

# Al in action: Creating better experiences with apps and agents



#### 01 /

#### Introduction

#### 02 /

#### Driving business value with Al

- **03** Generative Al apps and the evolution of agents
- **04** Agentic AI: Expanding what AI can do for your business
- 05 Example use cases for Al apps and agents
  - 5.1 Accelerating content delivery
  - 5.2 Transforming customer service
  - 5.3 Optimizing employee workflows
  - **6.1** Preventing fraud and detecting anomalies
  - 6.2 Unlocking organizational knowledge
  - 6.3 Automating document processing

#### 03 /

### Building a foundation for your Al estate

- 09 Create reliable and scalable Al solutions on Azure
  - 10 Build Al apps more easily
  - 11 Use a modern, managed app platform
  - 11.1 Ensure that you have the robust data capabilities to feed your Al
- 13 Take advantage of the wider capabilities of Azure

#### 04 /

### Start building the Al-powered future with Azure

## Introduction

Al in business has evolved from being a cutting-edge, experimental technology to a core strategic asset, delivering practical, real-world benefits. Sophisticated Al-powered solutions, ranging from productivity-boosting apps to customer-facing tools and powerful Al agents, are no longer conceptual or experimental. They're in production and creating measurable business impacts.

Enterprise-grade Al applications and agents require a comprehensive set of Al services, models, databases, application platform services, and developer tools that are only available on large cloud platforms. As you consider building, deploying, and operating these solutions, you must choose the best cloud provider and services to support the following key pillars, depending on your organization's needs:

- Al. Al services provide the core natural language, conversational, summarization, and creative response capabilities that make Al apps so powerful.
- Apps. Managed services need to provide the runtime environment, APIs, and user interfaces
  that allow AI apps to operate and must integrate them with other pieces of your infrastructure.
- **Data.** High-quality, well-managed data is a prerequisite for your Al apps to provide accurate, context-aware, and relevant functionality, while also helping to ensure compliance and governance.

Together, these components facilitate AI solutions that are not only powerful but also reliable, secure, and aligned with business goals.

In this e-book, we explain how AI apps and agents can drive your business growth and how you can use the Azure platform's trusted tools and infrastructure to create these kinds of innovative AI-powered solutions.

# Driving business value with Al

With Al-powered apps and agents, your business can deliver experiences that are more streamlined, personalized, and consistent for your customers. They can give instant support, offer tailored product recommendations, and even help users complete tasks more efficiently. Around the clock, Al solutions help reduce wait times and often anticipate customer needs based on past behavior, creating smoother and more satisfying interactions.

Al-powered solutions also offer advantages for your business. Internally, they can reduce operational costs and increase staff productivity by streamlining tasks like customer support, document handling, and data analysis. For technical and development staff, coding copilots, automated testing, and other Al-powered tools can speed up

innovation and time to market for new products. For decision-makers, AI can improve business intelligence capabilities, surfacing insights from across the organization and helping leaders to make faster, better-informed decisions.

#### **Key Al outcomes**

Businesses generally adopt Al-powered solutions to achieve any or all of four key outcomes:

- **Enhancing employee experiences.** By automating repetitive, mundane tasks, Al empowers employees to get more done and to focus on more complex, creative, and higher-value work.
- **Transforming customer engagement.** Al enables more personalized and tailored customer interactions, while reducing workloads for employees.
- Reimagining business processes. Al is revolutionizing operations across marketing, supply chain, and finance, optimizing processes and uncovering growth opportunities.
- **Accelerating innovation.** Al can speed up creative workflows and product development, streamlining time to market and helping companies stand out in competitive industries.

### Generative Al apps and the evolution of agents

Generative AI apps have fundamentally changed how we interact with technology. At their most basic, these apps, driven by powerful language models, can respond to questions posed in plain (or *natural*) language with useful answers based on their training data. Organizations are increasingly adopting generative AI–enabled apps for internal support, sales enablement, data analysis, and more.

As the adoption of AI apps has grown, so have their capabilities. They can now routinely do things like make decisions based off of pictures, automatically translate between languages, ingest and summarize large collections of documents, and more. These types of AI-powered features are increasingly integrated into directly into office productivity suites, human resources (HR) management tools,

customer relationship management (CRM) systems, and development environments. Modern AI apps are also increasingly able to perform more complex tasks, like scheduling meetings, generating reports from analytics dashboards, or assisting in software development, without explicit instructions.

