode and code review to streamline coding processes, while Microsoft 365 Copilot and Copilot Studio are utilized by hundreds of thousands of customers and over 230,000 organizations, including 90% of the Fortune 500, for research, brainstorming, and automation.

2025: The Age of AI agents and Building the Open Agentic Web

- **Microsoft Discovery:** A new platform to accelerate scientific discovery with agentic AI, transforming research and development processes across industries.
- **Open Standards and NLWeb:** Microsoft supports the Model Context Protocol (MCP) across platforms like GitHub, Copilot Studio, and Windows 11, and introduced NLWeb, likened to HTML for the agentic web, enabling conversational website interfaces.
- **Security and Governance:** Microsoft Entra Agent ID assigns unique identities to AI agents for secure management, and Microsoft Purview extends data security and compliance to AI apps, addressing risks like “agent sprawl.”
- **Developer Empowerment:** Microsoft emphasizes tools like Windows AI Foundry and GitHub Copilot’s asynchronous coding agent to enhance the software development lifecycle, enabling developers to focus on strategic tasks.

These announcements reflect Microsoft’s vision of an interconnected, secure, and developer-friendly agentic ecosystem, driving digital transformation across industries.

Begin Your Agentic AI Journey

From enabling hyper-personalized user experiences to automating intricate workflows across industries, Azure AI provides the scalable, secure, and versatile infrastructure needed to bring this vision to life. We will delve into the technical foundations, practical applications, and transformative potential of this technology, offering insights for developers, business leaders, and curious minds alike.

Through real-world case studies, technical blueprints, and forward-looking perspectives, this book unpacks how Azure AI’s capabilities—such as Azure Cognitive Services, Azure Machine Learning, and the Azure Bot Framework—are empowering a new generation of intelligent agents.

Azure AI Foundry: The AI App and Agent Factory

These agents are not merely tools but collaborative partners that learn, adapt, and act with unprecedented autonomy. Whether you're a developer seeking to harness Azure's tools to build cutting-edge applications or a strategist envisioning the future of digital ecosystems, this book will guide you through the principles, practices, and possibilities of the Agentic Web.

Join us on this journey to discover how Azure AI is not just shaping the future of the web but redefining what it means to interact, create, and thrive in a world where intelligence is woven into the very fabric of our digital lives.

The **latest Azure AI Foundry innovations** are central to Microsoft's vision of enabling the Agentic Web.

Azure AI Foundry is Microsoft's comprehensive platform for building, deploying, and managing AI solutions.

It is designed to help organizations navigate the rapidly evolving AI ecosystem by providing tools that enhance productivity, streamline operations, and deliver unique value.

The platform supports a wide range of AI applications, from generative AI to data analytics, and is built to address the needs of businesses across industries.

The Azure AI Agent Service is a flagship offering that enables businesses to create AI agents capable of automating routine tasks and enhancing productivity. These agents are designed to integrate seamlessly with enterprise knowledge bases, ensuring that their outputs are grounded in accurate, context-aware information. This is achieved by leveraging a variety of data sources, including unstructured data (e.g., documents, emails), structured data (e.g., databases), private enterprise data, licensed content, and public web data.

Azure AI Foundry: The AI App and Agent Factory

Microsoft Fabric Data Agents represent a significant advancement in integrating AI with data analytics. These agents are designed to enhance operational efficiency by combining the power of AI with Microsoft Fabric's robust data platform. Fabric Data Agents enable organizations to extract actionable insights from their data, automate decision-making processes, and optimize workflows.

The service is particularly valuable for organizations looking to streamline repetitive processes, such as customer support, data entry, or content generation. For example, an AI agent could be deployed to handle customer inquiries by pulling relevant information from a company's knowledge base and delivering precise, personalized responses.

The inclusion of **NVIDIA NIM (NVIDIA Inference Microservices)** in Azure AI Foundry marks a significant step toward optimizing AI performance and cost-efficiency. These microservices leverage NVIDIA's expertise in GPU-accelerated computing to enhance the speed and scalability of AI workloads.

Azure AI Foundry and the DevOps Toolchain to Infuse AI in All Your Apps

Microsoft's new innovations empower enterprises to infuse AI throughout their apps and DevOps life-cycle.

Azure AI Foundry, combined with a robust Dev Toolchain, offers developers a powerful set of tools and processes to seamlessly enhance your development workflow and deliver intelligent solutions.

The AI-Enhanced DevOps Toolchain plays a crucial role in streamlining the integration of AI into your applications. By incorporating AI capabilities into your development workflow, you can enhance the functionality and intelligence of your applications.

Understanding Azure AI Foundry

Azure AI Foundry is a comprehensive platform provided by Microsoft that empowers developers to build, deploy, and manage AI models at scale.

It offers a range of AI services, including machine learning, natural language processing, computer vision, and more. With Azure AI Foundry, developers can access pre-built AI models, customize them to suit their needs, and integrate them into their applications with ease.

Integrated Development Environment:

Azure AI Foundry seamlessly integrates with popular tools like Visual Studio, GitHub, and Copilot Studio, allowing developers to work in familiar environments. This reduces the learning curve and enhances productivity by embedding AI capabilities directly into existing workflows.

Azure AI Foundry and the DevOps Toolchain to Infuse AI in All Your Apps

Comprehensive Model Catalog: The platform offers access to over 1,900 models, including frontier and open-source options from providers like Microsoft, Azure OpenAI, DeepSeek, Hugging Face, Meta, Mistral, and xAI (e.g., Grok 3 and Grok 3 mini). Developers can explore, evaluate, and select models tailored to their use case via a unified API, simplifying model integration and experimentation.

End-to-End AI Development: Azure AI Foundry supports the full development lifecycle—define and explore, build and customize, observe and improve. Developers can prototype in intuitive playgrounds, fine-tune models with techniques like vision fine-tuning and distillation, and deploy production-ready applications with built-in scalability. The platform's SDKs (available in languages like Python, C#, Java, and JavaScript) streamline coding and resource management.

Agentic Capabilities: The Azure AI Foundry Agent Service, now generally available, enables developers to build, deploy, and scale AI agents for complex workflows. Tools like the Semantic Kernel and AutoGen simplify multi-agent orchestration, reducing code complexity. For example, Fujitsu used these tools to improve sales proposal creation by 67%.

By combining these features, Azure AI Foundry empowers developers to build innovative, secure, and scalable AI applications efficiently, bridging cutting-edge technology with practical business needs.

Companies like C.H. Robinson, PIMCO, BMW Group, and Dentsu have leveraged Azure AI Foundry to automate logistics, enhance client services, improve data delivery, and reduce time-to-insights by up to 90%, showcasing its practical value across industries.

Building a Digital Workforce with Multi-Agents in Azure AI Foundry Agent Service

As Microsoft writes [here](#) the multi-agent orchestration capabilities of Azure AI Foundry Agent Service enable developers to build scalable, modular, and intelligent AI systems.

It highlights the shift from single-agent to multi-agent architectures, which are better suited for complex, multi-step business processes like customer support, market research, and financial analysis.

Key features include:

- **Connected Agents:** These allow developers to break down tasks into specialized roles, enabling seamless collaboration without custom orchestrators. For example, a customer authentication agent verifies identities, while a billing agent retrieves account data, ensuring modularity and extensibility.
- **Interoperability:** Support for open protocols like Model Context Protocol (MCP) and Agent2Agent (A2A) enables agents to communicate across platforms, with A2A APIs facilitating multi-turn conversations and context handoffs.
- **Real-World Impact:** Companies like JM Family use the service to streamline software development through collaborative AI ecosystems, such as the BAQA Genie, demonstrating enhanced efficiency and compliance.

Building a Digital Workforce with Multi-Agents in Azure AI Foundry Agent Service

- **Enterprise Features:** The platform offers long-term state management, human-in-the-loop controls, and integration with tools like Semantic Kernel and AutoGen for simplified orchestration.

The article emphasizes how Azure AI Foundry Agent Service empowers developers to create robust, enterprise-grade digital workforces that automate complex workflows with minimal human intervention, fostering innovation and scalability.

What is the Agent2Agent (A2A) Protocol

The [Agent2Agent \(A2A\) protocol](#) is an open standard designed to enable seamless communication and collaboration between AI agents across different platforms, clouds, and organizational boundaries. Integrated into Azure AI Foundry and Microsoft Copilot Studio, A2A facilitates the creation of interoperable, multi-agent systems that can exchange goals, manage state, invoke actions, and return results securely and observably.