This progression has laid the foundation for the emergence of Al-powered agents. Simply put, agents are systems that use AI to automate and execute business processes. They can provide a spectrum of capabilities. Augmented retrieval agents are on the simpler side of this spectrum, with the agent interacting with external data sources to improve its Al-based responses. Task-based agents, connected to specific workflows or processes, are more sophisticated. On the more advanced end of the spectrum are autonomous agents (agentic Al systems), which can operate independently, dynamically plan their actions, and use tools to perform specific actions.



# Agentic AI: Expanding what AI can do for your business

Autonomous or semiautonomous agentic Al systems are more independent than traditional Al apps. They combine language models with goal-oriented logic to carry out complex tasks across systems and data sources, and they're capable of handling complicated behind-thescenes logic and tasks.

Agentic AI systems consist of several components:

- Intent recognition and planning.
   The agent monitors its environment to identify and interpret user input or system triggers, using a language model's logic to determine appropriate tasks or goals and to break them into actionable steps.
- Tool use and action execution. Agents are equipped with predefined sets of tools, APIs, or plugins (like calling a REST endpoint, running a query, or accessing CRM data) for performing actions and interacting with external systems.
- Memory and state management. Language models are generally not state-aware. Agents, by contrast, can track conversation history, task state, and results, enabling context awareness across interactions and longrunning processes.



 Reasoning and feedback loops. Using the capabilities of language models, agents make decisions at each step, adjust based on real-time feedback, and retry or escalate if needed.

To handle more complex workflows, these types of agents can be combined into larger multi-agent systems. They can assign specialized roles to different agents that can operate concurrently, collaborate effectively, and scale independently. By mimicking the dynamics of human teams and delegating tasks, sharing insights, and coordinating actions, multi-agent systems can drive smarter automation across workflows like customer service, logistics, and data analysis, delivering more robust and scalable solutions.

Agentic AI systems are well-suited to handling repetitive processes and can enhance your decision-making by providing

real-time data access and logic. With the ability to integrate securely into enterprise workflows and adapt to changing contexts, these agents can take on your organization's more tedious tasks, so your staff can focus on work that only humans can handle.

#### **Example use cases for** Al apps and agents

The types of Al apps and agents that we've discussed can help your business in many ways, including automating complex tasks, enhancing decision-making, increasing efficiency across operations, improving customer experiences, and more. Following are a few examples of how AIbased solutions are helping organizations meet their real-world business needs:

#### Accelerating content delivery

Scenario: An automotive business intelligence company helps car dealerships streamline content creation by automatically generating tailored product descriptions and marketing assets across digital platforms.

Al in action: A generative Al app uses vehicle data, market insights, and customer profiles to produce personalized text and images for online catalogs,

websites, and campaigns. This reduces manual work, speeds up the content creation and delivery process, and helps improve customer engagement.

#### **Transforming customer** service

**Scenario:** A global airline improves service efficiency and scalability by automating customer support and empowering agents with real-time insights.

Al in action: A generative Al assistant handles routine customer queries through a natural language interface, routes complex cases to human agents, and summarizes prior interactions to assist with faster resolution. It also automates repetitive tasks, like call logging. This solution helps to reduce support costs, increase agent productivity, and enhance the overall customer experience.

#### **Optimizing employee** workflows

**Scenario:** A healthcare services provider boosts clinician retention and job satisfaction by automating administrative tasks and improving access to essential information.

Al in action: By organizing unstructured data and creating a conversational search experience, an Al-powered solution enables

employees to quickly find and access relevant information. It also automates routine tasks, like report generation, and streamlines document management, allowing clinicians to focus on higher-value work. This results in improved productivity, increased staff retention, and higher job satisfaction.

#### Preventing fraud and detecting anomalies

Scenario: A risk-reduction company that specializes in combating money laundering enhances its fraud detection and decisionmaking speed by taking advantage of AI to identify suspicious activities and patterns in real time.

**Al in action:** To detect potentially fraudulent activity, an intelligent application enables real-time AI-enabled monitoring of vast numbers of transactions. By making the most of predictive models in machine learning and AI algorithms, this application identifies patterns and anomalies associated with fraudulent behavior, automates risk scoring, and accelerates decision-making. This helps individual analysts to be more efficient, speeds up fraud detection and investigations, and ultimately helps minimize financial losses.

#### Unlocking organizational knowledge

Scenario: A tax services provider boosts productivity by helping professionals instantly access relevant client history and insights from large volumes of documents.

Al in action: Al models are used to extract. summarize, and organize information from unstructured client data, including various forms, conversations, and reports. This makes the data easily searchable through natural language interactions, enables faster decision-making, reveals trends and patterns, and enhances the value of institutional knowledge across the organization.

#### **Automating document** processing

Scenario: A global manufacturer reduces manual effort and accelerates operations by automating document-intensive workflows, like invoice and claims processing.

Al in action: An Al agent classifies, extracts, and summarizes information from documents by using natural language prompts, making historical data easier to access and apply. By streamlining repetitive tasks, this solution improves individual employee productivity, saves time, and enhances the organization's operational efficiency.

# Building a foundation for your Al estate

Al-powered apps and agents offer many ways to add powerful new capabilities for your business. You also have many choices for how you build these Al-powered apps and agents. However, stitching together Al-based services with different application-hosting services and data infrastructure can be tremendously complex and often results in steep learning curves and costly, time-consuming development efforts.

Enterprise-grade Al applications do have some common requirements that affect your cloud technology choices, including:

- Robust AI services that support the full model lifecycle, including training, deployment, monitoring, and retraining, often at scale.
- Highly scalable, secure, and resilient cloudnative infrastructure to support continuous availability and global reach.
- Seamless access to high-quality, governed data sources.

## The benefits of an integrated Al application platform

An integrated Al application platform is a comprehensive environment that combines tools and services for building, deploying, and managing Al-powered applications from end to end. There are many benefits to building your Al apps on these types of platforms, including the ability to:

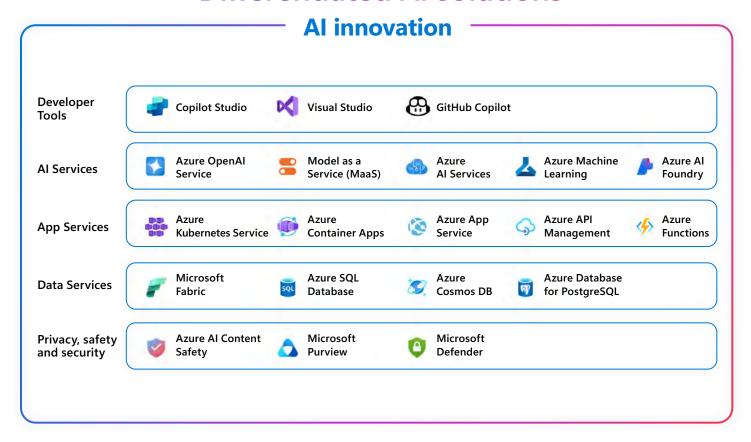
- Accelerate time to market. Go from idea to production faster with a streamlined platform that can reduce development timelines.
- Iterate faster with feedback. Test, deploy, and scale Al applications rapidly, refining them based on user insights to help ensure continual improvement.
- Securely integrate Al with enterprise data. Connect Al models to enterprise data sources for advanced personalization, search, and analytics, while maintaining robust security.
- Empower developers with AI tools.
  Enable developers to build AI apps faster using their preferred development tools for a seamless workflow.

Operationally, you need to ensure that your Al apps can reliably access the data and services they need while deploying and scaling quickly to meet demand. From a business perspective, you also need to manage costs and make sure that the solution is meeting your business goals. Additionally, you may be subject to

regulations, such as data sovereignty, privacy laws, bias mitigation, and AI safety standards.

To address these issues and help simplify the creation of Al-based solutions, you need an integrated Al application platform that provides the hosting, data, and Al services you need to quickly and efficiently build trusted, scalable, and performant Al solutions.

#### **Differentiated AI solutions**



# Create reliable and scalable AI solutions on Azure

Azure offers an integrated AI application platform that brings together a comprehensive suite of cloud services to support the full lifecycle of building, deploying, and managing AI-powered applications. It streamlines app development, data management, and AI services within a unified environment, so you can create AI-powered applications more efficiently.

High-performance AI applications depend on the app-hosting, data, and AI capabilities that your underlying infrastructure provides. Azure supplies your baseline for all three of these areas, with scalable, cloud-native application hosting, enterprise-grade data services and governance, and managed tools for building, training, and operating AI models.

#### A commitment to trustworthy Al

Microsoft is committed to ensuring that its Al services are secure, safe, private, and aligned with ethical principles. These efforts are organized around three main pillars:

- Security. The Microsoft Secure Future Initiative (SFI) ) reflects our company-wide commitment to security, with secure-bydesign principles implemented throughout the Microsoft internal development lifecycle. It also includes internal governance improvements, accountability, and a dedicated security workforce of over 34,000 engineers.
- Safety. The principles of responsible AI at Microsoft, established in 2018, guide the company in safely building and deploying AI, including properly developing, testing, and monitoring systems to avoid undesirable behaviors, such as harmful content, bias, misuse, and other unintended risks.
- Privacy. Another critical Microsoft priority is to help ensure that customer data is protected and compliant with Microsoft privacy principles, as itemized in Microsoft Trust Center: Security and privacy in the age of Al. These include user control, transparency, and legal and regulatory protections.