Building a Digital Workforce with Multi-Agents in Azure AI Foundry Agent Service

The A2A protocol, announced by Google and supported by Microsoft as of May 2025, is a standardized framework that allows AI agents—semi-autonomous programs designed for specific tasks—to communicate and collaborate across ecosystems, regardless of the framework or vendor. Unlike traditional AI systems that operate in isolation, A2A enables agents to work together as a team, mirroring human collaboration across departments or organizations.

Key features of A2A include:

- **Structured Communication:** Agents exchange goals, manage state, invoke actions, and return results in a standardized format, ensuring seamless interaction.
- **Cross-Platform Interoperability:** A2A allows agents built on different platforms (e.g., Azure AI Foundry, Google Vertex AI, SAP Joule) to collaborate without custom integrations.
- **Security and Governance:** Every A2A call is routed through enterprise-grade safeguards, including Microsoft Entra for identity management, mutual TLS for secure communication, Azure AI Content Safety for content filtering, and full audit logs for accountability.
- **Tool Compatibility:** A2A supports popular frameworks like Semantic Kernel and LangChain, enabling developers to use familiar tools while maintaining interoperability.

Building a Digital Workforce with Multi-Agents in Azure AI Foundry Agent Service

A2A Integration with Azure AI Foundry

Azure AI Foundry leverages A2A to enhance its Agent Service, which is generally available as of May 2025 and used by over 10,000 organizations, including Fujitsu and NTT DATA. The integration empowers developers to build multi-agent systems that span internal tools, partner systems, and external infrastructure while maintaining governance and service-level agreements (SLAs). Here's how A2A integrates with Azure AI Foundry:

Connected Agents: A2A enables “connected agents” (in preview), allowing point-to-point interactions where one agent can call another as a tool to handle specialized tasks. For example, in the retail system from Chapter 1, the QueryAgent could invoke an external A2A-compatible agent to retrieve real-time inventory data from a partner's system. This eliminates the need for external orchestrators, simplifying workflows and reducing development complexity.

Multi-Cloud Orchestration: A2A supports multi-cloud architectures, allowing Azure AI Foundry agents to interact with agents on platforms like Google Vertex AI or SAP Joule. For instance, an Azure-based OrderAgent could collaborate with a Google-based logistics agent to optimize delivery schedules. This is facilitated by A2A's open-source nature, with Microsoft contributing to the A2A working group on GitHub to refine the protocol and tooling.

Building a Digital Workforce with Multi-Agents in Azure AI Foundry Agent Service

Unified SDK and APIs: Azure AI Foundry's Agent Service integrates A2A with Semantic Kernel and AutoGen through a single SDK, providing a unified API surface for defining and managing multi-agent workflows. Developers can build, simulate, and deploy agents locally or in the cloud with consistent behavior. The A2A API head enables open-source orchestrators to use Foundry agents seamlessly, supporting multi-turn conversations without custom integrations.

Enterprise-Grade Security: A2A calls in Azure AI Foundry are secured with Microsoft Entra for identity verification, mutual TLS for encrypted communication, and Azure AI Content Safety to detect harmful content. Audit logs ensure compliance, making A2A suitable for mission-critical enterprise applications. For example, in the retail system, A2A ensures that sensitive customer data shared between the QueryAgent and an external payment agent is encrypted and logged for compliance.

Prebuilt Templates: Azure AI Foundry offers prebuilt templates for common A2A use cases, such as customer service escalation and IT incident resolution, which developers can customize for their needs. These templates accelerate the development of multi-agent systems like the retail example.

A2A, alongside the Model Context Protocol (MCP), represents a shift toward “agentic computing,” where AI systems dynamically adapt to workflows spanning human and machine collaborators.

With a public preview of A2A in Azure AI Foundry and Copilot Studio expected soon, developers can anticipate broader adoption and enhanced capabilities. The protocol's open-source nature and support from major players like Microsoft, Google, and SAP signal a move away from isolated AI systems toward interconnected ecosystems, mirroring the collaborative nature of global business.

Building a Digital Workforce with Multi-Agents in Azure AI Foundry Agent Service

The Agent2Agent protocol transforms Azure AI Foundry into a hub for interoperable, multi-agent systems, enabling developers to build collaborative AI workflows that span platforms and organizations. By integrating A2A, Azure AI Foundry simplifies the development of complex systems like the retail example, enhances security and compliance, and fosters innovation through open standards.