Azure Al Foundry directly support Al safety by integrating comprehensive tools and practices that align with principles of responsible Al. Features like Azure Al Content Safety and Risks & Safety monitoring can help to identify issues with Al output and to prevent problematic output or hallucinations from reaching end users. You can also configure custom guardrails and filters to meet your organization's specific needs.

#### **Build Al apps more easily**

When building Al apps, managed cloud Al services are critical because they support overall scalability, performance, development speed, and long-term maintainability. The right services provide a choice of models for every use case, including pretrained and open-source models. They also give you the tools for managing the full AI lifecycle—from data preparation and training to deployment and monitoring. Plus, these services help to ensure enterprise-grade security, compliance, and integration with existing systems.

On the Azure platform, Azure Al Foundry is a central hub for accessing the tools, services, and models you need to build and operate Al applications and agents. It provides easy access to powerful tools, such as Azure Al Foundry models, Azure OpenAl Service in Azure Al Foundry models, Azure Al Agent Service, Azure Al Search, and Azure Machine Learning, all from a single portal.

Providing a unified development environment that supports the entire AI app lifecycle, Azure Al Foundry offers a rich set of tools, SDKs, and APIs. It also helps streamline the development process, integrating with popular tools, such as GitHub and Visual Studio, and allowing easy access to complementary tools, like Microsoft Copilot Studio.

Azure Al Foundry can help you manage data integration, model training and customization, deployment orchestration, and experimentation. It also provides enterprise-grade governance

and management, and it integrates AI safety tools, helping to ensure that your Al apps are secure and compliant. By unifying data, models, and operations into a single platform, Azure Al Foundry radically simplifies the process of creating intelligent agents and integrated Al solutions that deliver real-world business benefits.

#### **Empowering your developers with Al**

Generative AI technology is not just a feature that's included in end-user apps. It also can help your developers build useful applications that solve complex business problems—faster and moe easily.

Within the Microsoft AI development ecosystem, developers can choose tools and environments to suit their skills and preferences and to reduce integration complexities. Al assistance, including the powerful GitHub Copilot, is seamlessly integrated into core development tools. Azure Al Foundry can also simplify your developers' work by offering broad support for an array of programming languages, models, and Al frameworks.

With Azure Al Foundry, developers can integrate available AI models into their apps using SDKs for Python, .NET/C#, Java, and JavaScript/TypeScript. Developers can also take advantage of many popular AI frameworks, including LangChain, LlamaIndex, Semantic Kernel, and AutoGen.

Other Azure Al Foundry features also integrate with your development cycle in multiple ways, further improving your dev processes. This can include automated testing and debugging, smart documentation, continuous integration/continuous delivery (CI/CD) optimization, and project management and planning.

By embedding intelligence throughout your development lifecycle, Azure Al Foundry can help reduce friction, shorten feedback loops, and create more seamless and productive development workflows.

#### Use a modern, managed app platform

Modern Al apps and agents are built around containerized microservices architectures. As a result, they're flexible, scalable, modular, and better suited for creating these kinds of Al-powered solutions.

On Azure, Azure Kubernetes Service (AKS) serves as the cloud-native app-hosting backbone that you need to deploy, scale, and manage Al-powered applications and agents with speed, flexibility, and reliability. AKS offers key features that AI-powered applications require, including:

- Scalable infrastructure. AKS can automatically scale compute resources up or down, based on need, so it's ideal for handling demanding enterprise Al workloads.
- Microservices and API management. Because it's Kubernetes-based, AKS supports containerized microservices architectures. Each component of your Al app can be deployed and managed independently as a container, so the system is modular and resilient.
- Integration with AI and machine learning pipelines. AKS integrates well with Azure Machine Learning for seamless model training, deployment, and monitoring. You can deploy machine learning models as containerized inference endpoints, retrain them in the background, and manage their lifecycle in production.

- Enterprise-grade security and governance. AKS includes built-in capabilities for identity management, role-based access control (RBAC), network policies, and integration with Azure Policy and Microsoft Defender for Containers, helping you to be sure that your Al workloads meet enterprise security and compliance standards.
- **DevOps readiness.** AKS can integrate into your organization's DevOps processes and into features such as CI/CD pipelines, canary deployments, and infrastructure-as-code (IaC) tools, like Terraform and Bicep.

For alternative app architectures, Azure offers other managed app platforms, including Azure Functions for serverless applications, Azure Spring Apps for Java-based Spring apps, and Azure App Service for traditional web apps.

#### Ensure that you have the robust data capabilities to feed your Al

Powerful database services are critical to storing, managing, and accessing the vast amounts of data that power intelligent systems. Al models rely on high-quality, well-organized data, both structured and unstructured, for training, validation, and real-time inference. Scalable data services help to ensure that data can be ingested from various sources, transformed, and retrieved efficiently, even under heavy workloads. They also support essential regulatory compliance and governance capabilities.

Azure Cosmos DB is ideally suited to power your Al applications, offering a flexible, enterprise-grade data platform that meets the demanding requirements of modern intelligent systems. Some of the Azure Cosmos DB capabilities that can help you effectively build your AI solutions include:

- Global distribution and low latency. Azure Cosmos DB delivers single-digit millisecond read and write latencies and is distributed globally for high availability and fast performance of AI apps used across different regions.
- Elastic scalability. Automatic scaling allows Azure Cosmos DB to handle fluctuating workloads, which is essential for Al applications that experience variable user traffic or inferencing demands.
- Support for multiple data models and APIs. Azure Cosmos DB supports document (SQL), key-value (Table, Cassandra), and graph (Gremlin) models, giving developers the flexibility to store and query data in the most appropriate format for their Al use case.
- Real-time data access. Azure Cosmos DB supports real-time ingestion and querying of data, which is critical for tasks like user personalization, chat history management, or IoT-driven Al.
- Seamless integration with Azure **services.** Integration with services like Azure Machine Learning, Azure Functions, and Azure Synapse Analytics let Azure



Cosmos DB support real-time inferencing, event-driven pipelines, and advanced analytics.

- Ideal for AI use cases, like Retrieval-Augmented Generation (RAG) and agents. Azure Cosmos DB serves as a fast-access data store for RAG and other Al workloads that require instant access to large, structured content.
- Enterprise-grade security and reliability. Enterprise-grade trust and compliance features are built into Azure Cosmos DB, providing your Al apps with builtin encryption, multiregion failover, and service-level agreement-backed (SLAbacked) guarantees for performance, availability, and consistency.

This combination of performance, flexibility, and integration makes Azure Cosmos DB a powerful data backbone for AI-powered solutions. However, Azure also offers support for common database technologies, including Azure SQL Database, Azure Database for PostgreSQL, Azure Database for MySQL, and Azure Database for MariaDB, for customers who need these options.

#### Take advantage of the wider capabilities of Azure

The services discussed in the previous section are all built on the powerful, globally distributed infrastructure of Azure, Modern Al applications, built around Azure services, can benefit from the many advantages that the Azure platform offers, including:

- **Service availability.** Most Azure services offer redundant systems and automated failover capabilities that help keep your workloads operational. With over 300 datacenters, the multiregional deployment options of Azure help maintain business continuity. Proactive platform maintenance and updates help minimize disruption to the services you depend on.
- Global reliability. Azure datacenters are distributed across more than 60 geographical regions around the world. This worldwide Azure infrastructure delivers consistent performance and offers region-specific availability zones to increase resilience. Geographic distribution of workloads also helps improve reliability.



- Performance SLAs. Microsoft guarantees 99.9% uptime across all services. Clear remediation paths and support commitments help address potential service degradation.
- **Dynamic scaling.** Azure services emphasize intelligent capacity management that adjusts to workload demands in real time. Automated resource allocation helps prevent performance bottlenecks.
- Support framework. Microsoft provides all day, every day enterprise support with defined response SLAs. Dedicated technical resources handle critical issues, and regular service health updates and maintenance notifications keep you informed.

# Start building the Al-powered future with Azure

The future of business is Al-powered, and you can start building that future today. Whether you're creating powerful customer-facing products, building Al apps to boost team productivity, designing custom Al agents to automate complex workflows, or empowering your developers with best-in-class tools, Azure gives you the integrated platform, flexibility, and trust to confidently embrace Al technologies.

Explore what's possible with Azure services, like Azure Al Foundry, Azure Cosmos DB, and AKS, and start building the intelligent solutions that can define your future.



**Explore Azure Innovate offerings** 

Talk to an Azure expert about your organization's needs.

© 2025 Microsoft Corporation. All rights reserved. This document is provided "as-is." Information and views expressed in this document, including URL and other Internet website references, may change without notice. You bear the risk of using it. This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes.