

Microsoft Partner Project Ready

# Agentic AI Accelerator

■ Day 2 of 4

<Presenter Name>



# Course plan and learning objectives

## Day 1

### Showcasing AI Potential with Agentic AI

#### Module 1: Innovate with Microsoft 365 Copilot and agents

- M365 Copilot and agents
- How Copilot works
- Semantic index for Copilot
- M365 Copilot Chat
- Agents Use cases
- New agents in Microsoft 365
- Unlock more value with SharePoint agents

#### Module 2: Extend Microsoft 365 Copilot with Agents

- Microsoft 365 Extensibility Planning and approach
- Declarative agents and agent tooling
- Explore Copilot Studio Agent Builder
- Build declarative agents with Microsoft 365 Agents Toolkit
- Build custom agents with Copilot Studio
- Autonomous agents overview
- Agent Governance - Overview
- Gen-AI decision guide – when to build, buy or extend

#### Hands-on Labs

- Explore Copilot Studio Agent Builder
- Build HR Assistant Agent with Copilot Studio
- Incorporate actions in HR Agent
- Enable Autonomous Capabilities in Microsoft Copilot Studio for HR Activities

## Day 2

### Architecting Success with Multi-Agent AI Systems

#### Module 3: Customize Agents with Gen AI in Copilot Studio

- Customizing your agents – Orchestrator, UI, Knowledge, Actions, Autonomy
- Copilot Studio implementation guidance for architects
- Generative AI in Copilot Studio
- Copilot Studio + Power Platform
- Building voice-enabled agents
- AI Foundry integration
- Developing agents using Microsoft 365 Agents SDK

#### Module 4: Innovate with Azure AI Platform

- How language models work
- AI Foundry and SDK introduction
- AI Foundry Model Catalog
- Azure AI Services
- Azure OpenAI Service and model guidance
- Models-as-a-Service
- Azure AI Foundry Agent Service
- Safeguard with Trustworthy AI

#### Hands-on Labs

- Setup AI Project and perform Chat Completion from VS Code
- Build a simple AI Agent
- Develop a multi-agent system

## Day 3

### Multi-Agent AI: Advanced Agent Dev in Azure AI Foundry

#### Module 5: Customize, orchestrate and experiment with Azure AI Foundry

- Retrieval Augmented Generation (RAG)
- Customizing models – Fine tuning, distillation
- Responses API (preview)
- Azure AI Foundry Agent Service - Orchestrate and debug AI workflows

#### Module 6: Build your own multi agents with Semantic Kernel or AutoGen

- Multi-agent applications
- Understanding Semantic Kernel
- Understanding AutoGen Agents Framework
- Multi-Agent Collaboration & Orchestration with AutoGen / Semantic Kernel

#### Hands-on Labs

- Set Up Azure AI Foundry SDK and Provision Resources
- Build a Retrieval-Augmented Generation(RAG) Pipeline
- Evaluate and Optimize RAG Performance
- Semantic Kernel Fundamentals
- Semantic Kernel Plugins

## Day 4

### Enterprise Grade: Optimization and production at scale

#### Module 7: Enterprise grade production at scale

- Scaling challenges and agent controls
- Manage AI performance in production
- Observability Tools
- Enabling Enterprise governance and management
- Enterprise grade security and data protection
- Monitoring and observability

#### Module 8: Advanced AI risk evaluation and mitigation

- Identifying risks
- Azure AI Content Safety
- Evaluation and GenAIops
- Identity and access management
- Network Security for AI apps
- Continuous security for AI

#### Hands-on Labs

- Understanding the Lifecycle of Flow Development
- Building and Customizing Prompt Flows
- Evaluation Flow Setup
- Fine-Tuning Prompts for Optimal Performance
- Implementing Chat Flow and Tool Integration
- Ensuring Responsible AI Practices with Content Safety

Journey A – Deal-ready

Deal-ready assessment 

Journey B – Solution Design

Solution Architecture assessment 

Journey C – Project-ready

Capstone Project 



# Architecting Success with Multi-Agent AI Systems

⌚ 2 Days (5 hours/day)



⌚ 2 hours

## Module 1: Innovate with Microsoft 365 Copilot and agents

- M365 Copilot and agents
- How Copilot works
- Semantic index for Copilot
- M365 Copilot Chat
- Agents Use cases
- New agents in Microsoft 365
- Unlock more value with SharePoint agents

## Module 2: Extend Microsoft 365 Copilot with Agents

- Microsoft 365 Extensibility Planning and approach
- Declarative agents and agent tooling
- Explore Copilot Studio Agent Builder
- Build declarative agents with Microsoft 365 Agents Toolkit
- Build custom agents with Copilot Studio
- Autonomous agents overview
- Agent Governance - Overview
- Gen-AI decision guide – when to build, buy or extend



⌚ 3 hours

## Hands on labs

- Explore Copilot Studio Agent Builder
- Build HR Assistant Agent with Copilot Studio
- Incorporate actions in HR Agent
- Enable Autonomous Capabilities in Microsoft Copilot Studio for HR Activities



⌚ 2 hours

## Module 3: Customize Agents with Gen AI in Copilot Studio

- Customizing your agents – Orchestrator, UI, Knowledge, Actions, Autonomy
- Copilot Studio implementation guidance for architects
- Generative AI in Copilot Studio
- Copilot Studio + Power Platform
- Building voice-enabled agents
- AI Foundry integration
- Developing agents using Microsoft 365 Agents SDK



⌚ 3 hours

## Hands on labs

- Setup AI Project and perform Chat Completion from VS Code
- Build a simple AI Agent
- Develop a multi-agent system

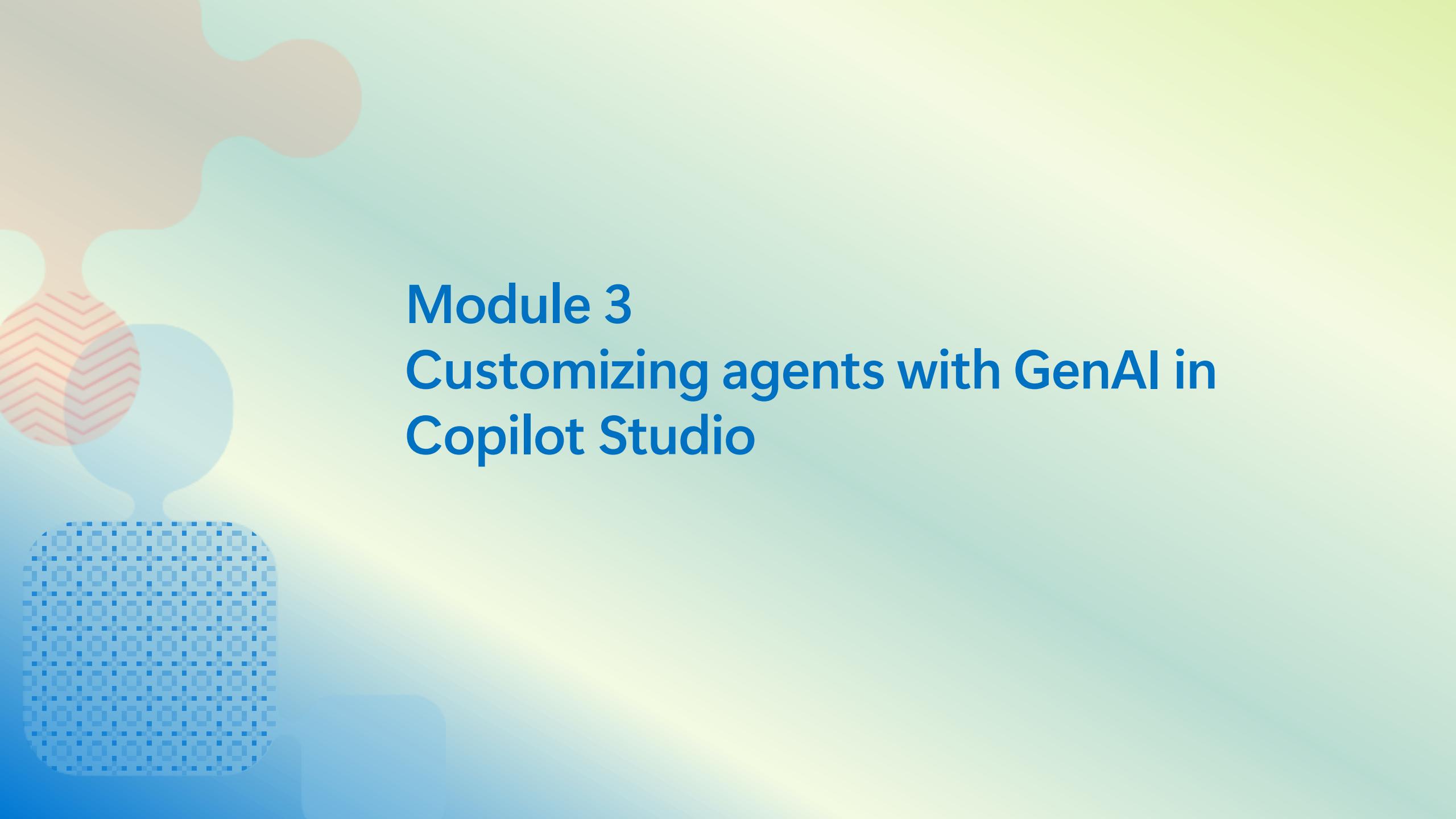


## Solution Architect Assessment Purpose:

Evaluate the Solution Architect's competency in rapidly designing, architecting, and clearly communicating technical solutions leveraging Microsoft Copilot, Azure AI Foundry, Semantic Kernel, AutoGen, and related technologies.

Live- Instructor led

On demand



# **Module 3**

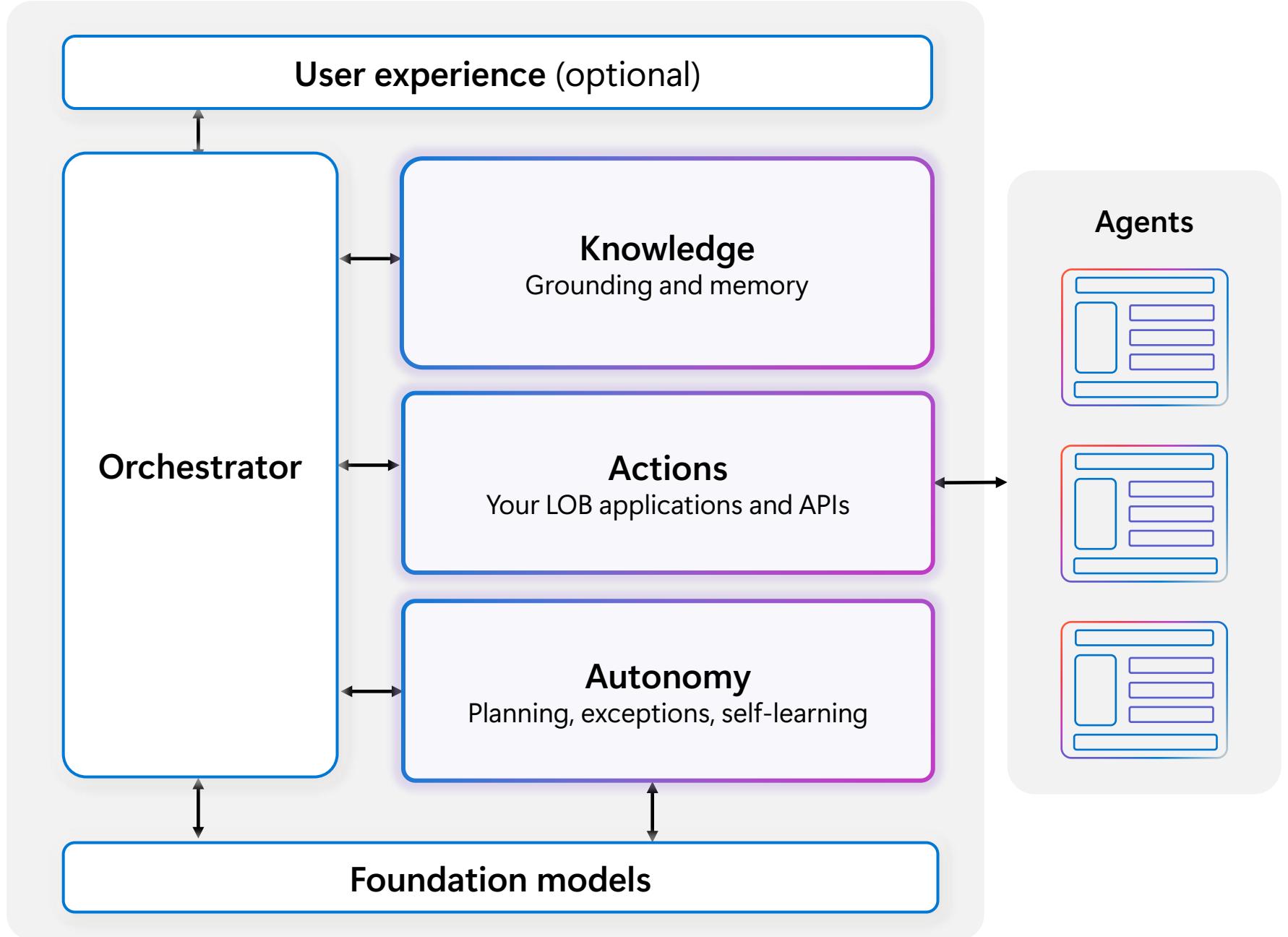
## **Customizing agents with GenAI in Copilot Studio**

# Customizing agents with Copilot Studio



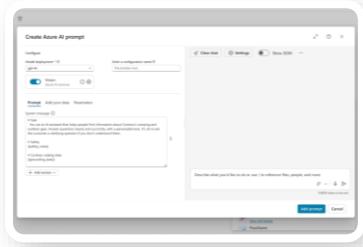
# Customize your agent

Easily tailor your agent's building blocks to meet your unique business needs in a comprehensive, end-to-end studio



**Design, enhance, and manage agents**  
Build an agent and **go live quickly**, all from one  
easy-to-use, E2E SaaS product

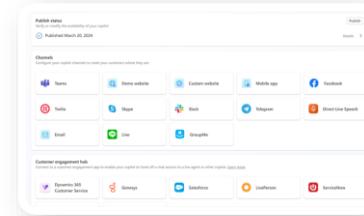
**Integrate with AI services**  
Integrate with Azure AI Studio,  
Azure Cog Services, Bot  
Framework and other Microsoft  
conversational services



**Monitor and improve**  
Access rich out-of-the-box  
insights and **analytics**



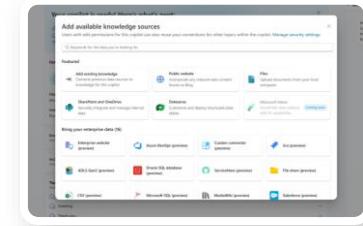
**Publish to multiple channels**  
Publish in 20+ languages and  
deploy to your channel of choice  
with a single click, including  
**Microsoft 365 Copilot**



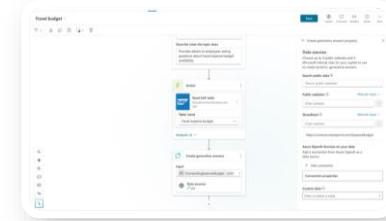
## Build & Publish



**Copilot Studio**

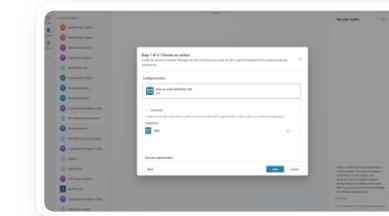


**Chat over knowledge with gen AI**  
Get enterprise-specific answers using  
your files, websites, Dataverse, third-  
party apps and more



## Create specific topics

Supplement gen AI responses  
with **specific, curated topics** for  
when you want more control

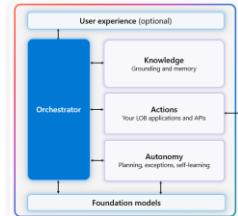


## Analyze & Improve

## Build actions

Create actions and use **pre-built  
connectors** or **agent flows** to  
call your backends and APIs or  
complete long-running tasks

# Orchestrator



Easily mix and manage both **generative** and **custom orchestration** in one agent

Have **complete control over critical scenarios** by designing specific step-by-step topics.

Enable your agent to **automatically select the most appropriate action or topic** to respond to a user using generative AI.

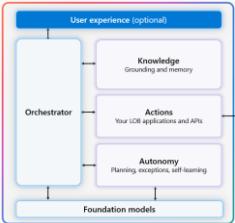
**Connect to your contact center** so your agent can escalate and hand off the conversation with full context to a human agent.

The screenshot shows the Microsoft Copilot Studio interface for creating a generative flow. The main workspace displays a flowchart with three main steps:

- Describe what the topic does:** Provides details to employees asking questions about travel expense budget availability.
- Action:** Read SAP table (Nora@northwindtraders.com, SAP). Table name: Travel expense budget.
- Create generative answers:** Input: OutstandingExpenseBudget table; Data sources: https://contoso.sharepoint.com/ExpenseBudget.

On the right side, there are panels for managing data sources and connections:

- Data sources:** Choose up to 4 public websites and 4 Microsoft internal sites for your copilot to use to create dynamic, generative answers. Includes sections for "Search public data" and "Public websites" (Manual input).
- Azure OpenAI Services on your data:** Add a connection from Azure OpenAI as a data source.
- Custom data:** Enter or select a value.



# Meet you where you are

Deploy to...



Microsoft Copilot



Your applications



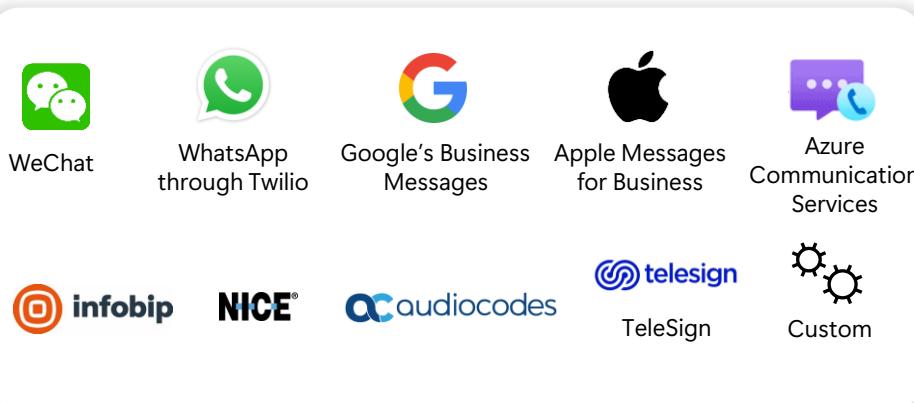
Your websites

# Channels

Publish and deploy to your channels of choice with a single click.

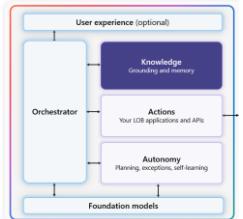
Add your agent to a custom app built with Power Apps or a custom website built with Power Pages.

You can access even more channels through ISVs, including:



The screenshot shows the 'Expense Budget Assistant' application interface. At the top, there are tabs: Overview, Knowledge, Actions, Topics, Analytics, Channels (which is underlined), and a user icon. A green success message states: "Your copilot was published at 9:42 AM on 3/36/2024! Your users will see the new content soon." Below this is a 'Publish status' section with a checked checkbox for 'Published March 20, 2024' and a 'Publish' button. The main area is titled 'Channels' with the sub-instruction: 'Configure your copilot channels to meet your customers where they are.' It features a grid of 15 channel icons: Teams, Demo website, Custom website, Mobile app, Facebook, Twilio, Skype, Slack, Telegram, Direct Line Speech, Email, Line, and GroupMe. At the bottom, there is a 'Customer engagement hub' section with icons for Dynamics 365 Customer Service, Genesys, Salesforce, LivePerson, ServiceNow, ZenDesk, and Customer engagement hub.

# Knowledge



Add your **public and enterprise data** sources using agent connectors.

Your agent will be able to **dynamically generate multi-turn answers** in real time using your enterprise data.

Allows you to create an **immediately useful agent**.

Supported data sources include:

Public websites      SharePoint / OneDrive

DataVerse            Microsoft Fabric (*coming soon*)

File uploads        Microsoft Graph connectors

**Add available knowledge sources** (Powered by Copilot connectors)

Users with edit permissions for this copilot can also reuse your connections for other topics within the copilot. [Manage security settings](#)

Keywords for the data you're looking for

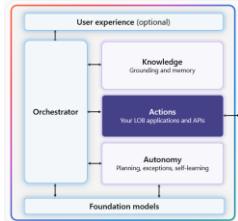
**Featured**

- Add existing knowledge**  
Converts previous data sources to knowledge for this copilot
- Public website**  
Incorporate any relevant web content found on Bing
- Files**  
Upload documents from your local computer
- SharePoint and OneDrive**  
Securely integrate and manage internal data
- DataVerse**  
Customize and deploy structured data tables
- Microsoft Fabric**  
Accelerate data analysis with AI capabilities

**Bring your enterprise data (16)**

Enterprise website (preview)	Azure DevOps (preview)	Custom connector (preview)	Jira (preview)
ADLS Gen2 (preview)	Oracle SQL database (preview)	ServiceNow (preview)	File share (preview)
CSV (preview)	Microsoft SQL (preview)	MediaWiki (preview)	Salesforce (preview)
Confluence (preview)	Azure SQL (preview)	Zendesk (preview)	Power Platform connector

# Actions



Easily connect to your key line of business systems.

Enable your agent to automate your business processes and complete tasks.

Tell your agent when to execute an action by adding autonomous triggers.

## Types of actions:

<b>Prebuilt connectors</b>	Choose from 1500+ prebuilt Power Platform connectors to popular data sources and apps
<b>Custom connectors</b>	Create a custom connector for any publicly available API
<b>Agent Flows</b>	Create automated workflows
<b>Prompts</b>	Provide custom instructions to the GPT model using AI Builder
<b>Skills</b>	Add a bot built using Azure Bot Framework as a skill

## Step 1 of 3: Choose an action

Create an action or browse through our list of actions you want to use to get information from external sources.

[Learn more](#)

---

### Discover an action

Search for flows, skill actions, and commonly used connector actions

11 actions found

ConnectorsCustom ConnectorsFlowsSkillsDataverse



Untitled

Get information about industries, solutions, services and cont...



Run a flow built with Power Automate for desktop

Desktop flows



Delete a row

Excel Online (Business)



Run script

Excel Online (Business)



Get a row

Excel Online (Business)



Run script from SharePoint library

Excel Online (Business)

Cancel



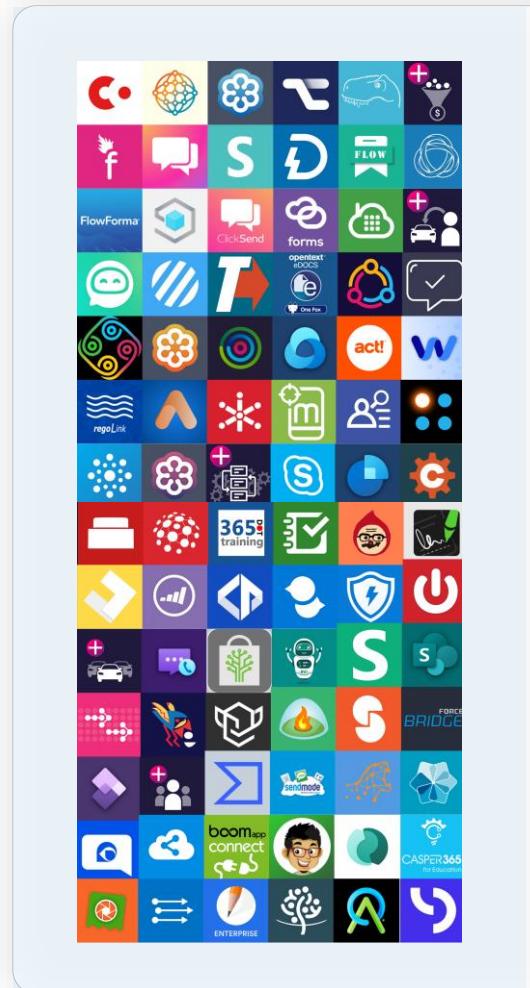
# Agent flows and connectors

Use agent flows to build, edit, and extend automation with 1,500+ pre-built connectors or custom APIs

For a smooth user experience, agent flows triggered with your agent to execute quickly to minimize delays in response times

Agent flows from Copilot Studio have a 100-second limit to return results before timing out

Make HTTP requests directly from Copilot Studio to bypass agent flows and enhance efficiency



1500+ prebuilt data connectors



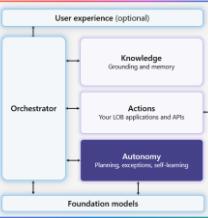
Copilot  
Studio

Multi-Channel

-  Website
-  Mobile
-  Teams
-  Facebook
-  Telephony
-  Email
-  Slack
-  Direct line Speech
-  Microsoft Power Platform
-  Many more

# A spectrum of agents

Agents vary in levels of complexity and capabilities depending on your need



## What are autonomous agents?

Expert systems that **manage and automate complex business tasks**, enhancing efficiency and innovation across your organization.

### Proactive

Automatically responds to signals across your business and initiate tasks, configured to react to events or triggers without human input.

### Independent

Seamlessly completes tasks behind the scenes without human intervention.

### Adaptive

Non-deterministic and able to produce different outcomes from the same initial conditions.



# Copilot Studio implementation guidance for architects

Guidance for key stakeholders from the customer and partner: project leads, solution architects, functional and technical leads



# Success by Design



Success by Design is Microsoft's framework for Power Platform implementations

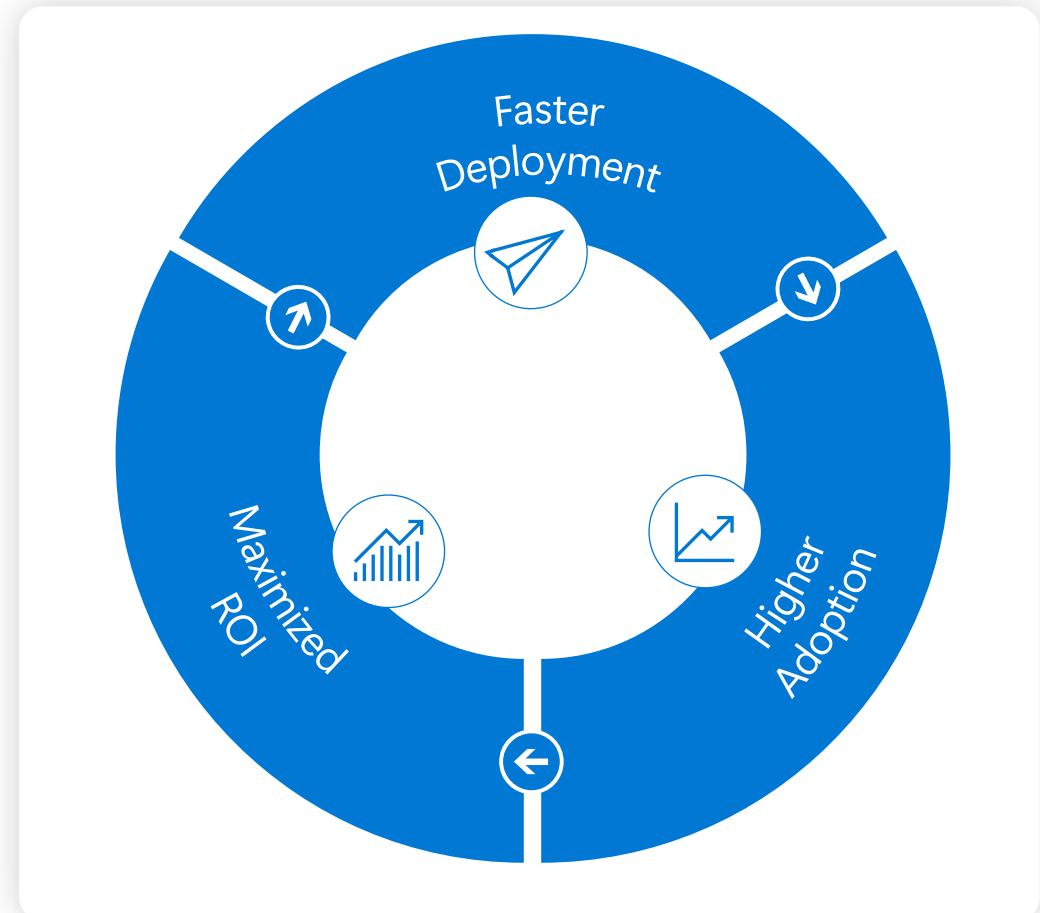


The framework is designed around **3 principles**:

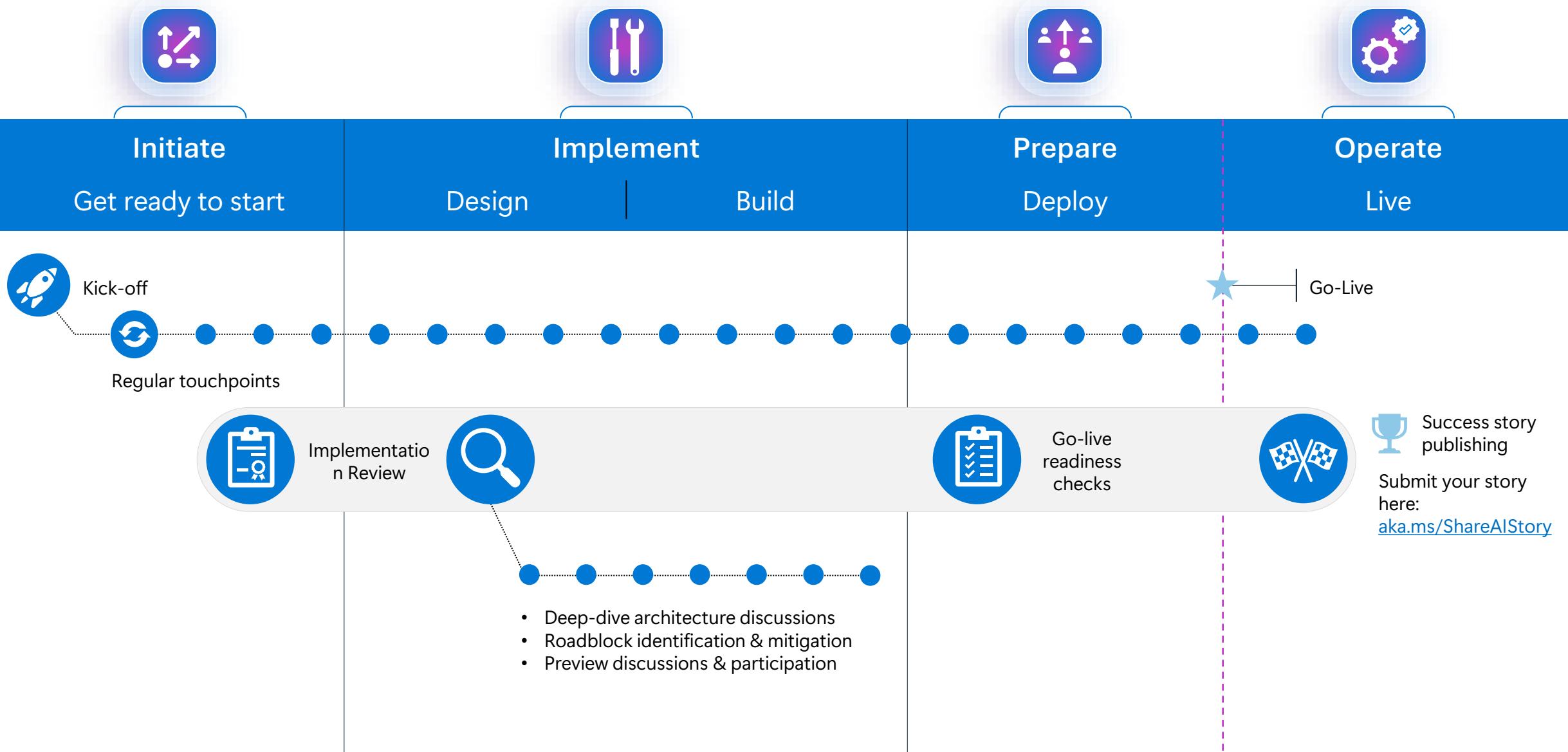
- Early Discovery
- Proactive Guidance
- Predictable Success



Success by Design brings the **learnings** and **experiences** from thousands of customer deployments to make your journey to the cloud **smoother, faster** and **successful**.



# Typical Success by Design engagement lifecycle



# Architecting a Solution

- Business overview
- Project overview & agent purpose
- Stakeholders
- Project planning
- Target KPIs
- Assumption & concerns



Download full implementation guide

The image shows the front cover of the 'Copilot Studio Implementation Guide'. The cover is black with the Microsoft logo at the top left. Below it is a blue stylized 'C' logo followed by the text 'Copilot Studio' and 'Implementation Guide'. There is a dashed box labeled 'Customer and project name' for customization. At the bottom right is the 'Power CAT' logo. To the right of the book cover is a photograph of three people (two men and one woman) sitting around a light-colored wooden conference table in a modern office setting. They are looking at documents and a laptop, engaged in a discussion. A potted plant is visible in the background.

# Project overview & agent purpose

Provide an overview of the project and the business problems the agent is trying to solve

## Example answer (you can delete this)

Enterprise information is disseminated across the organization and maintained in silos that different teams are responsible for. An internal employee survey revealed a high dissatisfaction about the time spent to find answers internally.

The Employee agent is planned to be available to all 40k Contoso Corporation employees in Microsoft Teams as an app and in the global internal portal as a chatbot.

The agent is expected to leverage AI and existing knowledge sources to answer employee questions on HR, IT, Sales and Finance topics. Answers must be grounded and sourced in enterprise data locations (SharePoint, ServiceNow, Dynamics 365, SAP).

Employees can also use the agent to perform common actions such as unlocking an account or booking meeting rooms.

## Why do we ask these questions?

- Understanding the project and what problem your agent is trying to solve is key to make relevant recommendations or share similar implementation patterns.
- Knowing who your end-users are will also be useful to understand the agent deployment channels requirements.

## Information we're looking for:

- What are the current business challenges that this project will help address?
- What is the purpose and main features of the agent?
- Who are the end-users of the agent?
- Is this a migration project?

# Other details to gather



## Key stakeholders

- Customer, Partner and Microsoft teams
- This helps us know who the key stakeholders are for the projects.



## Project Timeline

Understanding the project key milestones and planned go-live dates will help setup meetings accordingly.



## Target KPIs

- What are the success metrics for the project, and what are their targets?
- How is the agent expected to help your end-users and lower the support cost?
- How is end-user satisfaction expected to increase with the agent?



## Assumptions & concerns

- About the capabilities of your agent
- This will help you either validate or address them early in the project.
- Example: Limited control of the answers generated by a large language model (LLM).

# Architecture Overview



Solution architecture



Conversation volumes



Performance



Identified technical challenges

# Copilot Studio Overview

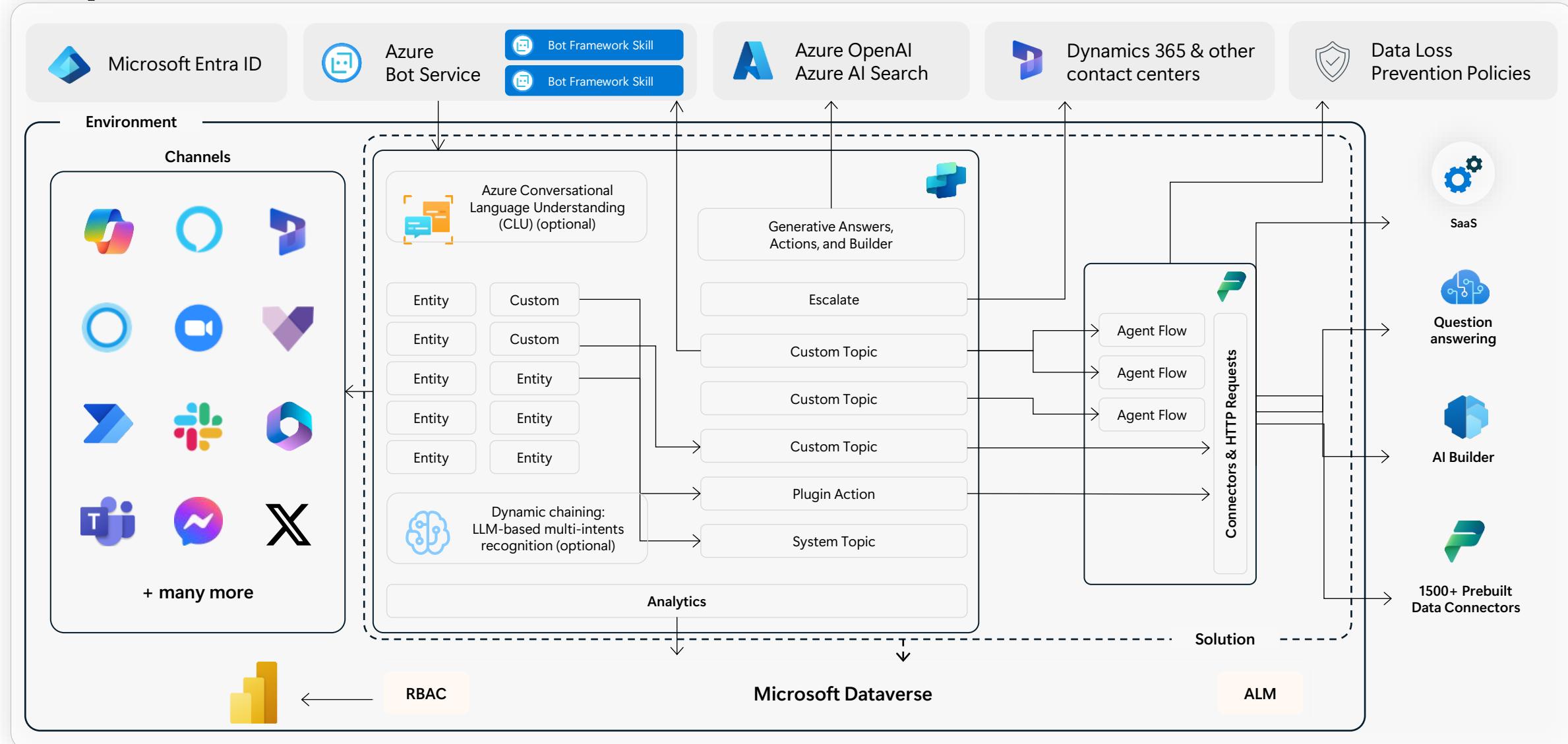
## Product capabilities

Intuitive UX	Visual canvas	Low code design	Pro code views	Real-time testing	Easy collaboration	Natural language to build
Conversation Design	Build + publish custom actions	Tailor specific topics	Create rich + dynamic responses	Multi-language	Multi-channel	Templates
Conversation Orchestration	Multi-turn conversations	Logic / variable management	Escalate to live agent	Multi-LLM routing	Dynamic content based on user	
Data Connectivity	Deep reasoning	1,400 pre-built data connectors	Custom data connectors	Agent flows	Generative actions	Generative answers
Pro-Dev Extensibility	DIY gen AI	Bring your own Model (BYOM)*	Custom Azure Bot Framework Skills	Knowledge base extension	Custom analytics	Azure app insights telemetry
Streamlined Management	Responsible AI checks	Trusted platform	Admin center	Compliance standards	Analytics	ALM automation
End-to-end conversational AI platform						

Why do we share this example?

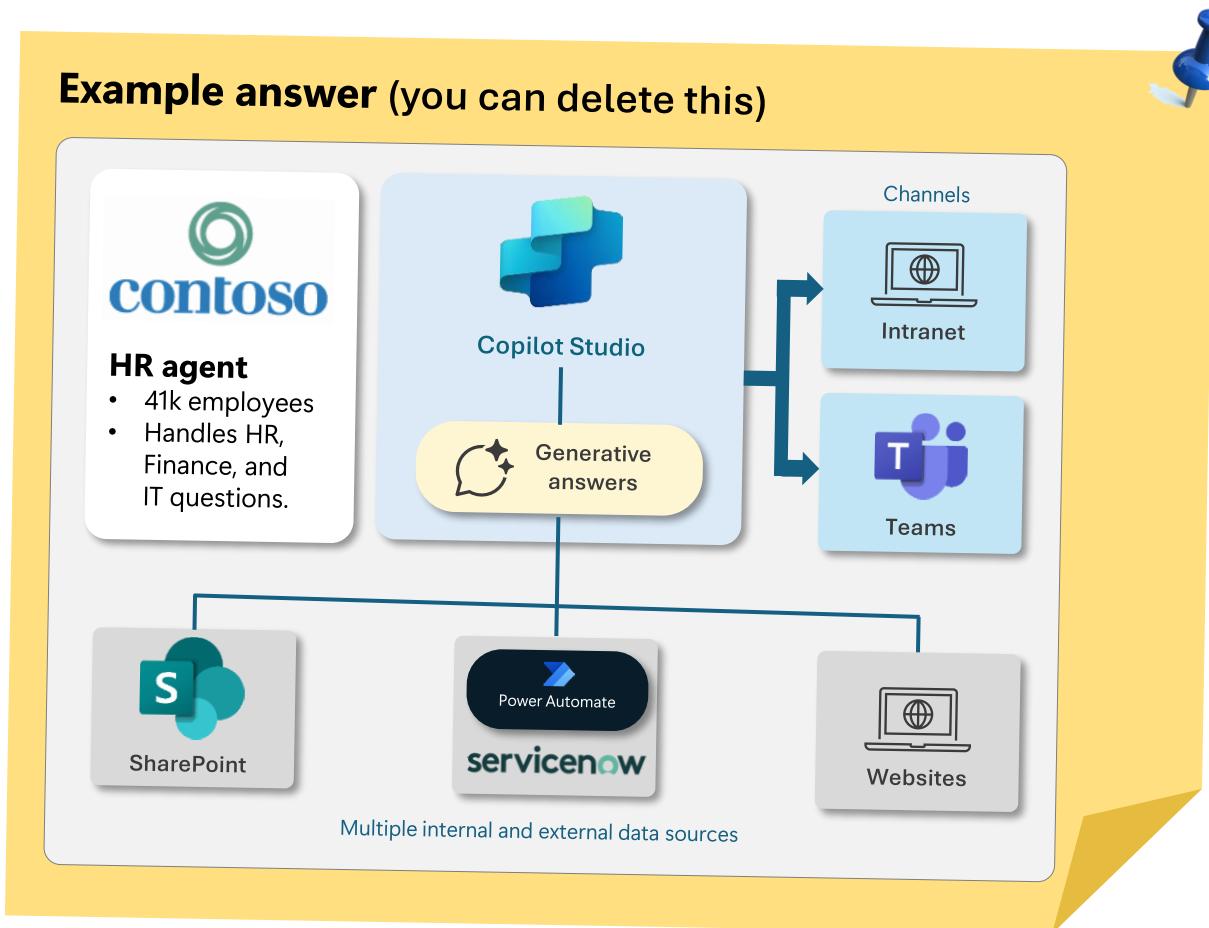
Overview of the native capabilities of Copilot Studio to help you assess the features you're planning to use in the project.

# Copilot Studio architecture



# Solution architecture

Provide an architecture diagram with the overview of the technical implementation



## Why do we ask these questions?

To have the big picture of your agent project and its integration in a broader technical and functional landscape.

## Information we're looking for:

- Architecture diagrams or blueprints
- Functional and technical overview of the agent and its integration in the broader technical landscape (other Power Platform or Azure services, internal APIs, etc.)

# Conversation volumes

## What are the expected volumes of chat messages or conversations?

Monthly target volume of :

- Total messages
- Generative actions
- Generative answers
- Power Platform requests (through connectors or cloud flows)

Expected seasonality impact

Expected maximum peak of concurrent conversations

### Example answer (you can delete this)

200,000 chat sessions per month (with an average of 8 messages per session), including 100,000 generative answers, 50,000 generative actions, and 1,000,000 Power Platform requests.

### Example answer (you can delete this)

800,000 monthly chat sessions in December and January.

### Example answer (you can delete this)

Maximum of 5,000 concurrent chats when sales season begins.

## Why do we ask these questions?

- Understanding target volumes helps validate the target architecture and scale.
- When integrating with external systems, for example through Power Automate or HTTP requests, it's important to validate that every part can handle the load.
- Target volumes also help validate the licensing aspects of the agent and the potential impact on Dataverse storage for the conversation transcripts.

## Useful resources:

- [Rate limits for agents](#)
- [Pricing plans](#)
- [Microsoft Power Platform Licensing Guide](#)

# Performance

## Detail the expected agent response times

First chat load time and first message expectation

### Example answer (you can delete this)

5 seconds for the web chat control to load and for the initial messages to be displayed to the user.

Expected maximum latency for the agent to answer user queries

### Example answer (you can delete this)

3 seconds after a user provides an input.

Approach for handling long-running actions (e.g., waiting for an external system to return data)

### Example answer (you can delete this)

A specific message is displayed to the user asking them to wait while the cloud flow runs and returns data. Typing indicator shows that the agent is doing something.

## Why do we ask these questions?

While there are no guaranteed service level agreements (SLAs) for Copilot Studio response times – mainly because they depend on complexity, integration patterns, etc. – it's still important to understand performance expectations for the agent response times to make sure the agent design optimizes for performance.

## Useful resources:

[Capture telemetry with Application Insights](#)

# Identified technical challenges

Please share the list of technical challenges or roadblocks you have already identified

Challenge #1

## Example answer (you can delete this)

Connecting to on-premises resources that are not exposed to the public internet.

Challenge #2

## Example answer (you can delete this)

Deploying to a WhatsApp channel.

Challenge #3

## Example answer (you can delete this)

Allowing agent end-users to download the conversation transcript at the end of the chat session.

## Why do we ask these questions?

To understand the challenges that have already been identified in the architecture that could get in the way of the agent success and for which it's important to define mitigation plans.

# Natural language understanding



## Standard NLU model

- ✓ Default, out-of-the-box, model that comes pre-trained, with many predefined entity types.
- ✓ Configuration is done by adding trigger phrases and custom entities (either closed lists with values and synonyms, or regular expressions).



## Custom Azure CLU model

- ✓ Supports additional languages, with native models.
- ✓ Allows to further customize the intent triggering model for better intent recognition or to address specific industry requirements.
- ✓ Allows for complex entity extraction (e.g., of the same type).
- ✓ Entity extraction can also leverage Copilot Studio standard NLU.



## Dynamic chaining

- ✓ Uses a GPT large language model.
- ✓ Can handle multiple intents and chain topics and/or plugins.
- ✓ Automatically generate questions for missing inputs and answers complex entities and questions from the conversation context.
- ✓ Configuration is done by simply describing topics, actions, and inputs/outputs.

A red circular icon with a red checkmark inside, indicating a negative or cautionary point.	<ul style="list-style-type: none"><li>✓ Single intent recognition per query.</li><li>✓ Cannot be extended.</li><li>✓ Slot-filling multiple entities of the same type in the same query requires disambiguation for each (e.g., from and to cities)</li></ul>
---	--

A red circular icon with a red checkmark inside, indicating a negative or cautionary point.	<ul style="list-style-type: none"><li>✓ Single intent recognition per query.</li><li>✓ Configuration is done in Azure and involves additional costs.</li><li>✓ Has its own service limits that need to be evaluated.</li><li>✓ Azure CLU intents and Copilot Studio topics must be carefully kept in sync.</li></ul>
---	--

A red circular icon with a red checkmark inside, indicating a negative or cautionary point.	<ul style="list-style-type: none"><li>✓ As it's a generative AI feature, the licensing burn rate of messages is higher than in regular topic triggering.</li></ul>
---	--

# Questions to discuss with customers



## Topic design

- What is your approach to designing topics?
- Do you plan to use disambiguation topics?
- How are you avoiding duplicating topic content multiple times?



## Languages

- What languages and markets should your agent support?
- Are you planning for a single multi-lingual agent or for one agent per language?
- Should the translations be set during configuration or real-time (i.e., provided at runtime)?



## Unrecognized intent

- How are you planning to manage unrecognized intents?
- Are you integrating with an external system as part of fallback? If yes, how?
- What is the expected % of conversations hitting fallback?



## Other

- Do you have specific requirements on session timeout (i.e., how long a agent session can last)?
- Do you have other requirements or questions on conversation design?

# Topic structure

## Creating and designing efficient topics

What is your approach to designing topics?

Do you plan to use disambiguation topics?

How are you avoiding duplicating topic content multiple times?

### Example answer (you can delete this)

A few custom topics for key scenarios with relevant trigger phrases and redirects, with a parent-child topic structure. Unrecognized intents will trigger generative answers and fallback.

### Example answer (you can delete this)

Yes, for user account operations with clarifying questions on the operation (e.g. create, unlock, suspend, etc.) and system (e.g. SAP, ServiceNow, Microsoft, etc.)

### Example answer (you can delete this)

Whenever a dialog path needs to be repeated, creating reusable topics that can be called by a parent topic before resuming the parent topic conversation logic once the child topic is complete.

### Why do we ask these questions?

Topics are discrete conversation paths that, when used together, allow for users to have a conversation with a agent that feels natural and flows appropriately.

While there's no one size fits all, given how topics can be triggered, it's a good practice to distinguish between:

- ✓ **Topics that will trigger based on user utterances.** These can almost be seen as your entry points topics. If you have trigger phrases that overlap multiple topics, consider having a catch-all topic and then redirect to other topics after clarifying questions. With entity extraction and slot filling, clarifying questions can be skipped if already answered.
- ✓ **Topics that will trigger when called from a redirect action, activity or event.** These can be called by multiple topics and can have input and output variables. They're ideally reusable, bite-size, topics.
- ✓ **A topic can also be both,** triggered through intent recognition or by an explicit redirect.
- ✓ **Conversational boosting and fallback:** topics that trigger when no matching topic is triggered based on a user query.

### Useful resources:

- [Topics best practices](#)

# Generative AI in Copilot Studio



# Generative AI in Copilot Studio

Powered by Azure

CONNECT AND EXTEND



M365 Copilot



Azure AI  
Foundry



Power Apps



Power Pages



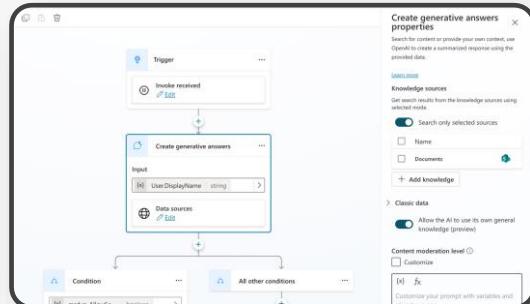
Teams



Dynamics 365

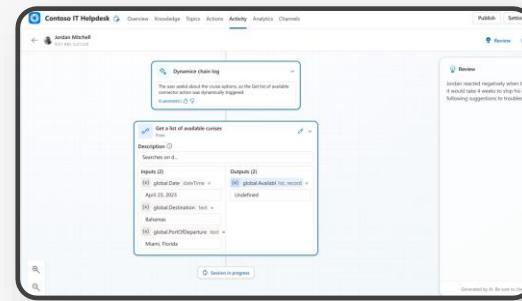
## Generative Conversations

### Generative Answers



Copilot Studio boosts conversational coverage by dynamically generating multi-turn answers based off an organization's content in real-time in response to unanticipated questions.

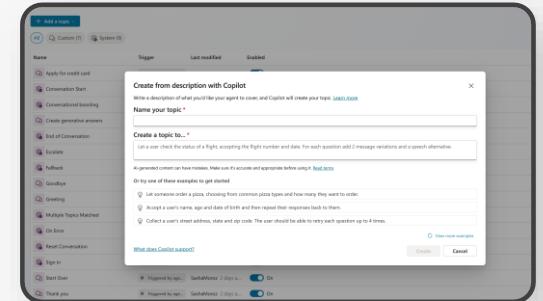
### Generative Actions



Copilot Studio generates dialog and takes action by dynamically chaining existing building blocks which can handle queries that were not anticipated or previously built.

## Generative Assistance

### Generative building



Assisted authoring to build, design and modify agents through natural language.

### External

3P Search APIs

External URL

### Internal

OneDrive URL

Documents

SharePoint URL

### Tools

Power Platform Connectors

Topics

Skills

Agent Flows

### Building

Topics

Trigger Phrases

Entities

Variables

Adaptive Cards

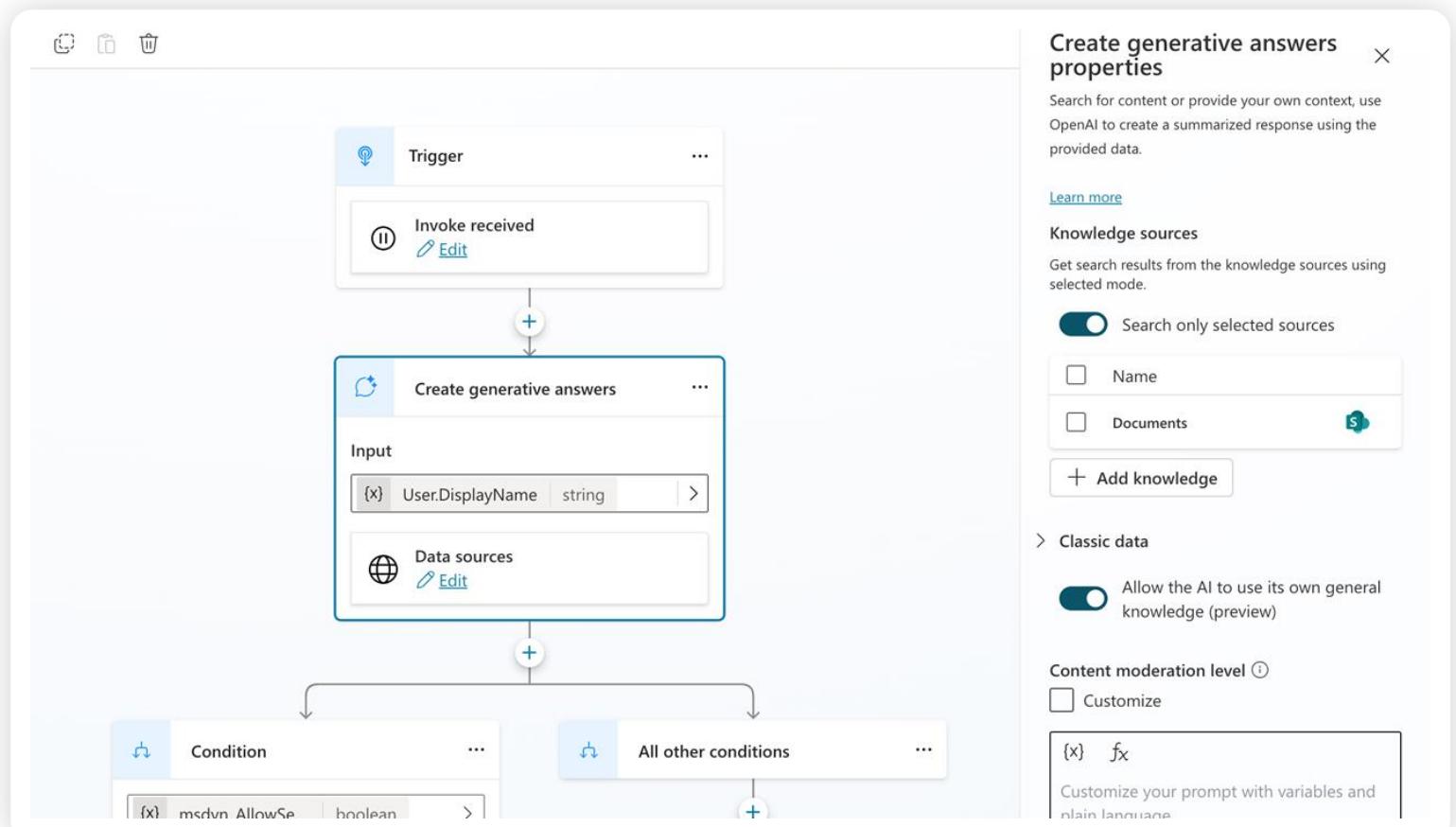
Branching

# Generative answers

Enable multi-turn chat over your own internal and external knowledge sources and sites with generative answers

Agents can answer thousands of questions out of the box in seconds

You can even pull data from an API or other backend system and enable generative chat over it



# Generative answers

Dynamically generate answers real-time based on content you choose

What data sources are you using for generative answers? Do you use custom instructions?

Where, in your topic structure, do you leverage generative answers?

Do you have more advanced scenarios that require your own large language models (LLMs)?

## Example answer (you can delete this)

Yes, over public websites, internal SharePoint sites, and ServiceNow knowledge bases.

Instructions to return friendly and brief answers (40 words max).

## Example answer (you can delete this)

Mostly in custom topics that will contain trigger phrases and entity questions that will direct the user towards the right data source. Conversational boosting will use public FAQ websites.

## Example answer (you can delete this)

Yes, as a last resort and to handle chitchat and unrelated queries, the agent will be able to answer in a ChatGPT manner (e.g. help write me a SQL query, etc.)

## Why do we ask these questions?

- By default, generative answers is available in the 'Conversational boosting' system topic that triggers just before 'Fallback', when an intent isn't recognized.
- Generative answers nodes can however be used throughout your custom and system topics, and the data sources be different and even set dynamically.

## Useful resources:

- [Generative answers](#)
- [Generative answers node](#)
- [Generative answers with Bing Custom Search \(video\)](#)
- [Generative answers with custom data – ServiceNow \(video\)](#)
- [Generative answers with uploaded documents \(video\)](#)
- [Customize responses with generative answers \(video\)](#)



# Prompt building with text & generative AI tools

A single interface to build, design, and test generative AI prompts that can be embedded across products

Prompts are saved in the prompt library and become a platform concept (supporting sharing, ALM, RBAC)

Prompts can be created and invoked in agents via topics or actions

Key capabilities:

- Multi-modal (images, files etc.)
- Low-code multi-model selector (40 mini, 40)
- Native Dataverse table data
- Templates available out-of-the-box

### Add a prompt action

Action details

Create prompt

Select action parameters

Review and test

Publish

**Prompt**

Extract the specified / provided as a comma-separated list, and match it up with the `Account`.`Account Name` , paying special attention to preserving listing them at the same time. For each entity, numerical amount all data. For example:

```
{  
  "extracted_data": {  
    "[entityToBeExtracted": entityToBeExtracted,  
    "[entityToBeExtracted": input text,  
    "[entityToBeExtracted": Data used (preview),  
    // Continue listing entities  
    ...  
  }  
}
```

**Inputs**

`Text`

`Image or document (preview)`

`entityToBeExtracted`

`input text`

`Data used (preview)`

`Account`

**Test prompt**

All inputs and data are available for this prompt.

**Prompt response**

Waiting for response generation.

**Back**



# Deep reasoning in prompts

Instruct agents to perform complex reasoning tasks

## Scenario:

Customer support agents need to analyze complex troubleshooting data (error logs or past support ticket) to pinpoint the root cause of an issue.

## Solution:

With deep reasoning in prompts, agents can rapidly identify patterns, suggest tailored solutions, and offer clear next steps for resolution.

A screenshot of the Copilot Studio interface. The main window title is "Supply Chain Prompt". On the left, there's a sidebar with icons for Home, Create, Agents, Flows, Tools, and an ellipsis. The main area has a "Prompt" section with a text input field containing placeholder text "Write or paste your text here." and a button "+ Add" with a "Test prompt" link. Below this is a modal titled "Write your prompt with Copilot" with a text input "Describe what your prompt should do" and several AI tool buttons: "Summarize text", "Extract information from text", "Classify text", "Sentiment analysis of text", "Respond to a complaint", and "Prompt library". To the right, there's a "Prompt response" section with the message "Waiting for response generation." and a note at the bottom: "AI-generated content may be incorrect. Make sure it's accurate and appropriate before using it. [Read terms](#)". At the bottom right are "Save" and "Cancel" buttons.

# Integrations



# Planning integrations

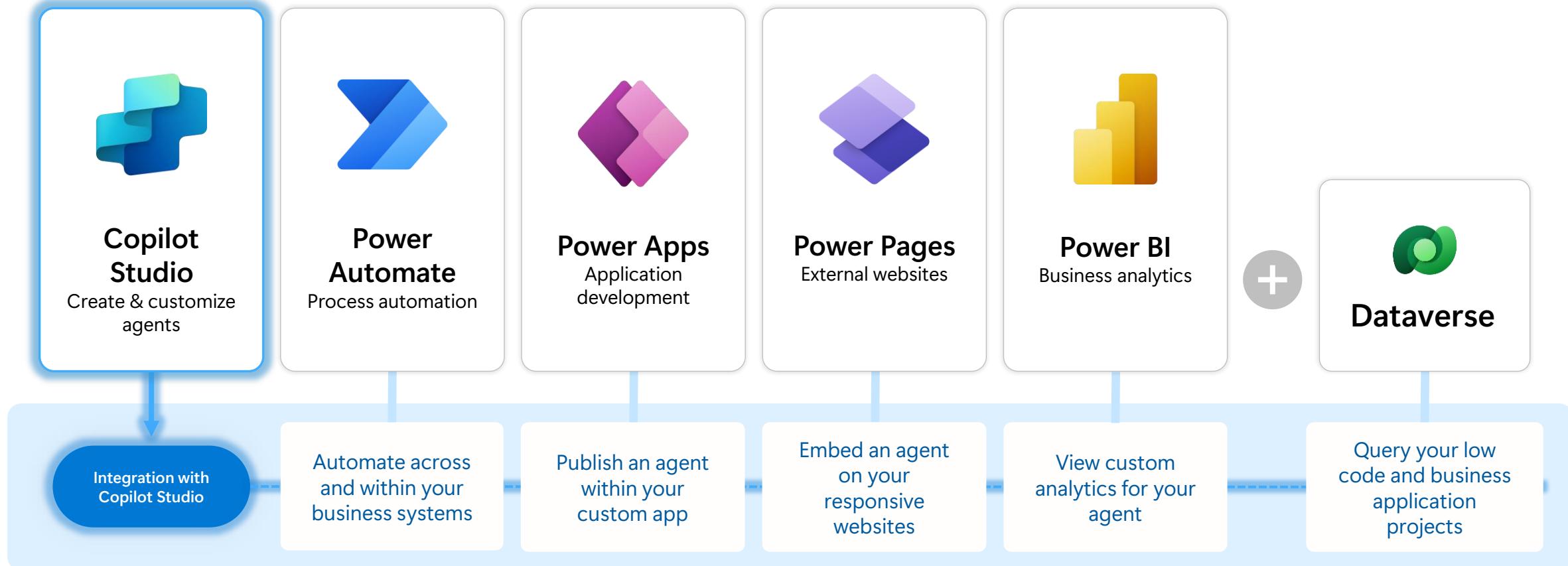
Provide an overview of all integrations with other systems/expected volumes

*Example answers:*

Connected system	Called by	Called with	Expected daily volume	Expected peak	Details / purpose
ServiceNow	Service Desk KB topic	Agent flow	5,000	1,000	The user query is used as an input to run a query on ServiceNow knowledge base using the standard connector. A JSON of results is returned to Copilot Studio for generative actions custom data.
Contoso website	Conversational Boosting topic	Generative answers	1,000	100	<a href="https://contoso.com/en-us/FAQ">https://contoso.com/en-us/FAQ</a> with /en-us/ being dynamically set based on the user locale variable.
Internal Directory API	Conversation Start topic	Agent flow	15,000	5,000	Use of the on-premises data gateway to connect to an internal API.
Weather API	Weather topic	HTTP	500	100	GET request to a REST-based API
Teams	Meeting Booking topic	Action	1,500	300	Create a Teams meeting based on conversation inputs.
Bot Framework Skill	Travel topic	Skills	2,000	400	Leveraging existing Bot Framework Skill to book travels.
SAP	Every 24h	Agent flow	20,000	-	Daily batch synchronization of the product catalogue from SAP to Dataverse.

# Microsoft Power Platform

Accelerate app development, automate workflows, visualize data, and reduce repetitive tasks with AI-powered tools

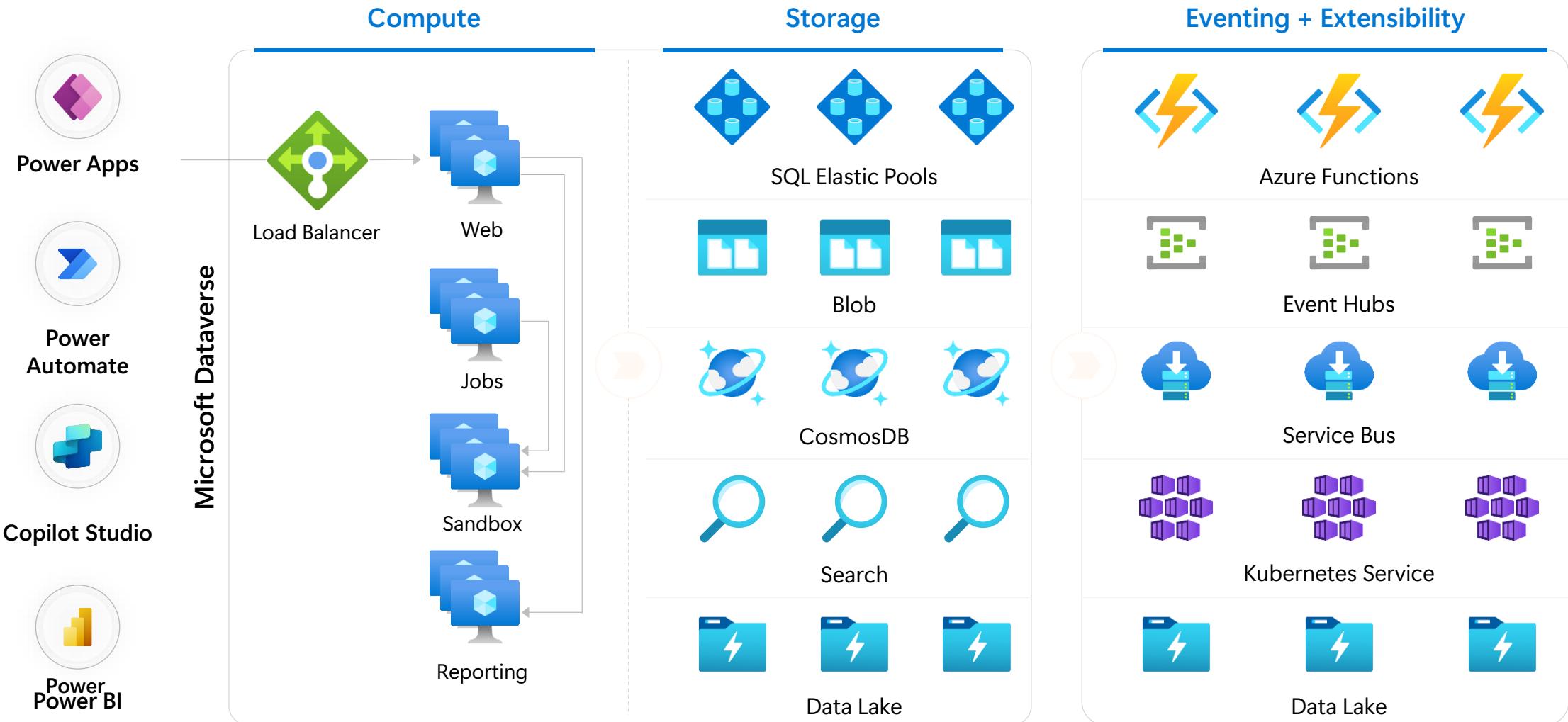


# Copilot Studio is hosted on Dataverse



Dataverse easily structures a variety of data and business logic to support interconnected applications and processes in a secure and compliant manner

# Dataverse runs on Azure and extends on Azure



# Telephony integration

## Copilot Studio & D365 Contact Center

- Build IVR dialog with Copilot Studio
- Natively integrated with Digital Contact Center (Requires D365 contact center)
- 1st party telephony via Azure Communication Service or bring your own phone number
- 1st party speech to text and text to speech via Azure Cognitive Service

## Copilot Studio & 3P contact center

- Build IVR dialog with Copilot Studio via event extensibility
- Integrates with 3rd party contact center such as Genesys via 3rd party solution AudioCodes
- AudioCodes provides telephony, speech-to-text and text-to-speech integration

# Use interactive voice response in your agents

Copilot Studio supports interactive voice response (IVR) capabilities, including speech and dual-tone multi-frequency (DTMF) input, context variables, call transfer, and speech and DTMF customization.

The screenshot shows the 'Settings' page in Copilot Studio. On the left, a sidebar lists various settings categories: Agent details, Generative AI, Security, Authoring Canvas, Entities, Skills, **Voice** (which is selected and highlighted in blue), Languages, Language understanding, Component collections, and Advanced. The main content area is titled 'Voice' and contains several configuration sections:

- Optimize for voice**: A note encouraging users to make their agent work well with voice-first features like advanced speech recognition and DTMF input. A 'Learn more' link and a toggle switch (set to on) are present.
- Use voice as primary authoring mode**: Another toggle switch (set to on).
- Customized automatic speech recognition**: A section for increasing accuracy with agent data, noting that it adds voice-specific data to the speech recognition model and increases publish time. A 'Learn more' link and a toggle switch (set to on) are included.
- Timeouts**: A section with four expandable items:
  - DTMF**: Allows choosing how the agent receives input from a dial pad.
  - Silence detection**: Allows choosing how the agent responds if it doesn't detect any input from the end user.
  - Speech collection**: Sets limits for how long the agent will try to detect and collect input.
  - Latency messaging**: Chooses how and when users hear a latency message when background operations are taking longer than expected. Notes mention latency messaging can be added in plugins or topics, and it only applies to telephony channels. An expand arrow is shown at the end of this item.

# Configure voice features

Collect user input via speech and DTMF



Control how your agent's voice sounds and behaves with end-users



Control the call flow by transferring calls or hanging up



Authoring capabilities when building a voice-enabled agent

The screenshot shows the 'Settings' interface with the 'Customized automatic speech recognition' section open. The 'DTMF' tab is selected, displaying options for 'Interdigit timeout' (set to 3000 ms) and 'Termination timeout' (set to 2000 ms, with the radio button for 'Termination timeout' selected). Below the DTMF settings, the 'Speech' section is partially visible, showing the 'Sensitivity' setting which controls the level of sensitivity to speech.

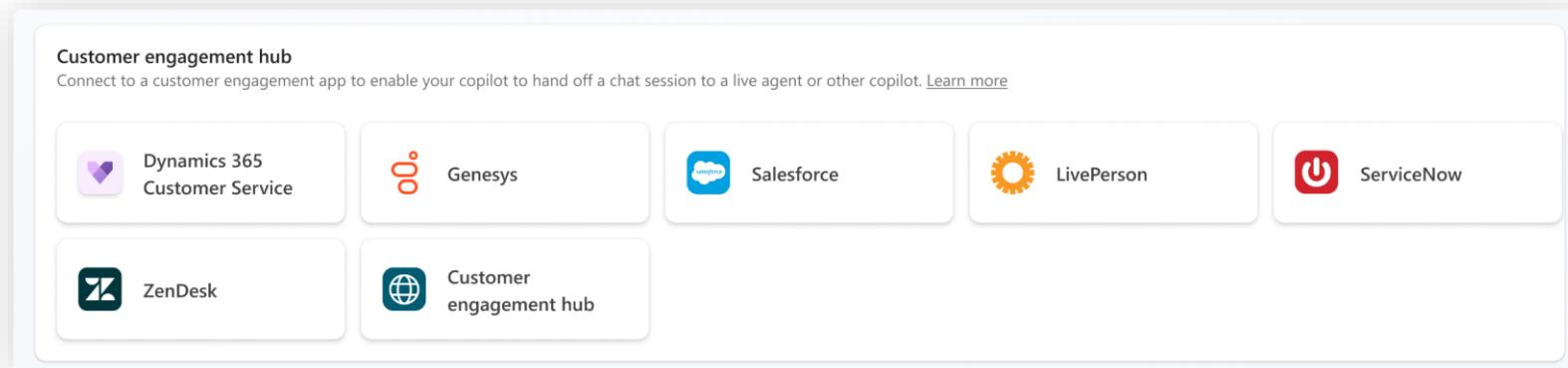
# Hand off to a live agent

When requiring hand off to an agent, the chat client must be one of the connected engagement hubs (that is placed at the front) and not Copilot Studio.

The engagement hub integrates with Copilot Studio through APIs, and that's how it's able to take over a conversation from an agent to escalate to a live agent.

A full hand off to an engagement hub follows this pattern:

- 1 An end user interacts with the engagement hub's chat canvas
- 2 The engagement hub routes the incoming chat through routing capabilities
- 3 A custom adapter relays the incoming chat messages to the Copilot Studio agent
- 4 Once the end user triggers hand-off, Copilot Studio starts hand-off with full chat context
- 5 The custom adapter intercepts the hand-off message and context, and seamlessly routes the conversation to a live agent
- 6 The end user's chat is handed off to a live agent who can resume the conversation

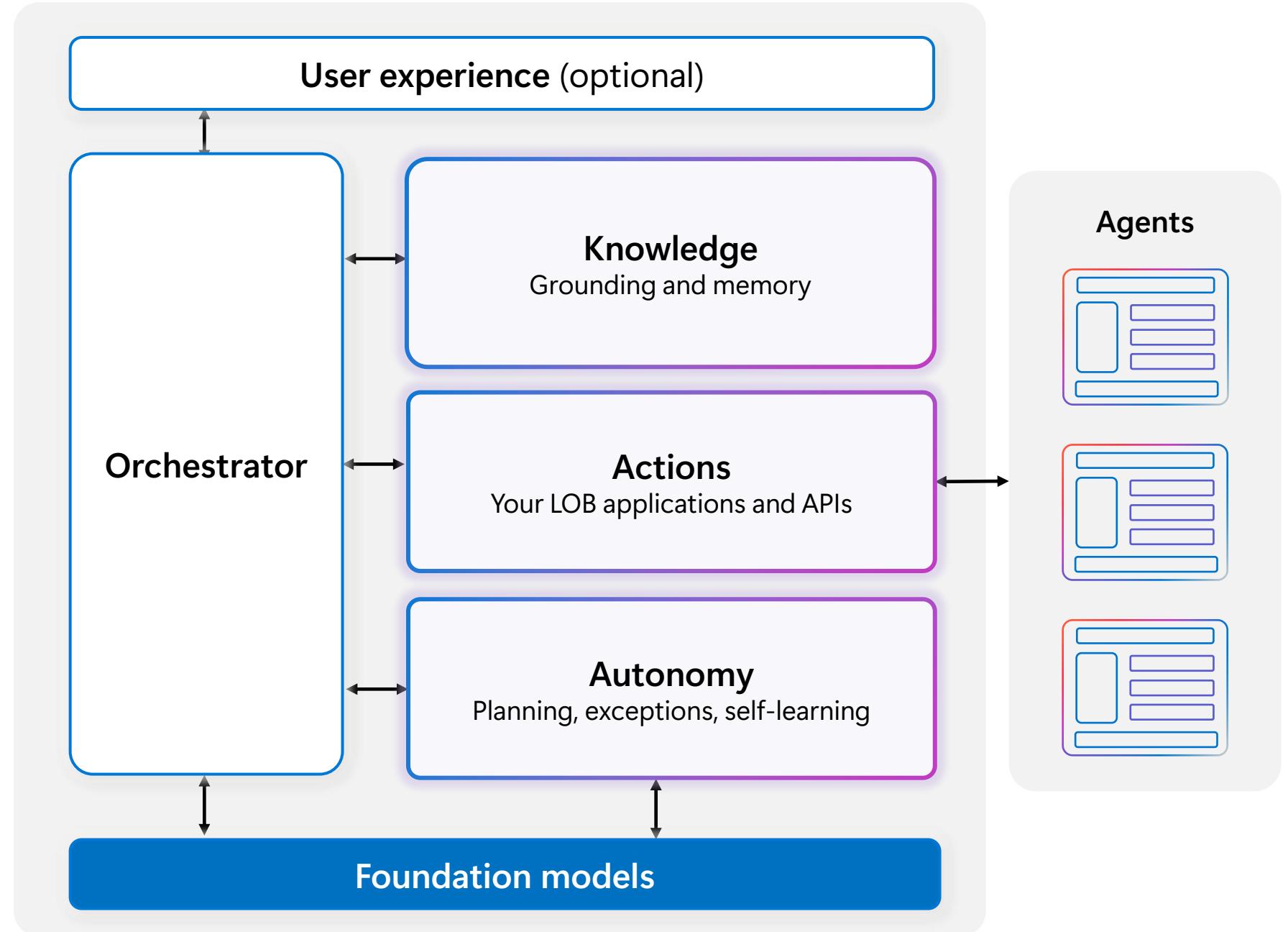


## Questions to discuss:

- Are you handing off conversations to Dynamics 365 Omnichannel for Customer Service? (*dedicated section*)
- Are you handing off conversations to another engagement hub solution? Like Genesys Cloud  
If yes, which one and how?
- What chat client are you placing in front of your end-users?

# Customize your agent

Easily tailor your agent's building blocks to meet your unique business needs in a comprehensive, end-to-end studio



# Better together

Copilot Studio +  
Visual Studio Code / GitHub,  
**all powered by Azure AI Foundry**



## Copilot Studio

Low-code managed environment for AI business process automation



## Visual Studio / GitHub

Pro-code environment for commercial AI solutions



## Azure AI Foundry

Comprehensive platform of capabilities, models, and tools to design, customize agents, and manage AI applications and agents



## Microsoft Fabric

Complete data platform to accelerate AI innovation

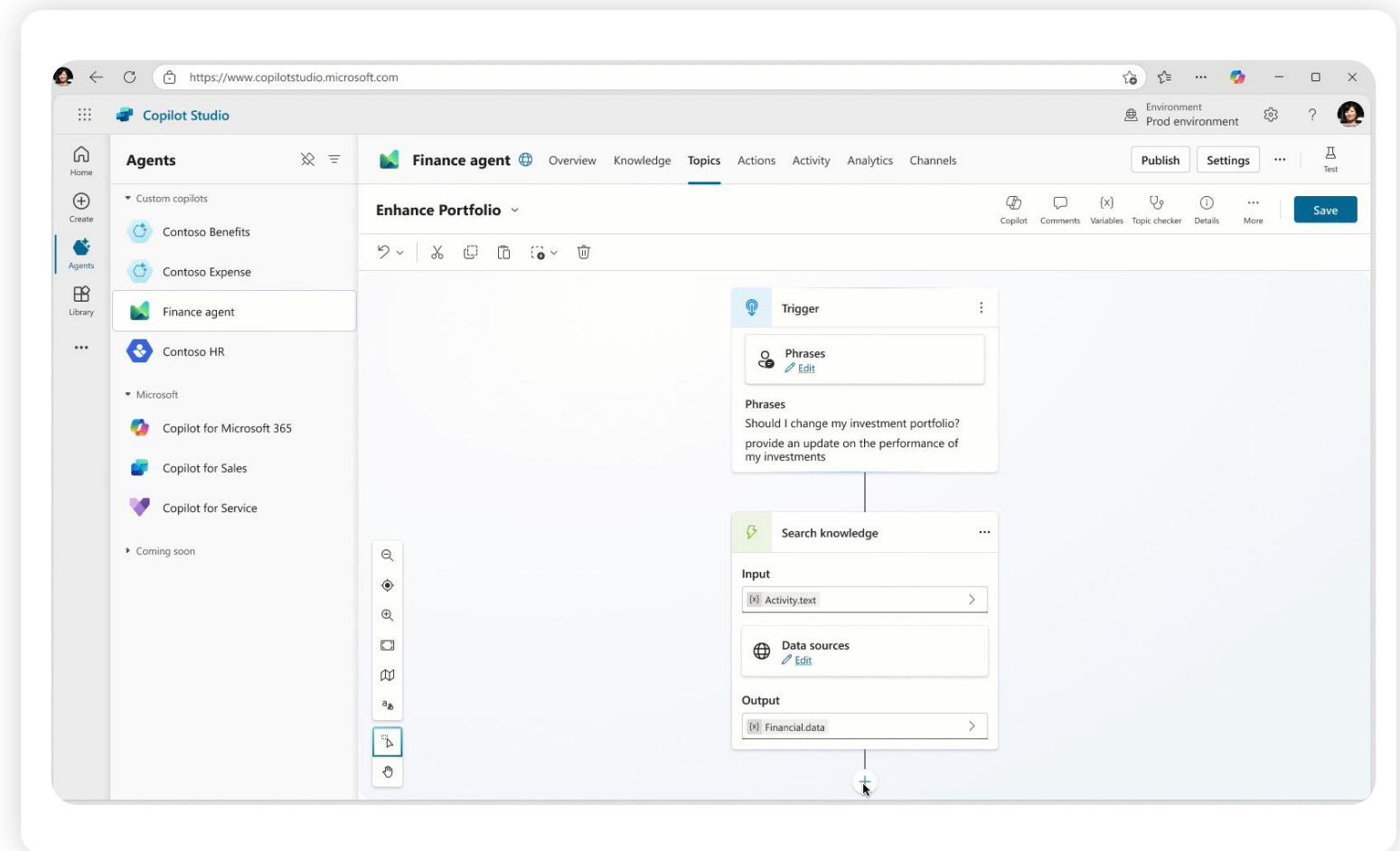
# Use models from Azure AI Foundry in Copilot Studio

Copilot Studio can call any of the 1,800+ Azure AI Foundry models

New search node on your knowledge

Search and call custom models from within your topic

Allows for more control and customization freedom



**Integrate with**

# Azure AI Foundry

## AI Search

Azure OpenAI's models enable you to build powerful conversational experiences over your data with Copilot Studio.

## Analytics

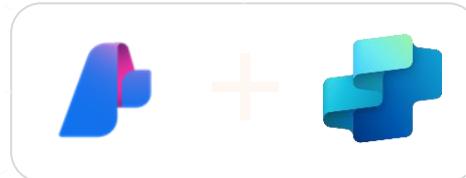
Build custom analytics by moving your data into extended data pipelines.

## API Calls

Copilot Studio supports any API / HTTP calls to Azure services.

## Language Services

Connect Conversational Language Understanding, including dialog triggering, interruptions, did-you-mean, and slot filling.



## Bot Framework

Access Bot Framework Composer capabilities natively in Copilot Studio, call existing Azure Bot Framework skills, and update and extend bots.

## Application Lifecycle Management

Works with Azure Devops for full ALM and solution management.

## Telemetry

Use Azure app insights for custom telemetry on copilot usage.

## Knowledge Base

Bring Azure Custom Question Answering into your agent.

# Developing agents using Microsoft 365 Agents SDK



# Developing agents using Microsoft 365 Agents SDK

Develop agents with Teams AI Library  
for Visual Studio Code

**Develop agents using Microsoft 365  
Agents SDK**

Develop your own agents with Azure  
AI Foundry

Develop multi agents with Semantic  
Kernel or AutoGen

Introducing

# Microsoft 365 Agents SDK



Your agent, your way for the  
most advanced needs



Build on your terms with  
Azure AI or your choice

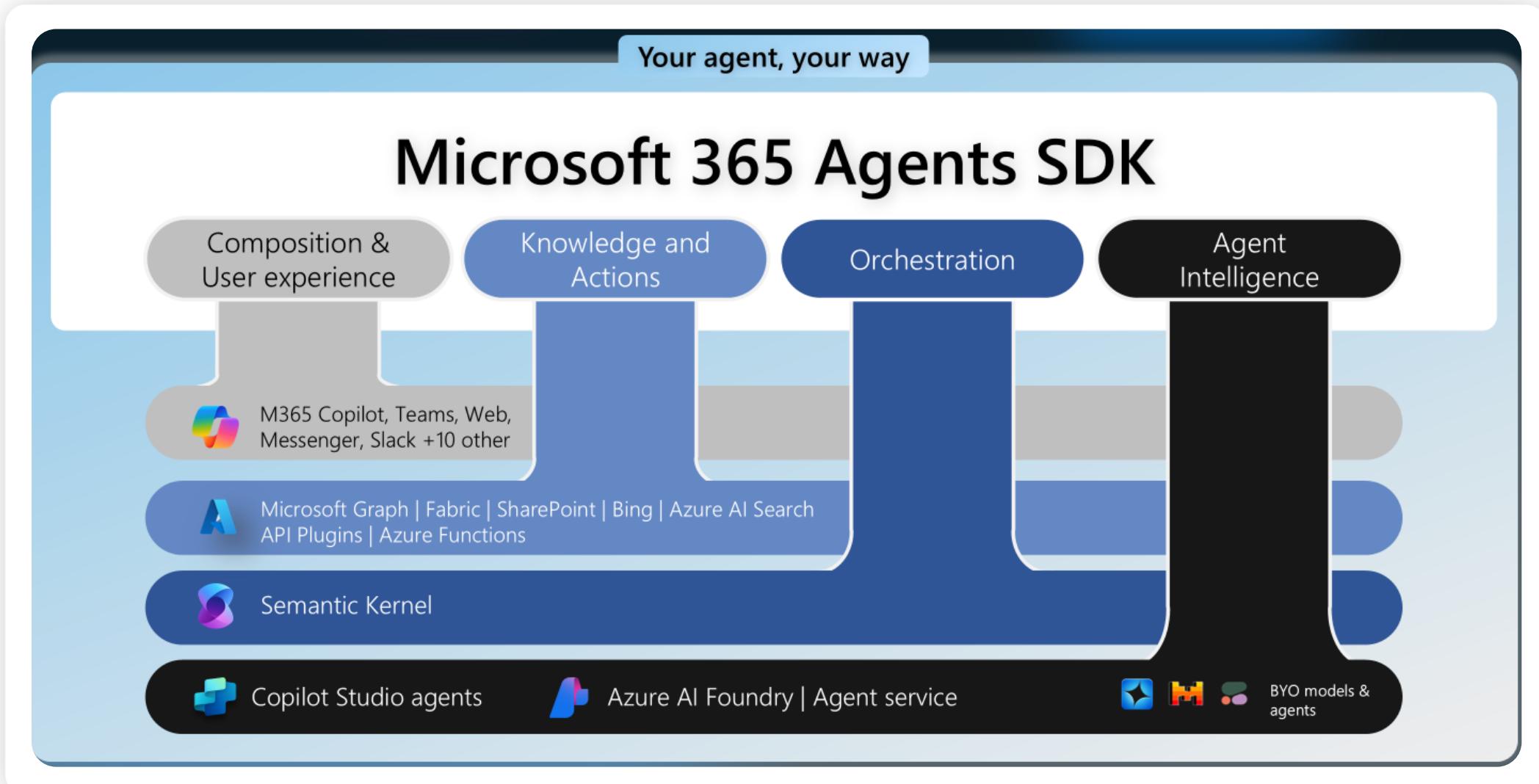


Seamlessly integrate into and  
from Copilot Studio



Publish across  
multiple channels

# Build on your terms with the AI services of your choice

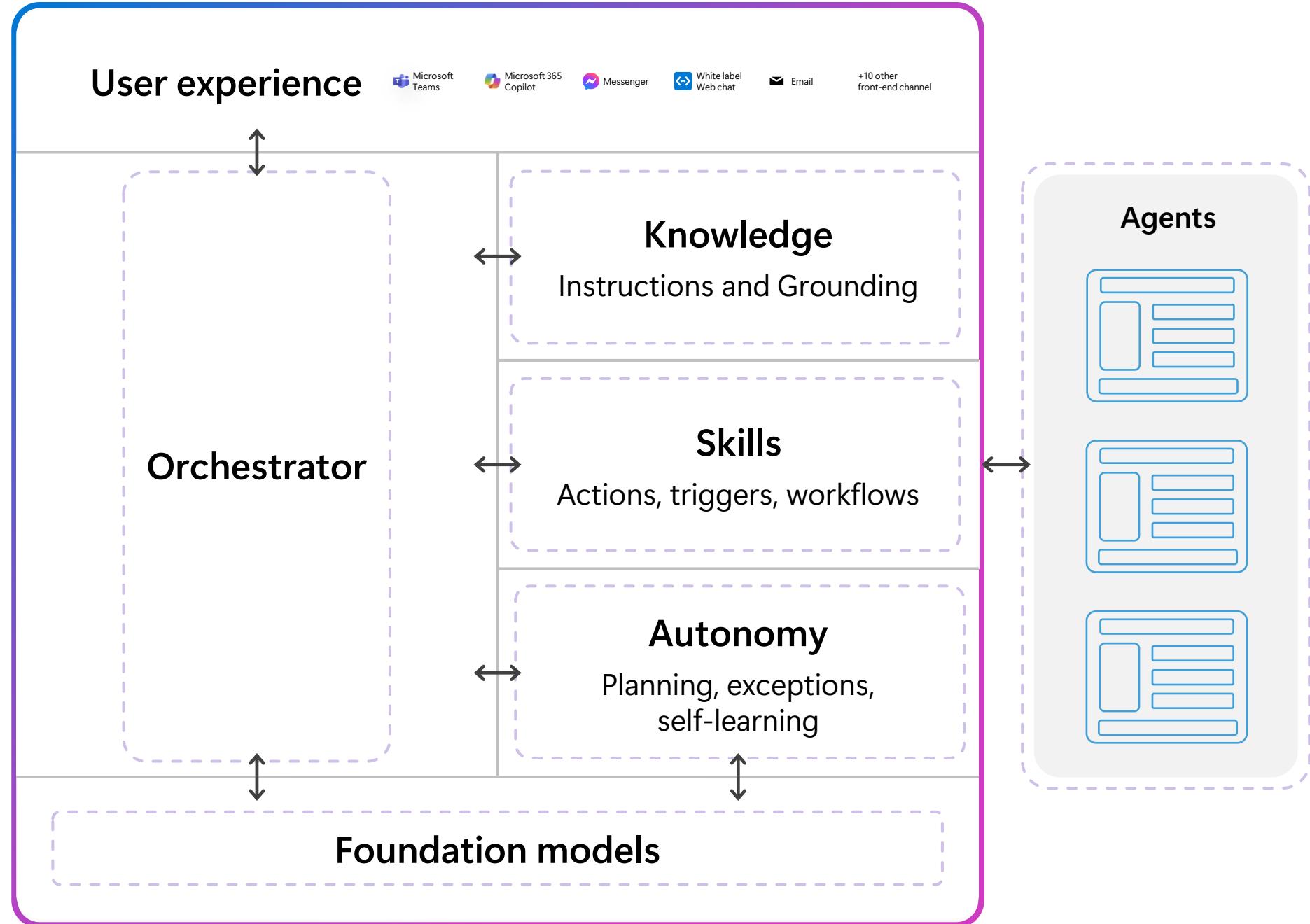


# Custom engine agents

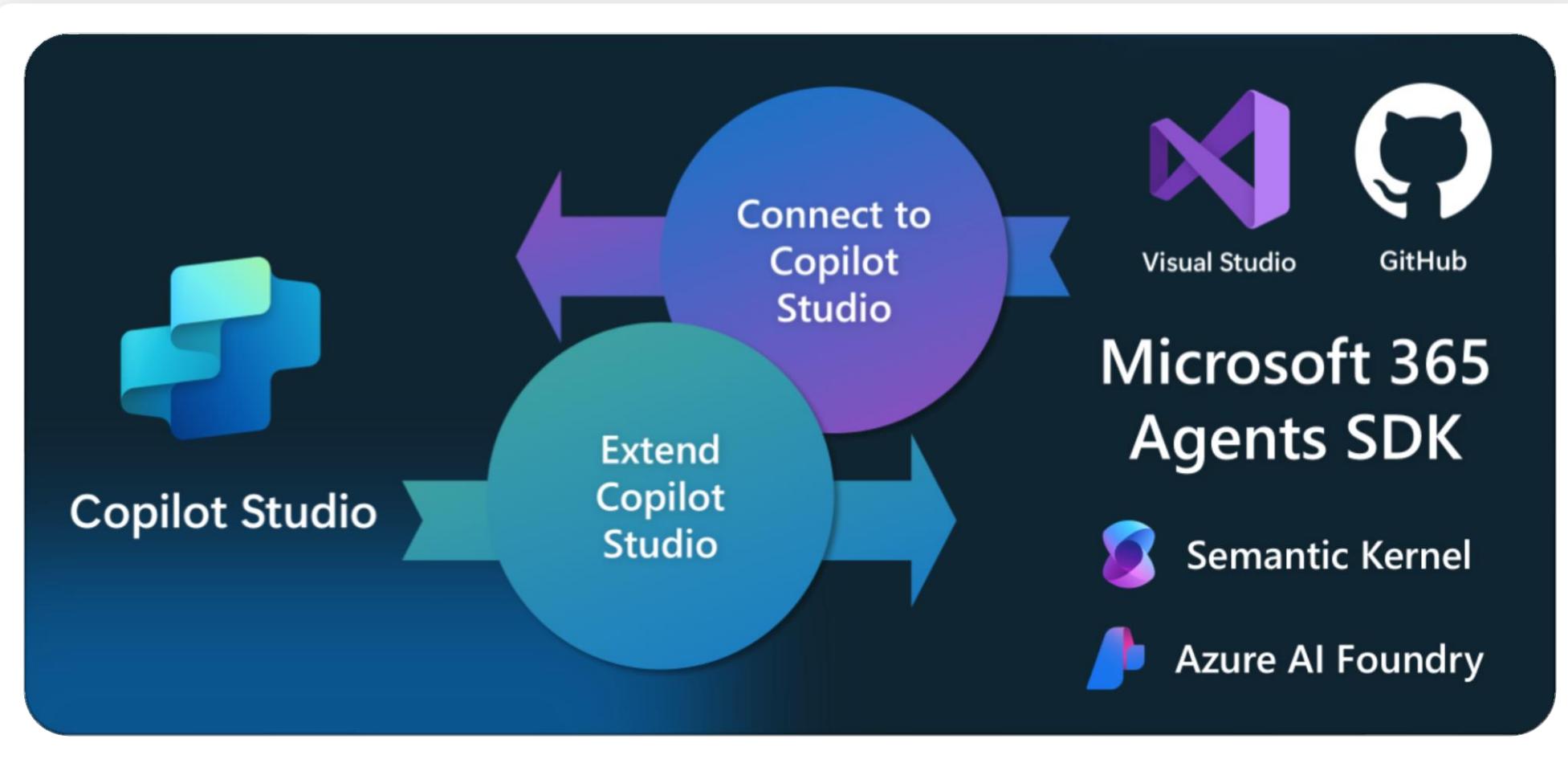
Microsoft 365 Agents  
SDK implementation

Microsoft provided

Developer provided



# Seamlessly build agents with low & pro-code



# Pro-code approaches to develop your own agents

Develop with [Microsoft 365 Agents Toolkit for Visual Studio Code](#)

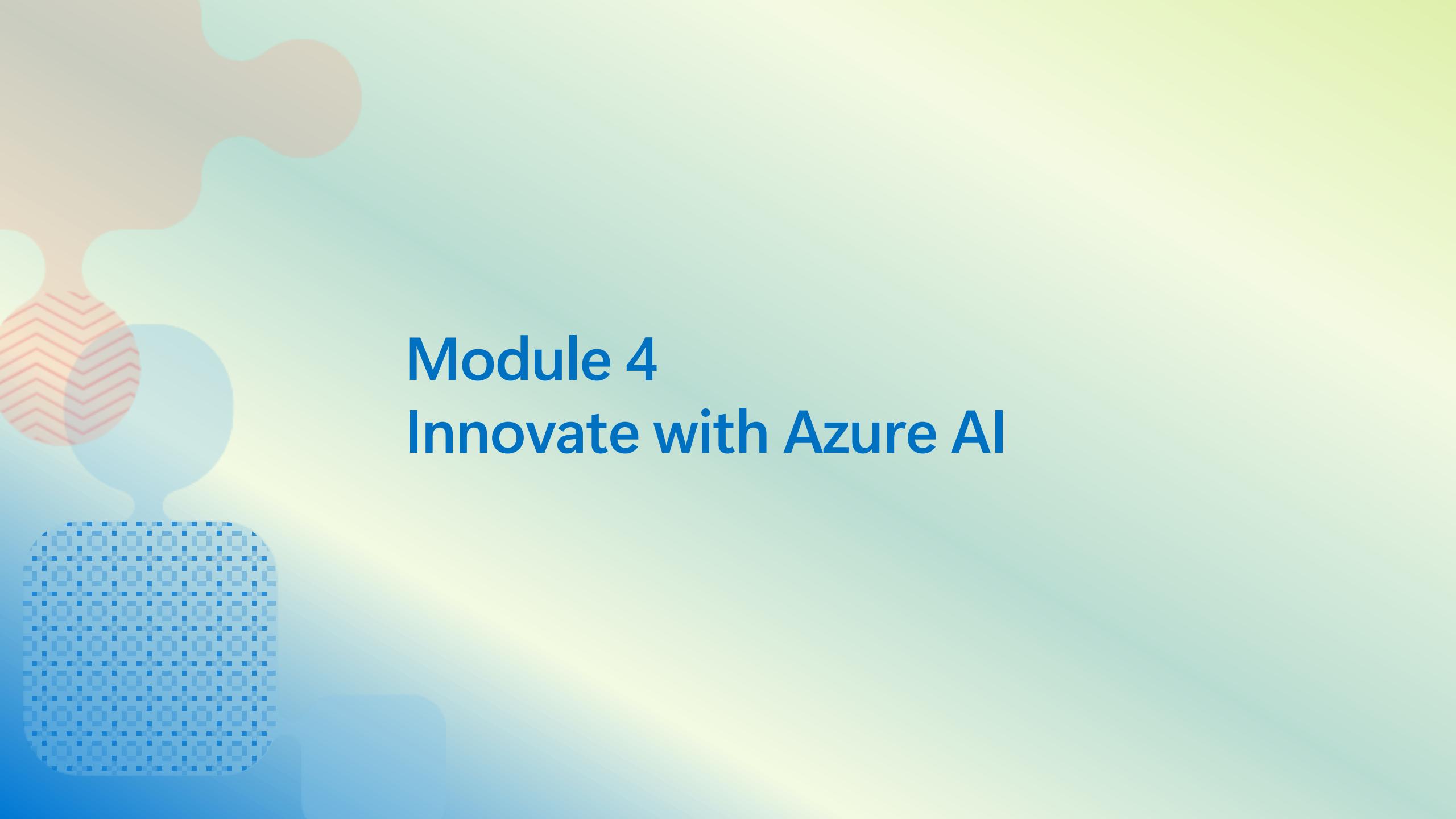
Develop using [Microsoft 365 Agents SDK](#)

Develop your own agents with [Azure AI Foundry](#)

Develop multi agents with [Semantic Kernel or AutoGen](#)

Coming up in Module 4

Coming up in Module 6



# **Module 4**

# **Innovate with Azure AI**

# Better together

Copilot Studio +  
Visual Studio Code / GitHub,  
**all powered by Azure AI Foundry**



## Copilot Studio

Low-code managed environment for AI business process automation



## Visual Studio / GitHub

Pro-code environment for commercial AI solutions



## Azure AI Foundry

Comprehensive platform of capabilities, models, and tools to design, customize agents, and manage AI applications and agents



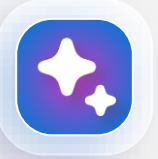
## Microsoft Fabric

Complete data platform to accelerate AI innovation

# Agents are already changing the way we...



Build and **manage** AI



Interact with **software**

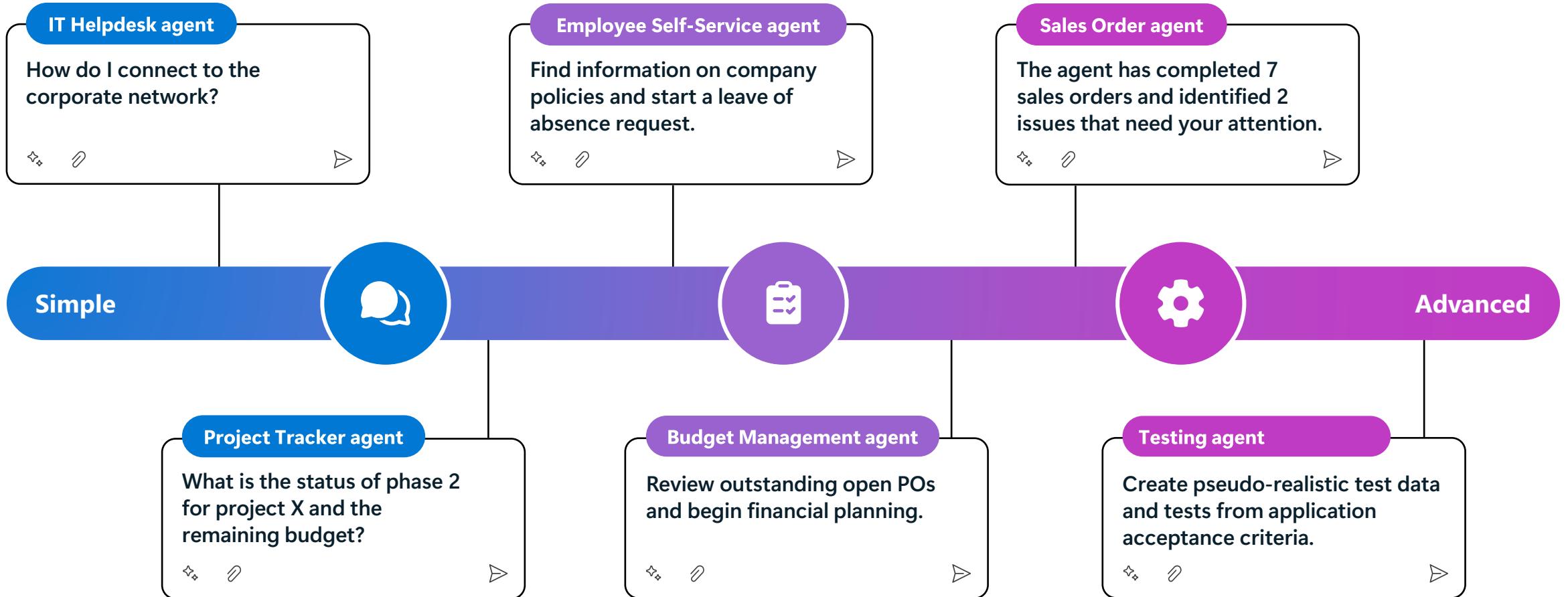


Gain value **from data**

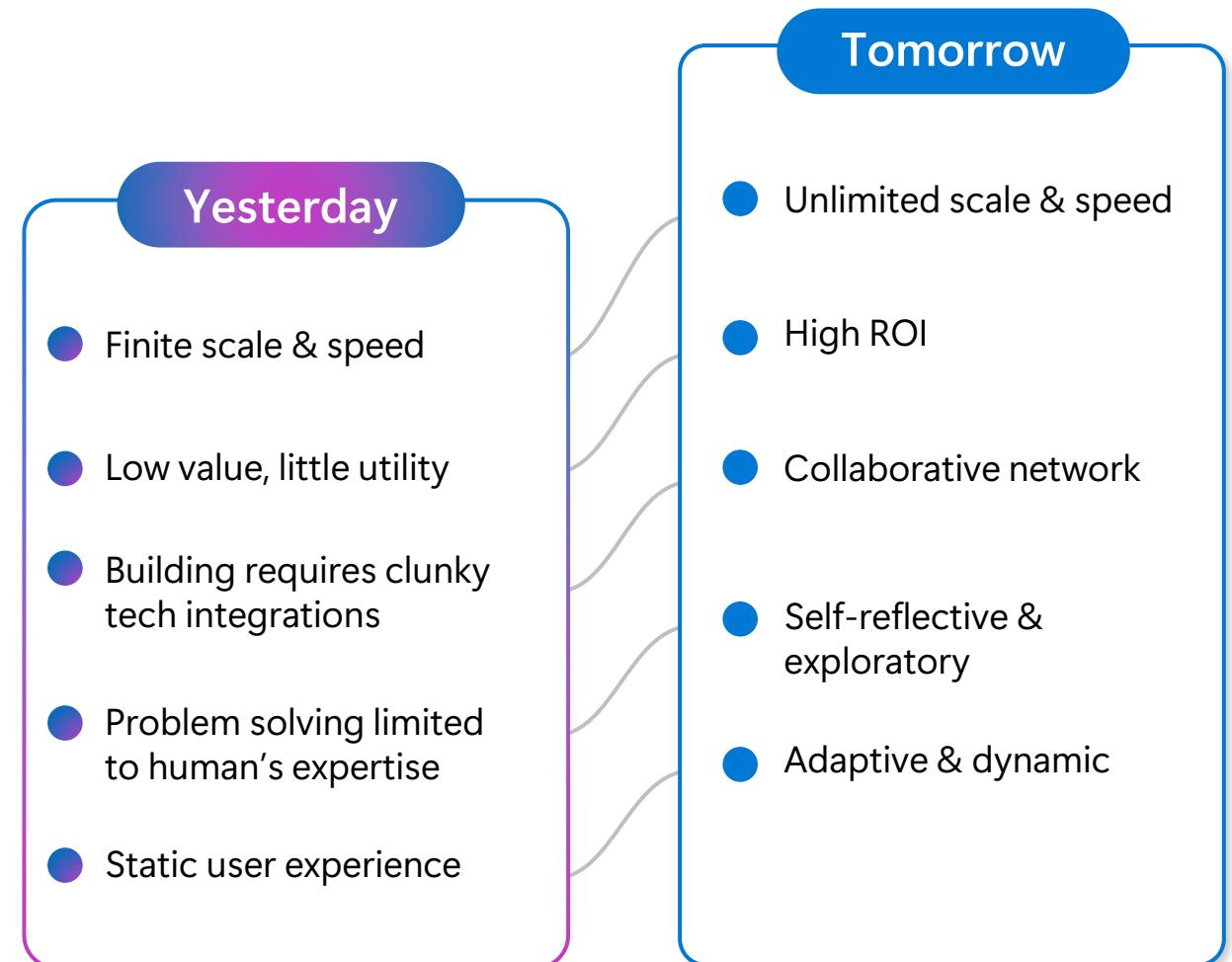


**Solve complex  
challenges**

# ...with customized solutions...



**...and they  
ask more  
of your applications**



**70%**  
or more generative AI  
experiments never  
make it to production<sup>2</sup>

## Agentic AI development challenges



Model selection &  
deployment



Complex workflows



Content safety

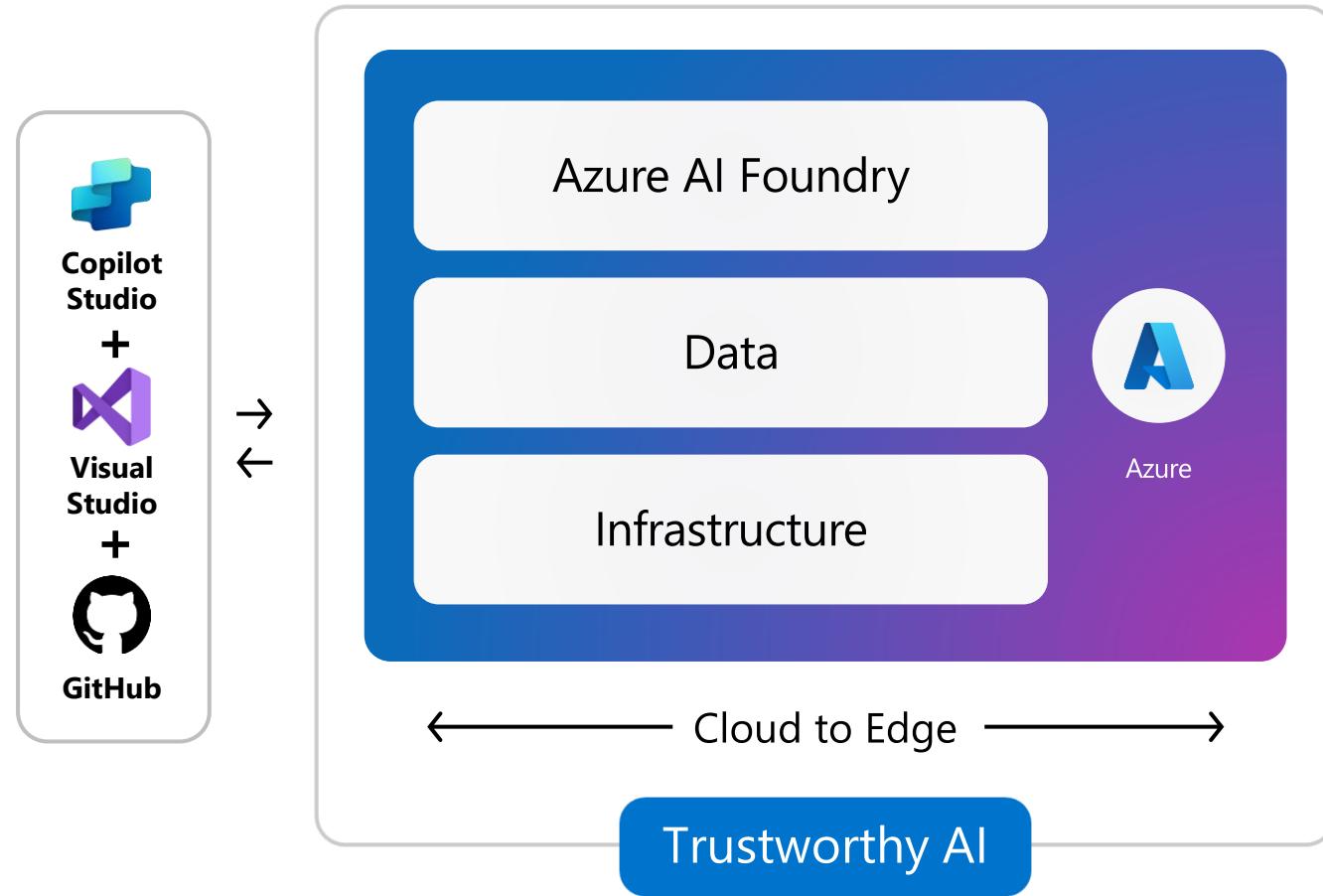


Observability and  
governance



Tool  
sprawl

# Copilot & AI stack

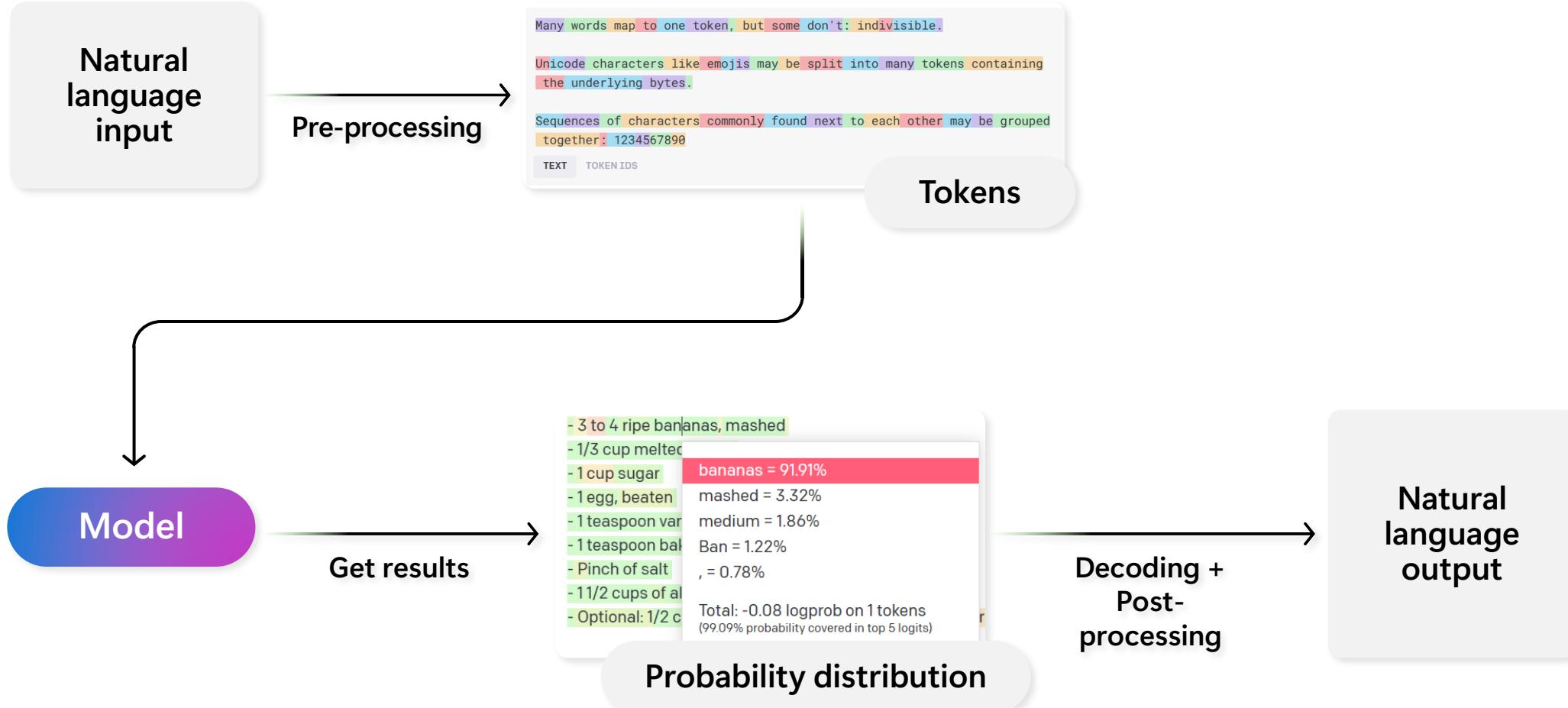


# How language models work

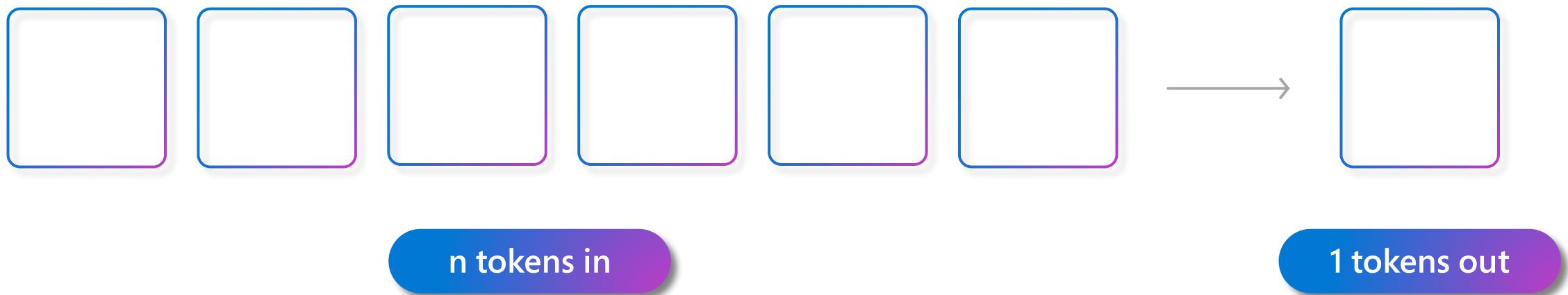
Predicting the next most likely token.



# How language models work



# How language models work



# How language models work

## Tokens

Tokens

11

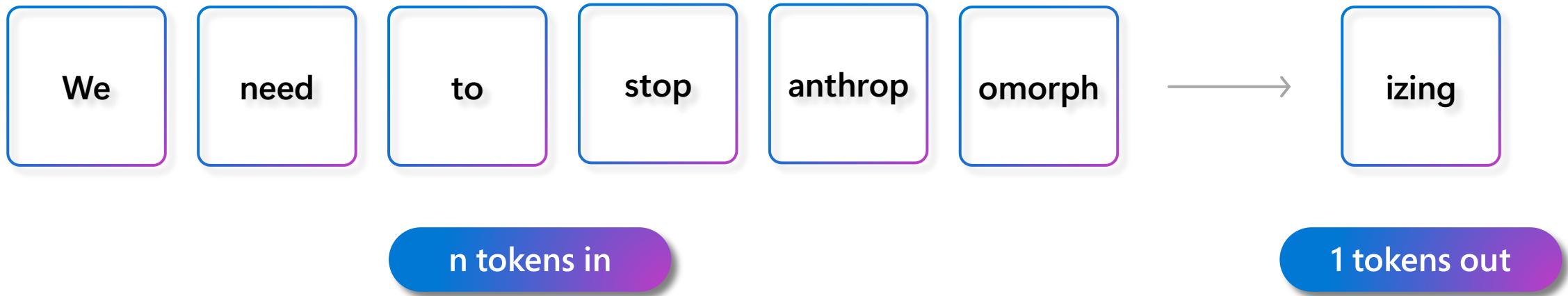
Characters

43

We need to stop anthropomorphizing ChatGPT.

[platform.openai.com/tokenizer](https://platform.openai.com/tokenizer)

# How language models generate text



# How language models generate text

in      out

We need to stop

We need to stop anthrop

We need to stop anthropomorph

We need to stop anthropomorphizing

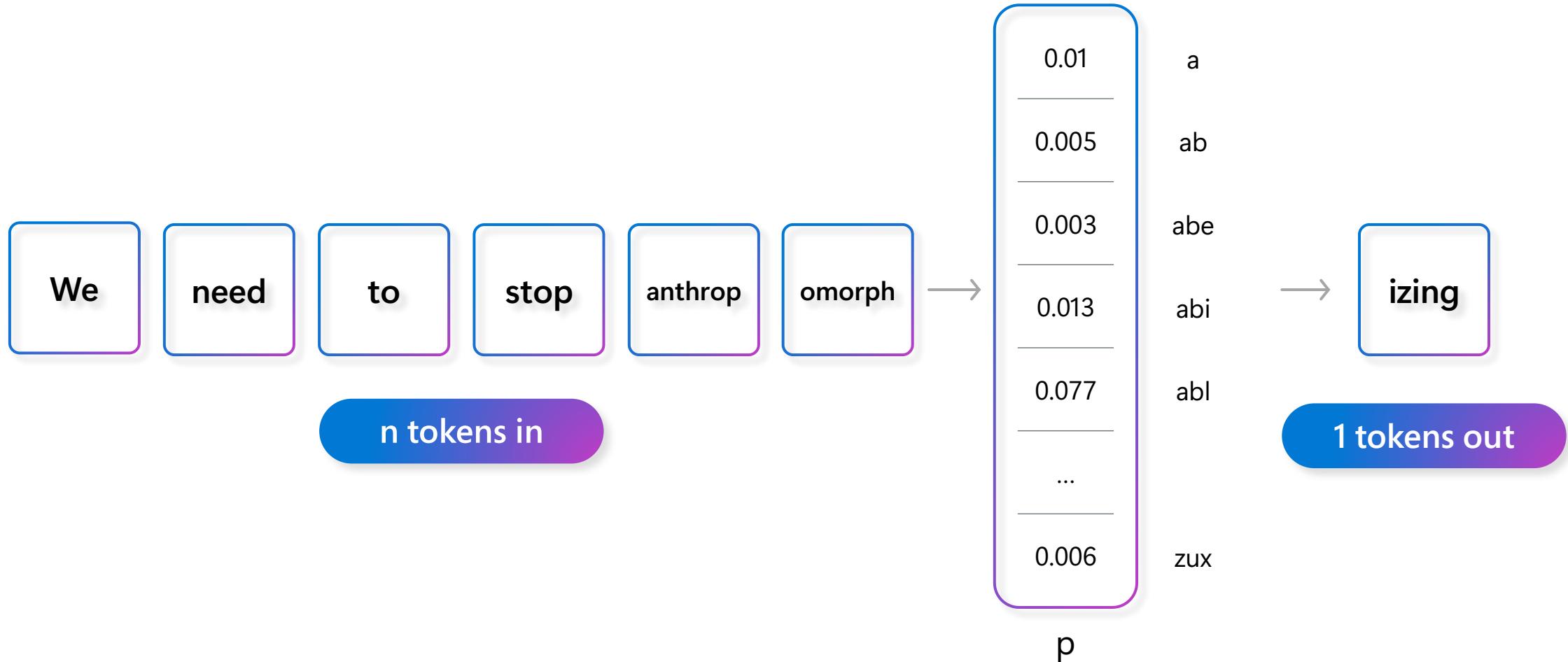
We need to stop anthropomorphizing Chat

We need to stop anthropomorphizing ChatG

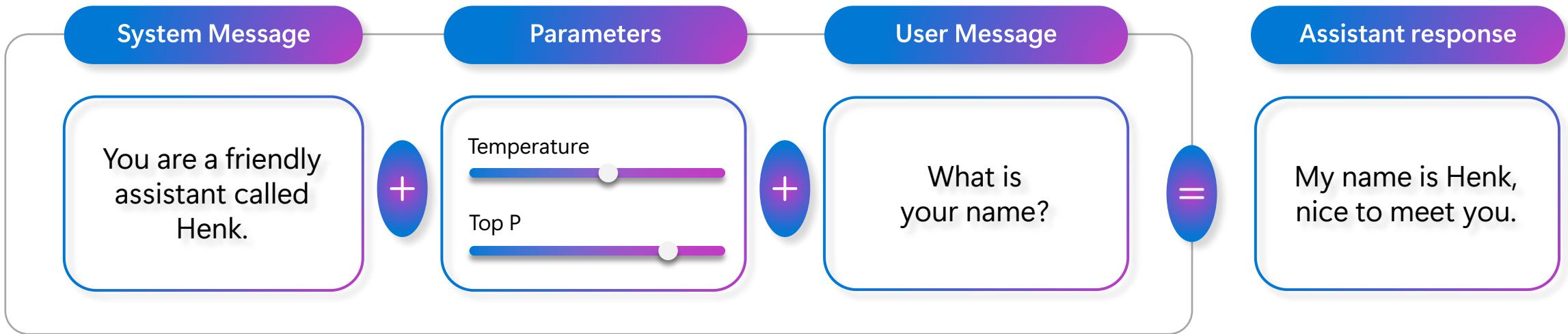
We need to stop anthropomorphizing ChatGPT

We need to stop anthropomorphizing ChatGPT.

# How language models generate text



# Interacting with LLMs





# Azure AI Foundry

-  Copilot Studio
-  Visual Studio
-  GitHub
-  Azure AI Foundry SDK

Foundational models

Open-source models

Task models

Industry models

Azure  
OpenAI Service

Azure AI  
Search

Azure AI Foundry  
Agent Service

Azure AI  
Content  
Safety

Azure  
Machine Learning

Evaluations

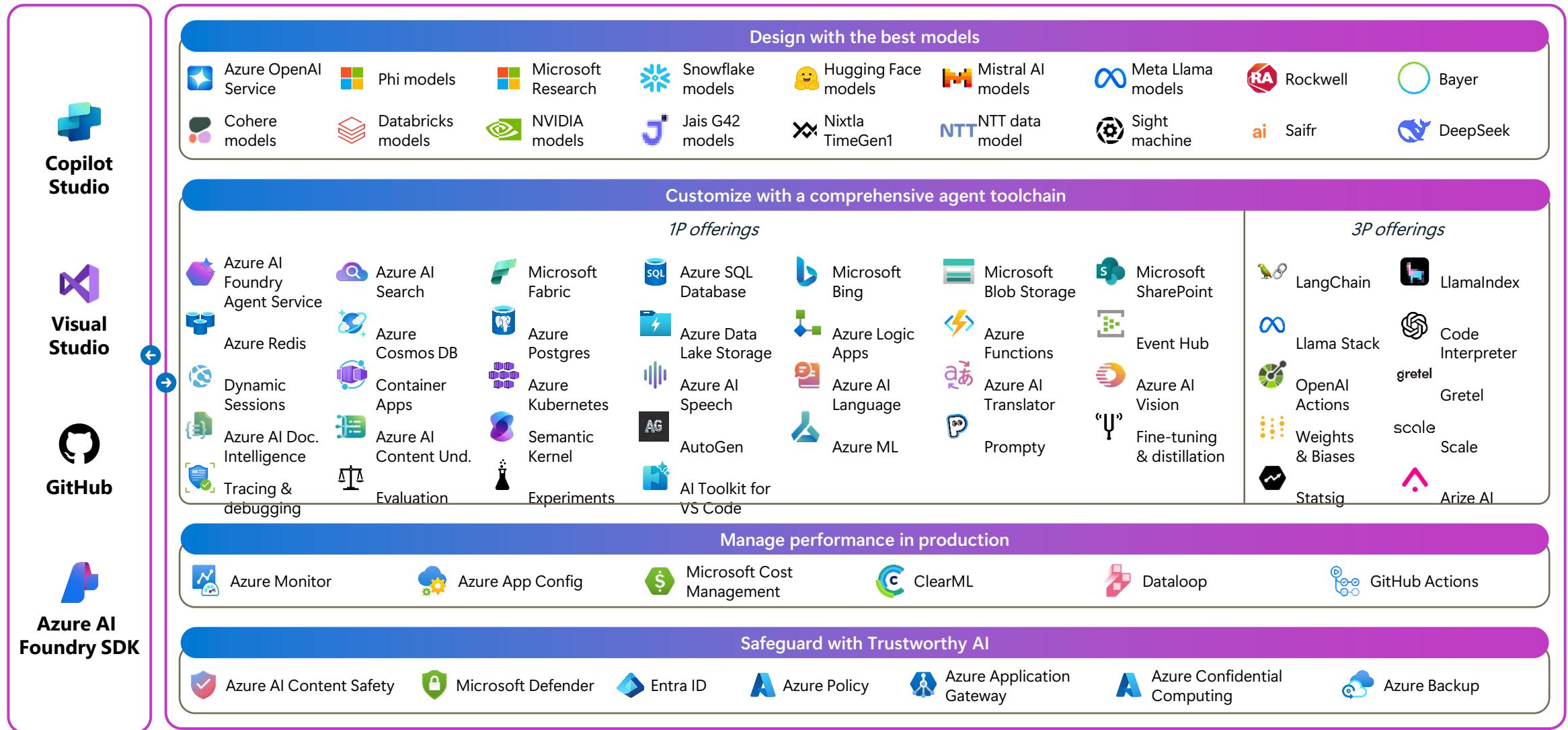
Customization

Governance

Monitoring

## Observability

# Azure AI Foundry ecosystem



# Azure AI Foundry SDK

Unified toolchain for AI application development

Access our leading models  
through a single interface

Easily integrate Azure AI capabilities into  
your applications

Develop faster with a simplified  
coding experience

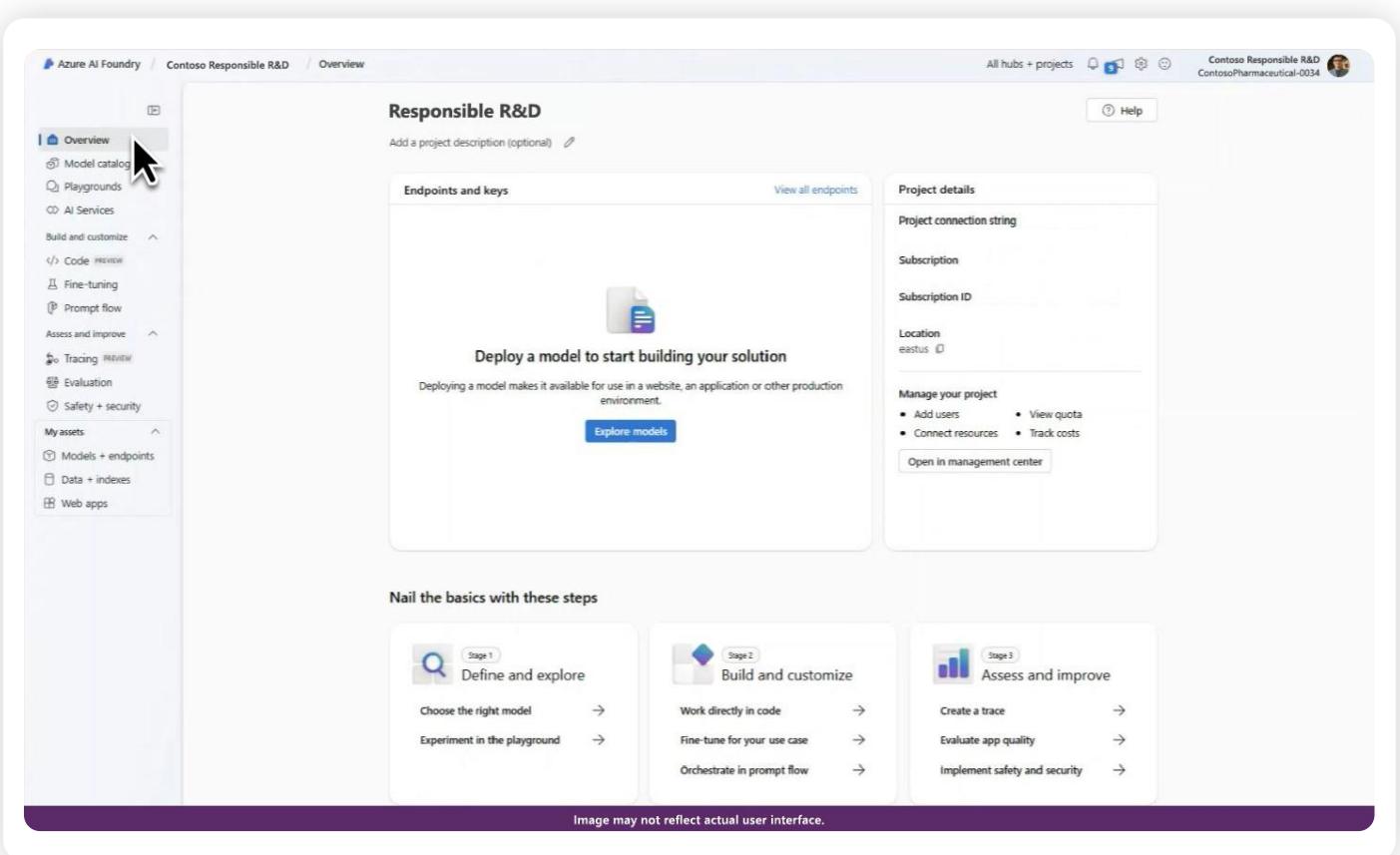
[AI.Azure.com](https://AI.Azure.com)



# Azure AI Foundry SDK

The Azure AI Foundry SDK is a comprehensive toolkit for developers, offering pre-built modules and resources to integrate AI functionalities seamlessly into applications.

- Access our most popular models through a single interface
- Easily integrate Azure AI Foundry capabilities into your application with a single project client
- Unlock another level of intelligence with Azure AI Foundry Agent Service
- Supports languages like Python and C#
- Evaluate your apps locally, in the cloud, and in production using state-of-the art safety and quality evaluators
- Review project logs with integrated tracing
- Access Azure AI App Templates beyond SDKs, to provision resources for popular agentic scenarios, hosted in web, container, function app, and more



 The Azure AI Foundry SDK provides a local developer experience that reduces the complexity of using multiple resources together in code when building AI apps and agents.

# Top use cases for

Azure AI Foundry



Build your own agent



Chat with your data



Analyze and summarize documents



Generate code and documentation



Deliver personalized recommendations



Improve information discovery



Generate engaging content



Process transactions and detect fraud



Improve forecasting

# Which model to use?

Easily find various types of models



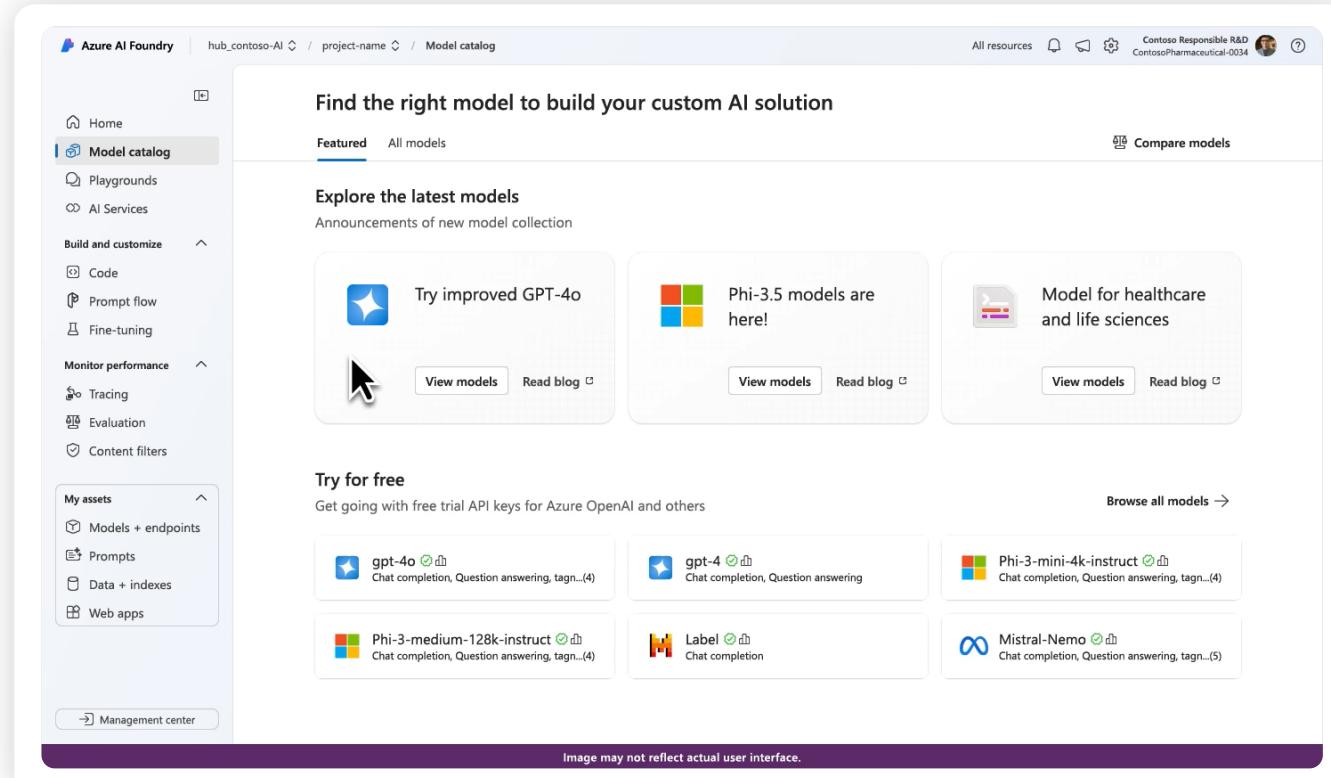


# Design with the best models

Discover models for your use case

Design intelligent apps with industry-leading task models

Deploy serverless models



The screenshot shows the Azure AI Foundry Model catalog interface. The left sidebar includes links for Home, Model catalog (which is selected), Playgrounds, AI Services, Build and customize (with options for Code, Prompt flow, Fine-tuning), Monitor performance (with options for Tracing, Evaluation, Content filters), and My assets (with options for Models + endpoints, Prompts, Data + indexes, Web apps). The main content area has a header "Find the right model to build your custom AI solution" and tabs for "Featured" and "All models". It features sections for "Explore the latest models" (announcements of new model collections) and "Try for free" (free trial API keys for Azure OpenAI and others). There are cards for "Try improved GPT-4o", "Phi-3.5 models are here!", and "Model for healthcare and life sciences", each with "View models" and "Read blog" buttons. Below these are cards for "gpt-4o", "gpt-4", "Phi-3-mini-4k-instruct", "Phi-3-medium-128k-instruct", "Label", and "Mistral-Nemo", each with their respective icons and brief descriptions. A "Management center" button is at the bottom of the sidebar, and a note "Image may not reflect actual user interface." is at the bottom of the main content area.



# Model catalog

The model catalog enables discovery of large and small language models packaged for out-of-the-box or custom usage in Azure AI Foundry.

- Models from OpenAI, Hugging Face, and Meta, Azure OpenAI.
- Filter model catalog by collection, industry, tasks, or model name to find the model that best suits your needs.
- Compare models by Benchmarks, now integrated into Model Catalog: benchmarks across models and public datasets to find the model that best meets your business scenario.
- Redesigned model card that highlights key information about the model.
- Leverage ready-to-use, curated fine-tuning pipelines, code-based inferencing, and evaluation of the model

Easily find the foundation model that best suits your needs while improving business outcomes, reducing operational expenses, and enhancing competitive advantage.



# Azure AI Foundry Models

Discover models for your use case

*Offering 11,000+ frontier and open models*

## Azure OpenAI

o1/o1-mini  
o3/o3-mini  
o4-mini  
GPT-4o/GPT-4o mini  
GPT-Image-1  
Sora  
GPT-4.1/ 4.1-mini/4.1-nano  
Dall\*E-3  
Whisper



## Microsoft

Phi-4  
Phi-4-mini/multimodal  
Phi-4-mini-reasoning  
Phi-4-reasoning/ reasoning-plus  
Bitnet  
Magma  
Aurora  
BioEmu  
OmniParser v2



## DeepSeek

DeepSeek-R1  
DeepSeek-V3 0324  
MAI-DS-R1 (Microsoft AI)



## xAI

Grok 3  
Grok 3 mini



## Black Forest Labs

Flux Pro 1.1



## Meta

Meta-Llama-3  
Meta-Llama-3.1-405B  
Llama-3.2  
Llama-3.3  
Llama4-Scout  
Llama4-Maverick  
CodeLlama  
Llama-Guard



## Mistral AI

Mistral Large  
Mistral Medium 3  
Mistral Nemo  
Mistral Small  
Mistral 7B  
Mixtral 8x7B – Mixture of Experts  
Codestral  
Mistral OCR



## Cohere

Command R+  
Command R  
Command A  
Cohere Rerank  
Embed v3-Multilingual  
Embed v3-English  
Embed 4



## Hugging Face

Qwen QwQ 32b  
DeepCogito  
Seethal Sentiment  
Roberta  
Google Flan  
Intel Bert  
Cross Encoder  
Intfloat Multilingual E5  
Cerebras



## NVIDIA

Nemotron-3-8B-4k  
Nemotron-3-8B-Chat-SFT/ RLHF/SteerLM  
Nemotron-3-8B-QA  
NIMS microservices inc:  
Mistral 7B Instruct v0.3  
DeepSeek R1 Distill Llama  
Mixtral 8x7B Instruct  
Llama 3.2 NV Rerank QA



## Stability

Stable Diffusion  
Stable Image Ultra

## Bria

Bria 2.3 Fast

## Deci

DeciCoder  
DeciDiffusion

## Databricks

Databricks/dbrx-base  
Databricks/dbrx-instruct



## Snowflake

Snowflake/arctic-base  
Snowflake/arctic-instruct



## Nixtla

Time GEN-1

## Industry Models

Fidelity/Saifr  
Sight Machine  
Bayer  
Cerence  
Paige AI  
Prism  
Rockwell  
Virchow  
RA-FT-Optix

## Core 42

JAIS

## Gretel

Navigator

## SDAIA

ALLaM-2-7b-instruct

## NTT

Data Tsuzumi

# Find the best model for every application

11000 frontier, task and open models  
in Azure AI Foundry model catalog



OpenAI  
Model Family  
(available day 1)



Phi SLM  
Model Family



Mistral AI  
Model Family



Meta Llama 2  
Model Family



Jais G42  
Model Family



Snowflake  
Model Family



Cohere  
Model Family



Databricks  
Model Family



Hugging Face  
Collection



NVIDIA  
Model Family



Deci AI  
Model Family



Nixtla  
Time-Gen1 Model



NTT Data  
Model



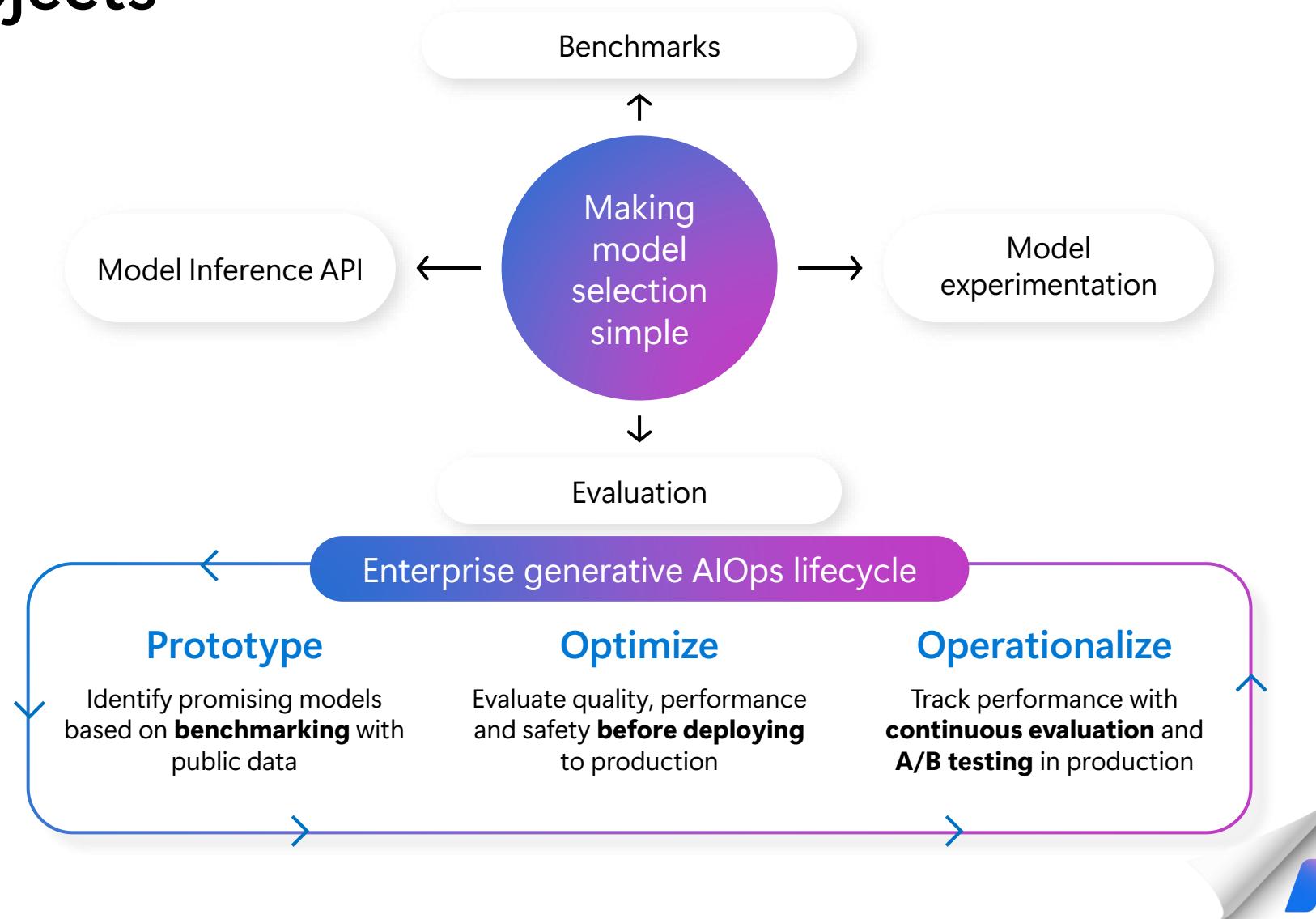
Industry models:



# Select your best models with support through every stage of your AI projects

## The Azure AI Foundry difference:

- Unified API for model swaps
- Built-in, full-lifecycle tooling
- Versatile deployment options



# Azure AI Services

Build cutting-edge, market-ready AI applications with out-of-the-box and customizable APIs and multi-modal models

The screenshot shows the Azure AI Foundry interface for the project "contoso\_a1". The left sidebar includes sections for Overview, Model catalog, Playgrounds, AI Services (selected), Build and customize, Fine-tuning, Prompt flow, Assess and improve, Tracing, Evaluation, Safety + security, My assets (Models + endpoints, Data + indexes, Web apps), and Management center.

The main content area is titled "Azure AI Services" and describes creating intelligent apps with small, task-specific models. It features a "Speech playground" section with a video player showing a translated video of a plane flying over mountains. A "Pronunciation score" of 92 is shown with a breakdown for Accuracy and Fluency. Other sections include "Content Understanding", "Document field extraction", and "Infuse your solutions with AI capabilities" (Speech, Language + Translator, Vision + Document, Content Safety). The "What's new" section highlights "Document translation", "Ensure content safety for generative AI", and "Extract PII". The "Learning resources" section provides links to Documentation, Watch a video, Get started with AI on Azure, and Microsoft Q&A.

# Design intelligent apps with Azure AI services

## Leverage out-of-the-box and customizable APIs and multimodal models

### Azure OpenAI Service

- Access to powerful AI models
- Scalable development
- Compliance & security
- Integration with other Azure Services

### Azure AI Search

- AI enrichment & semantic ranking
- Generative AI content creation
- Vector search for data organization

### Azure AI Speech

- Speech to text (including the Whisper model on Azure OpenAI Service)
- Text to speech
- Speech translation
- Speaker recognition

### Azure AI Vision

- Image and face analysis
- Custom model training
- Face detection and recognition
- Document text extraction

### Azure AI Content Safety

- AI-driven content moderation for enhanced safety
- Customize safety thresholds for diverse user types
- Detect and prevent Jailbreak Risk from XPIA attacks

### Azure AI Document Intelligence

- Automated documentation generation
- Documentation quality analysis
- Interactive documentation experiences
- Natural language understanding for documentation

### Azure AI Language

- Task-optimized AI models for text analytics
- Custom industry-specific AI for healthcare
- Custom, industry-specific models

### Azure AI Translator

- Multilingual text and speech translation
- Synchronous and asynchronous translation request support
- Native translation of documents and manuals

# Azure OpenAI Service (AOAI)

Discover models for your use case

Azure OpenAI Service empowers developers and businesses to harness the power of advanced AI technologies without the need for extensive AI expertise or infrastructure investment.

- Latest innovations with same-day access - Offering the latest models in the Azure OpenAI Service on the same-day with the same features.
- Widest Range of Deployment offerings - Standard, Provisioned and Batch available across 28 different regions, data zones and Global
- Your data is private and secure. Data residency at rest and at processing, and data protection from exfiltration.
- Safety is built-in with tools to responsibly build and implement AI.
- Enterprise Azure promises with private networking, managed identity, customer managed keys, azure monitoring, cost management, and more.
- Developer-Friendly Integration. Seamless integration with GitHub, Visual Studio and Copilot Studio.
- Choice of Models - Over 1800 models available to integrate into your application.

The screenshot shows the Azure AI Foundry Model catalog interface. At the top, there's a navigation bar with icons for settings, user profile, and help. Below it, a search bar and a 'Find the right model to build your custom AI solution' heading. The main area displays a grid of 25 model cards, each with a thumbnail, name, and brief description. The models listed include o1-preview, o1-mini, gpt-4o-realtime-preview, gpt-4o, gpt-4o-mini, gpt-4, gpt-4-32k, text-embedding-3-large, text-embedding-3-small, tts, tts-hd, whisper, dall-e-3, dall-e-2, text-embedding-ada-002, davinci-002, gpt-35-turbo-16k, gpt-35-turbo-instruct, gpt-35-turbo, babbage-002, openai-whisper-large, openai-clip-vit-large-patch14, openai-clip-vit-base-patch32, OpenAI-CLIP-Image-Text-E..., and OpenAI-CLIP-Image-Text-E... . A 'Models 25' label is in the top right corner of the grid. At the bottom, there are navigation arrows for 'Prev' and 'Next'.



Leverage OpenAI's state-of-the-art models and Azure's cloud platform to build innovative AI-driven applications and solutions across various industries and domains.



Cutting-edge AI models  
Pioneering research and advancements  
Unlocking new creative possibilities



Secure data and built-in responsible AI  
Global availability and flexibility  
Developer-friendly integrations

## Azure OpenAI Service

Connect your data | Customize your agents | Manage quota & deploy

**Reasoning and  
multimodal  
text, image, audio**

o-series models  
GPT 4.1 series  
gpt-4o  
gpt-4o mini

**Fine-tuning**

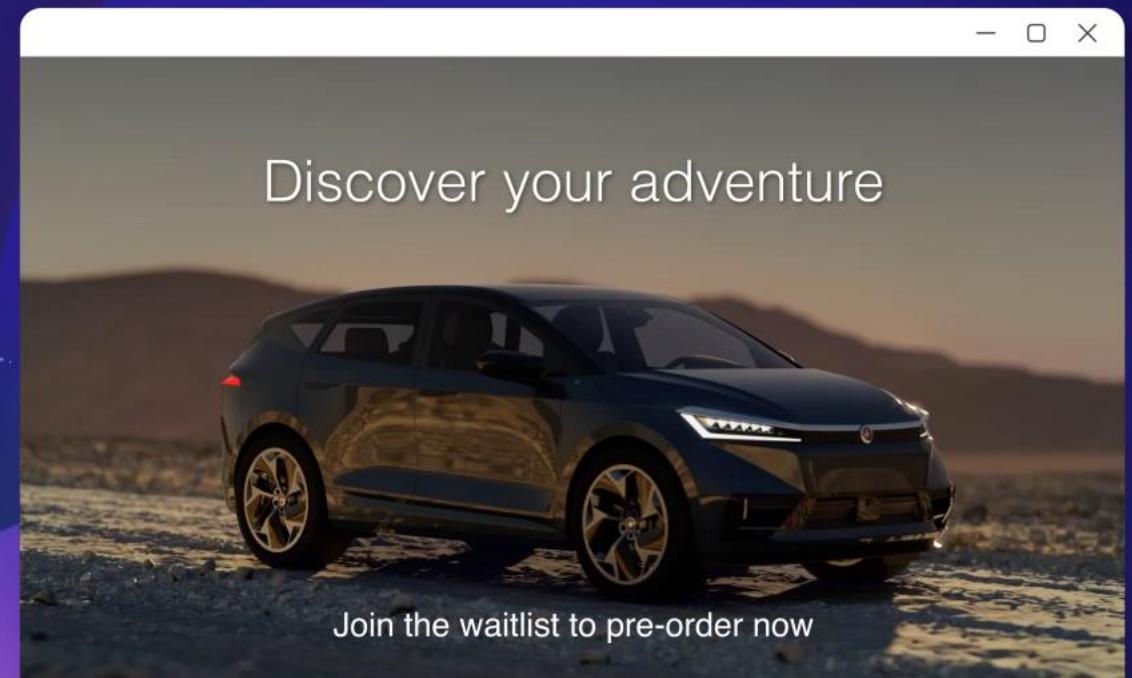
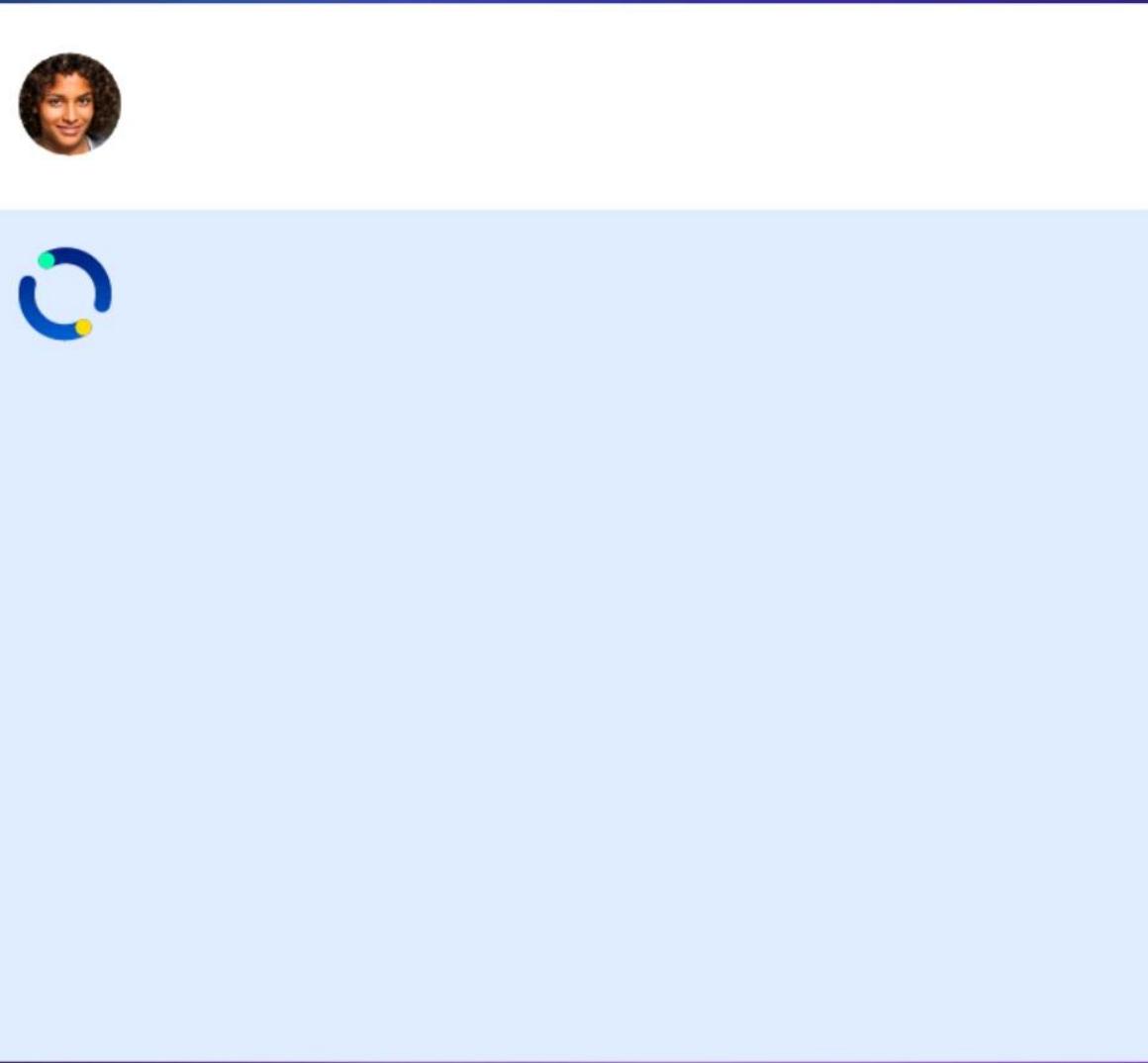
gpt-4o  
gpt-4o mini  
gpt-4.1  
gpt-4.1-nano

**Images**

gpt-image-1

**Transcription &  
translation**

Whisper  
gpt-4o-transcribe



# Azure AI Speech

## Speech-to-text

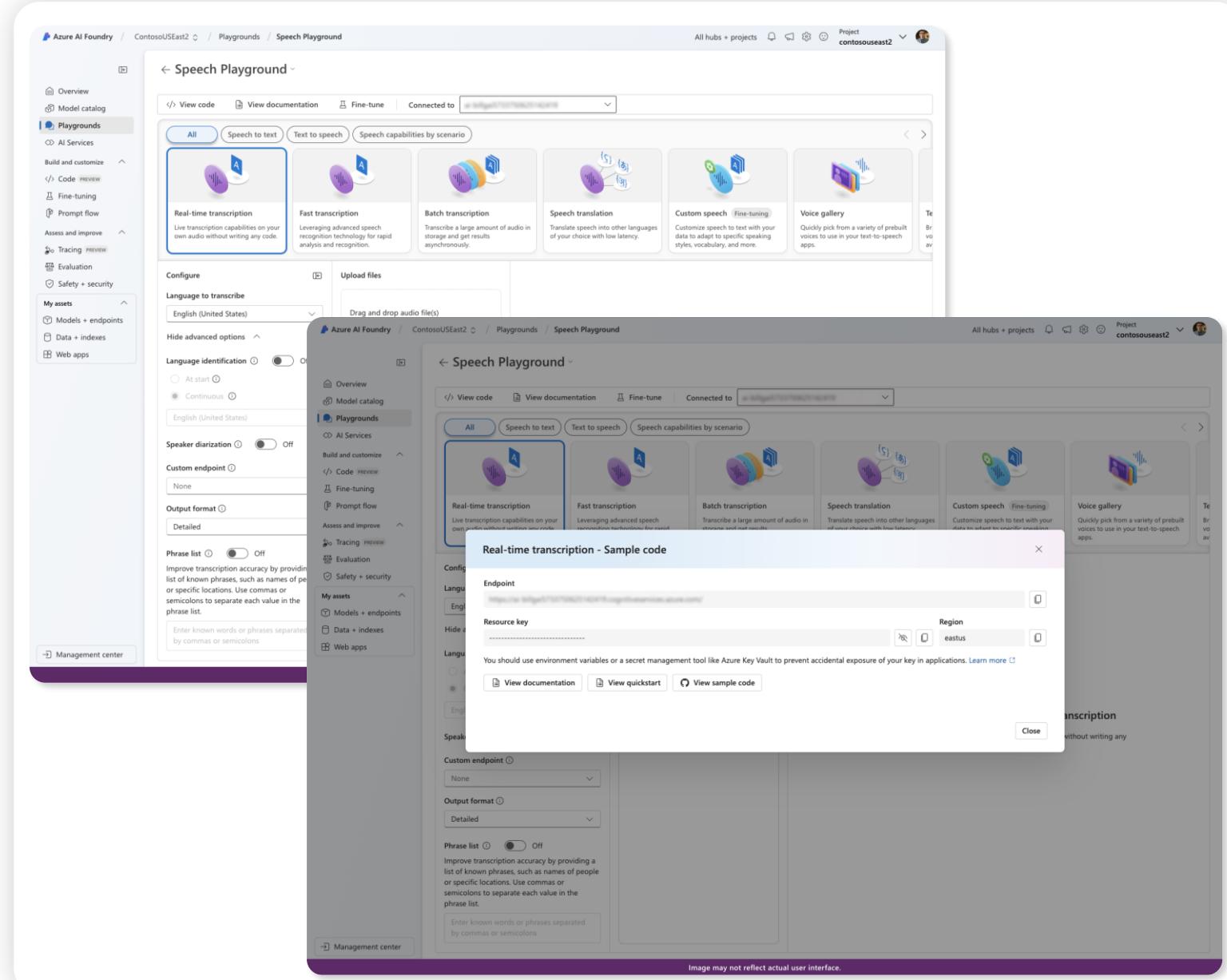
- Contact center – post call analytics and agent assist
- Meeting transcription & insights
- Audio/video transcription & insights
- Dictation
- Caption & subtitles for online and in-person meetings
- Audio and video captions & subtitles

## Text-to-speech

- Context-aware, highly expressive HD voices
- Professional voice/brand voice
- Avatars

## Speech translation

- Interpreter agent in Microsoft Teams
- Video translation API
- Real-time captions and subtitles
- Audio and video caption & subtitle
- Batch video

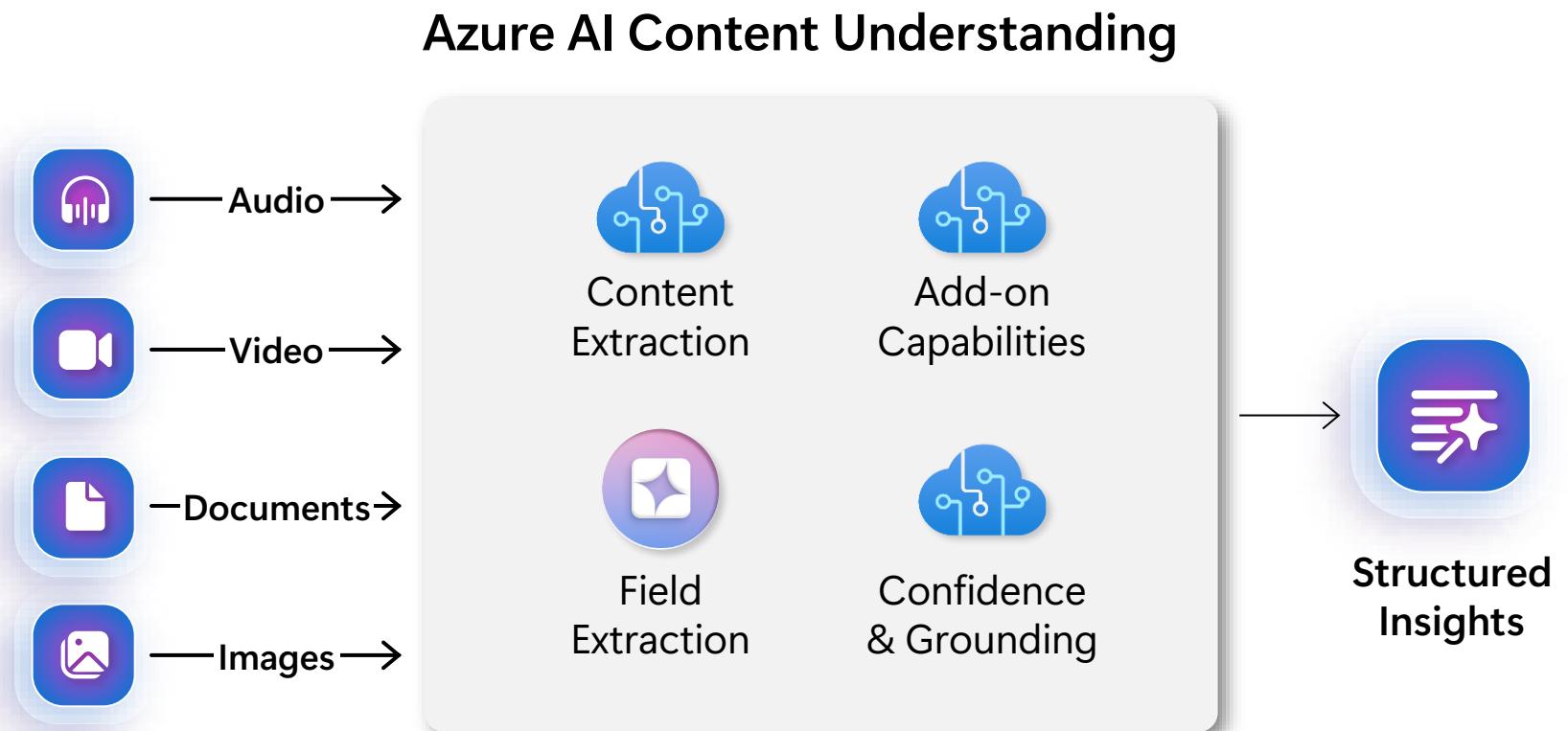


Available in [ai.azure.com](https://ai.azure.com)

# Generate insights from documents, text, audio, images, and video

## Differentiators:

- Streamline multi-modal app development
- Task specific structured output for agentic workflows
- Pre-built templates to speed development



# Models as a Service (MaaS)

## Offering within the model catalog

Models as a Service (MaaS) lets you fine-tune models using serverless APIs, making it easier for generative AI developers to build custom copilots.

- Ready-to-use APIs with pay-as-you-go billing based on tokens
- Customize models with your own data without the need to set up and manage GPU infrastructure
- Easily integrate the latest AI models as an API endpoint to applications
- Integrate with preferred orchestration tools like prompt flow, Semantic Kernel, or LangChain
- Achieve serverless fine-tuning without provisioning GPUs
- Fine-tune Llama 2 with your own data to enhance the model's ability to generate more precise predictions

The screenshot shows the Azure AI Foundry Model catalog interface. At the top, there is a purple button labeled "Deploy serverless models". Below the header, there is a section titled "Find the right model to build your custom AI solution" which includes a "What's new?" section with cards for "tsuzumi is now available!", "Announcing BRIA 3.2 Fast", "News from Cohere!", "New SLM from Mistral", and "Meta Llama 3.2 models are here!". Below this, there is a section titled "New model benchmarks available now in model catalog" with a sub-section "Compare with benchmarks" and "How model benchmarks are scored". The main area displays a grid of 1815 models, each with a thumbnail, name, and description. The models listed include o1-preview, o1-mini, gpt-4-realtime-preview, gpt-40, gpt-40-mini, tsuzumi-7b, Bria-2.3-Fast, Minstral-3B, Cohere-embed-v3-english, Cohere-embed-v3-multi, Llama-3.2-11B-Vision-Ins..., Llama-3.2-90B-Vision-In..., gpt-4, gpt-4-32k, jais-30b-chat, Phi-3.5-MoE-instruct, Phi-3-mini-128k-instruct, Phi-3-mini-4k-instruct, Phi-3-small-8k-instruct, and Phi-3-medium-128k-instruc... . At the bottom, there is a note "Image may not reflect actual user interface."



MaaS provides simplified management with ready-to-use GPU provisioning, lowering costs and reducing barriers to adoption by eliminating complexity.

# Grok 3 AI

Grok 3 AI is the latest breakthrough from xAI, now available in Azure.

- Advanced Reasoning
- Multimodal Capabilities
- Efficient Programming

The screenshot shows the Azure AI Foundry Model catalog page for the 'grok-3' model. At the top, there's a navigation bar with the Azure AI Foundry logo, 'Model catalog', and the specific model name 'grok-3'. Below the navigation, there's a back arrow labeled '← grok-3' and a 'PREVIEW' button. A horizontal line separates this from the main content area. In the main area, there are two buttons: 'Use this model' (which is highlighted with a blue background) and 'Fine-tune'. Below these buttons, there are two tabs: 'Details' (which is underlined, indicating it's active) and 'License'. The 'Details' tab contains a paragraph about Grok 3's capabilities, mentioning its non-reasoning nature, pre-training at supermassive scale, and its focus on enterprise domains like finance, healthcare, and legal. It also highlights its exceptional instruction following capabilities and purpose-built nature for common business use cases such as data extraction, coding, and text summarization. Another paragraph describes Grok 3's support for a 131,072 token context window, enabling it to handle extensive inputs while maintaining coherence and depth. The model was trained on a diverse dataset emphasizing high-quality, reasoning-rich content, making it particularly strong at drawing connections across domains and languages. Below this, there's a section for 'Model developer: xAI' and a 'Supported languages' section listing English, Spanish, French, Afrikaans, Arabic, Bengali, Welsh, German, Greek, Indonesian, Icelandic, and Italian. There's also a 'See more' link.

# Models-as-a-Service

## Integration with 3P Platforms



Dataloop



Arize AI



ClearML



Azure AI Foundry Models-as-a-Service

Exploring an exciting new frontier in Agentic AI with  
**Azure AI Foundry Agent Service**

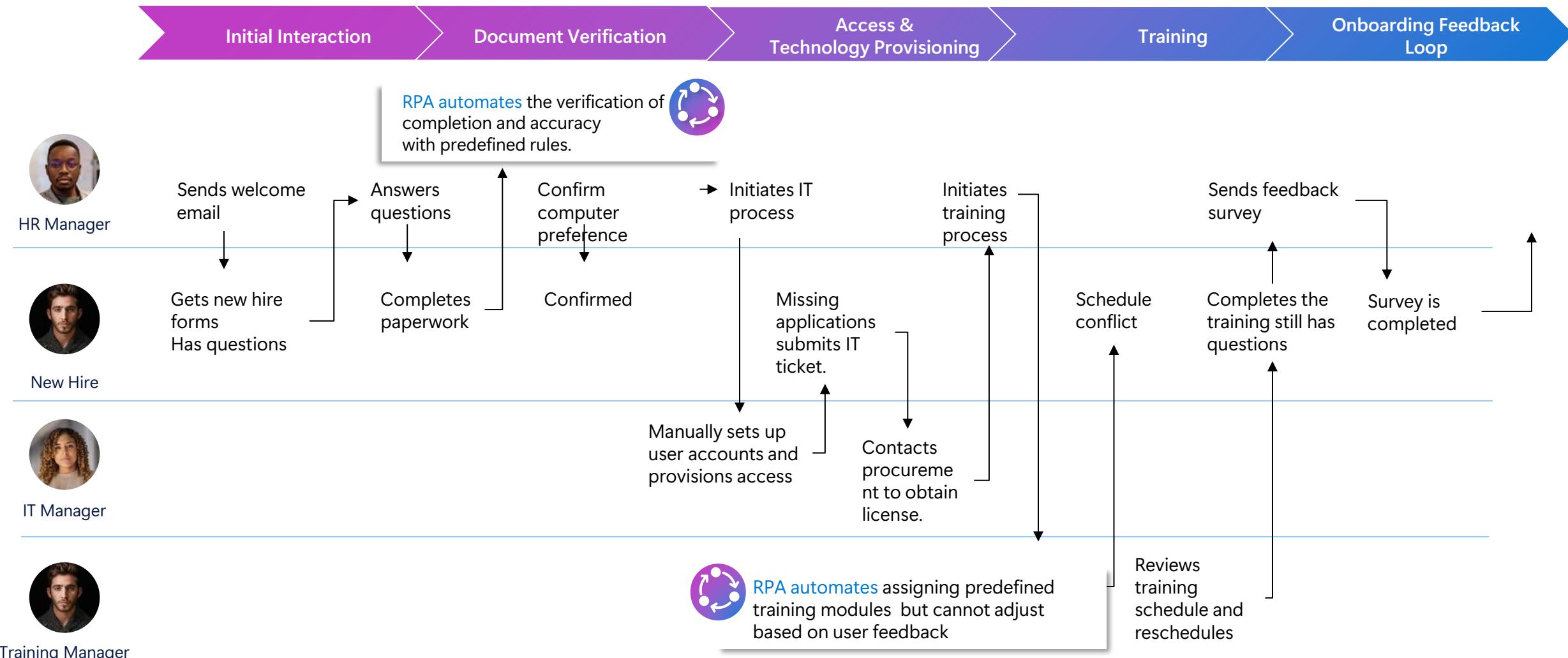


A new frontier:

# Agentic AI

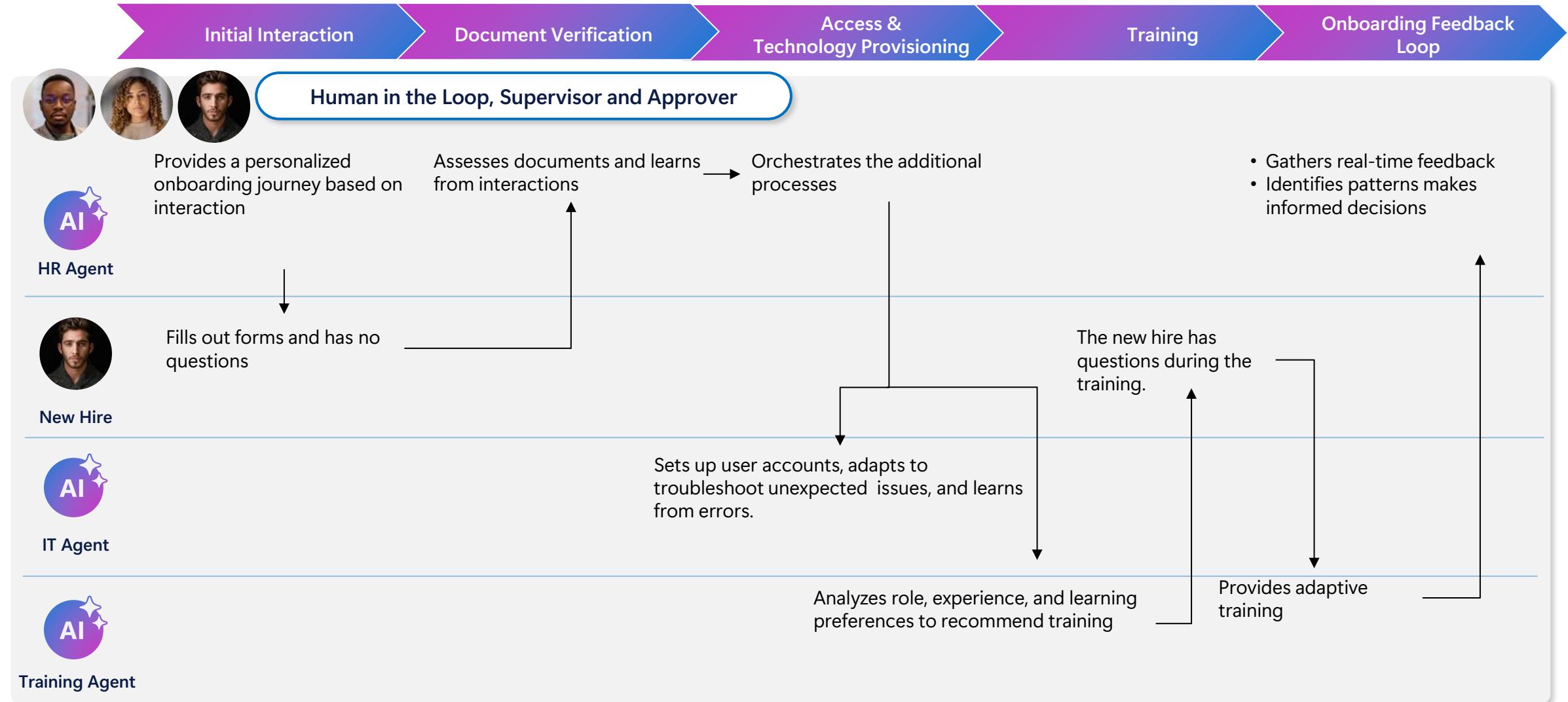
# TODAY

Existing solutions can only automate very specific tasks that have clear inputs and outputs

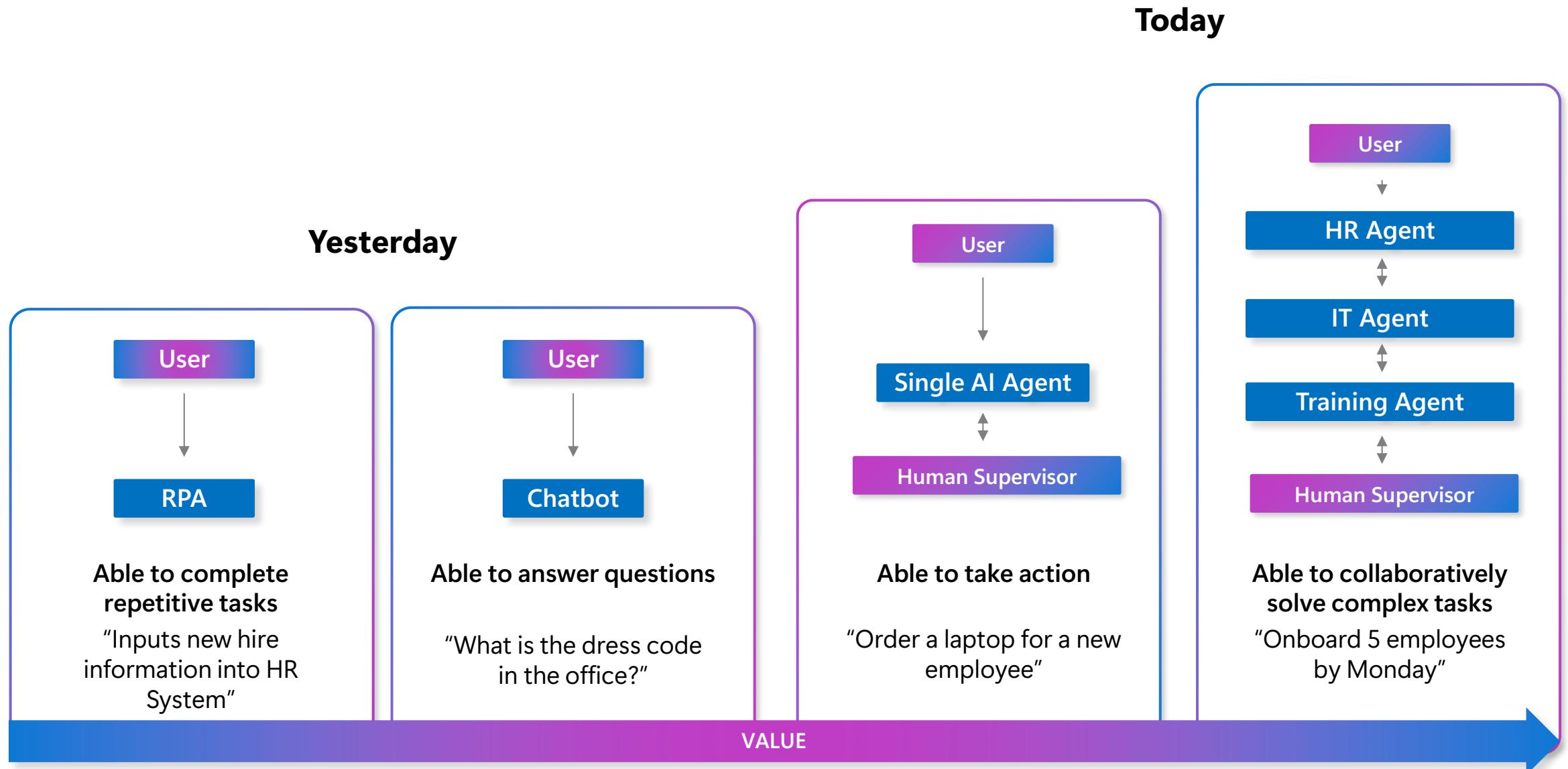


# TOMORROW

With AI Agents, these steps can be automated for the first time, while keeping human in the loop.

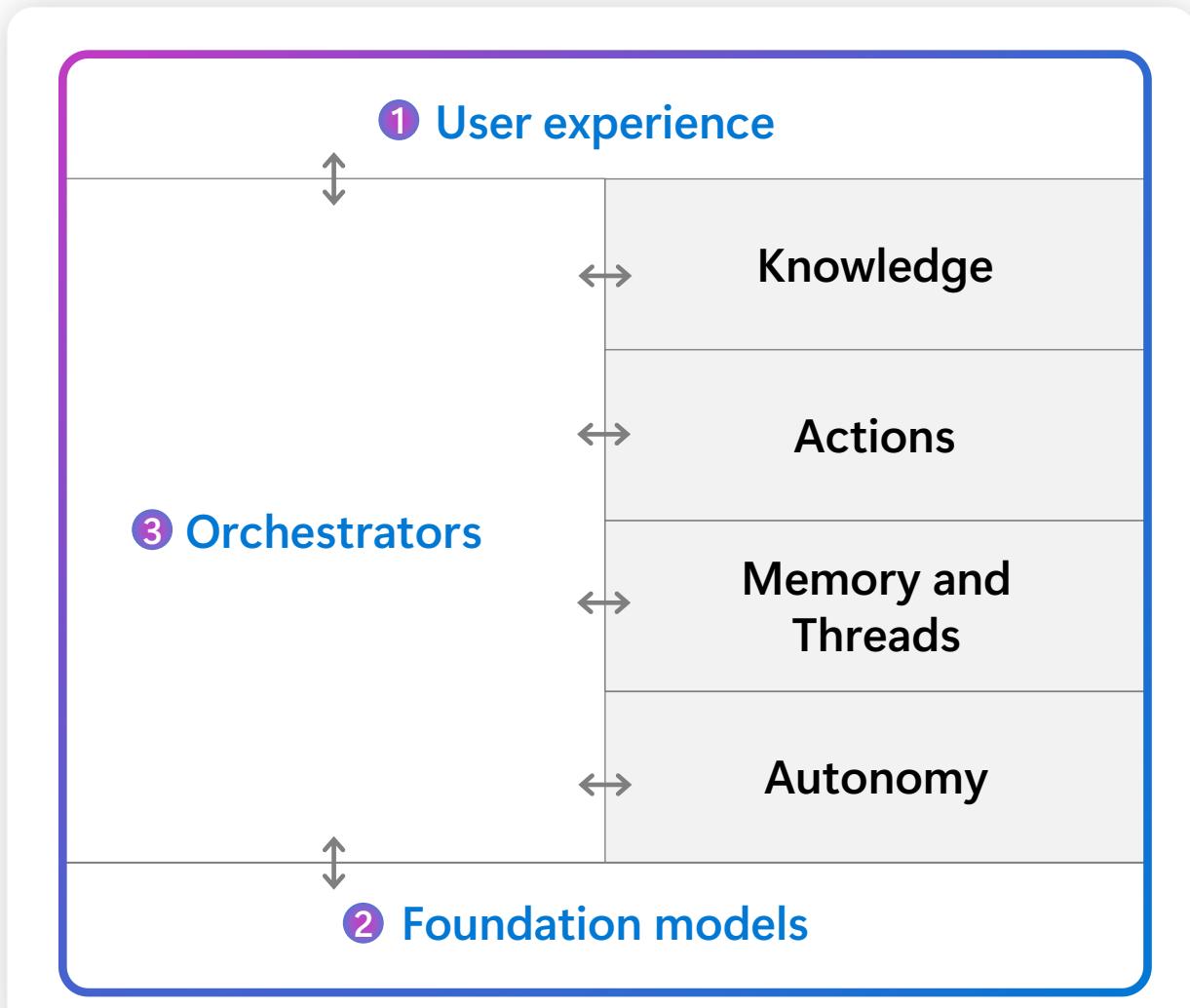


# Promising even more efficiency, value, and advantage

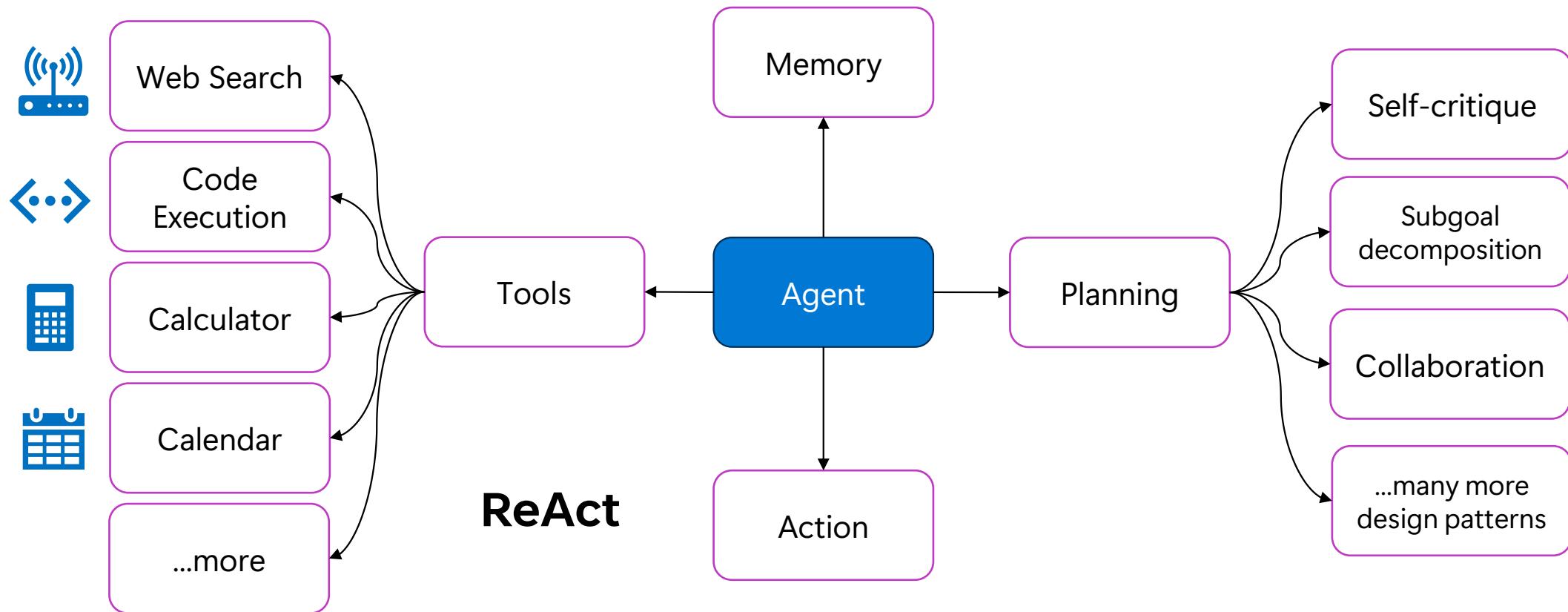


# Components of AI Agents

1. **User Experience:** The prompt that starts the whole process of the agent execution, as well as the human-agent interaction.
2. **Knowledge:** Search and retrieve information from online sources or company knowledge base, to ground the models.
3. **Actions:** Helps the agents perform certain actions (e.g. send an email, write a report) through connections to key applications.
4. **Memory and Threads:** Captures and stores the past interactions for hyper-personalisation and increased human-like interactions
5. **Autonomy:** Leaves the agents perform the tasks through an event-driven, trigger-based approach.
6. **Foundation Models:** Enables AI Agents to think throughout the process, helping to plan and reflect.
7. **Orchestrators:** Whether to be the client-side code or orchestrate across multiple agents across multiple cloud, orchestrators are key to bringing everything together.



# Agentic AI capabilities



# Common Design Patterns

RAG Agent



Code Generation  
Agent



Multi-agent  
Systems



Multi-domain  
Agent Systems



Learn more on Day 3

# Organizations need platforms that enable rapid development of performant, secure AI Agents

## Current Frameworks

Lack of integrated tools, insecure data grounding, challenging orchestration

Ineffective deployment of AI across websites, applications, and production environments

Restrictive pre-defined models that are challenging to customize

Security and data privacy risks

## What's Needed

Connected complex workflow automation grounded by seamless connection to enterprise data

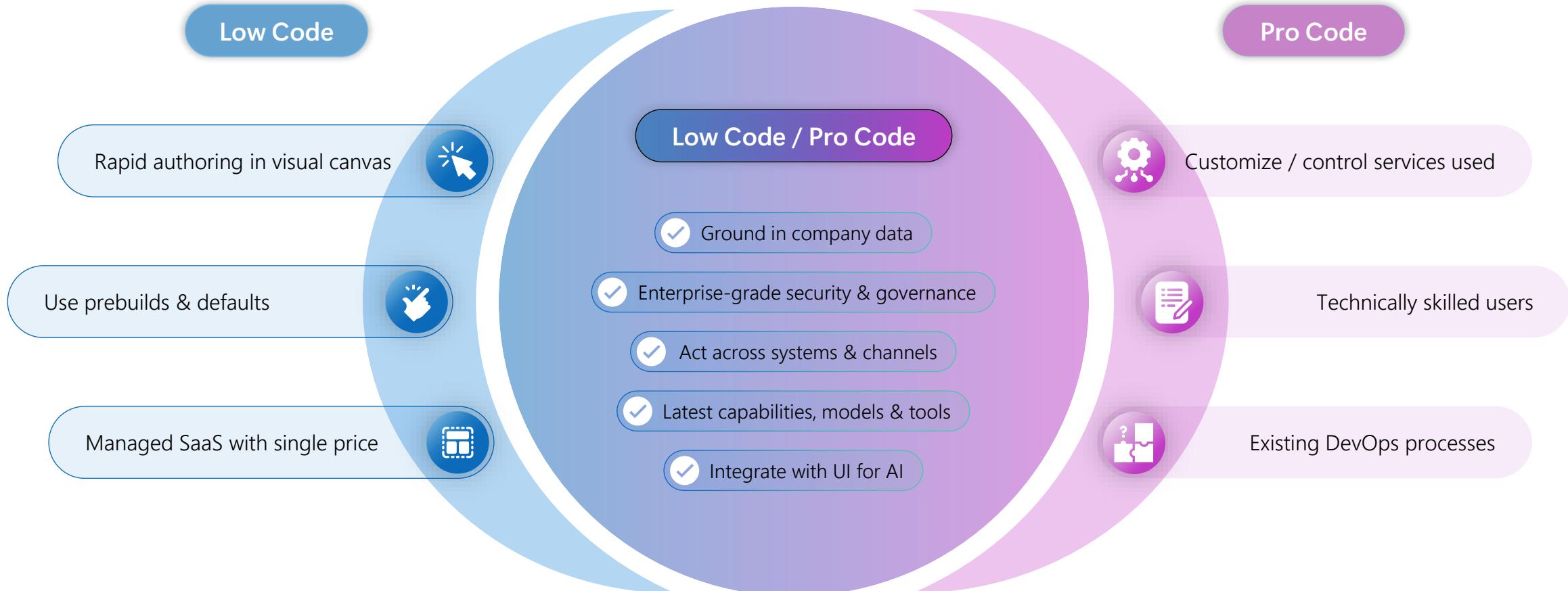
Tools and APIs that seamlessly integrate across enterprise applications

Flexible models that enable processing and integration of information from multiple modalities or types of data

Secure, responsible AI that protects sensitive information and behaves compliantly

# Choose the tool that best fits your needs

Most agents can be built using either of these differentiated tools – all powered by Azure AI Foundry

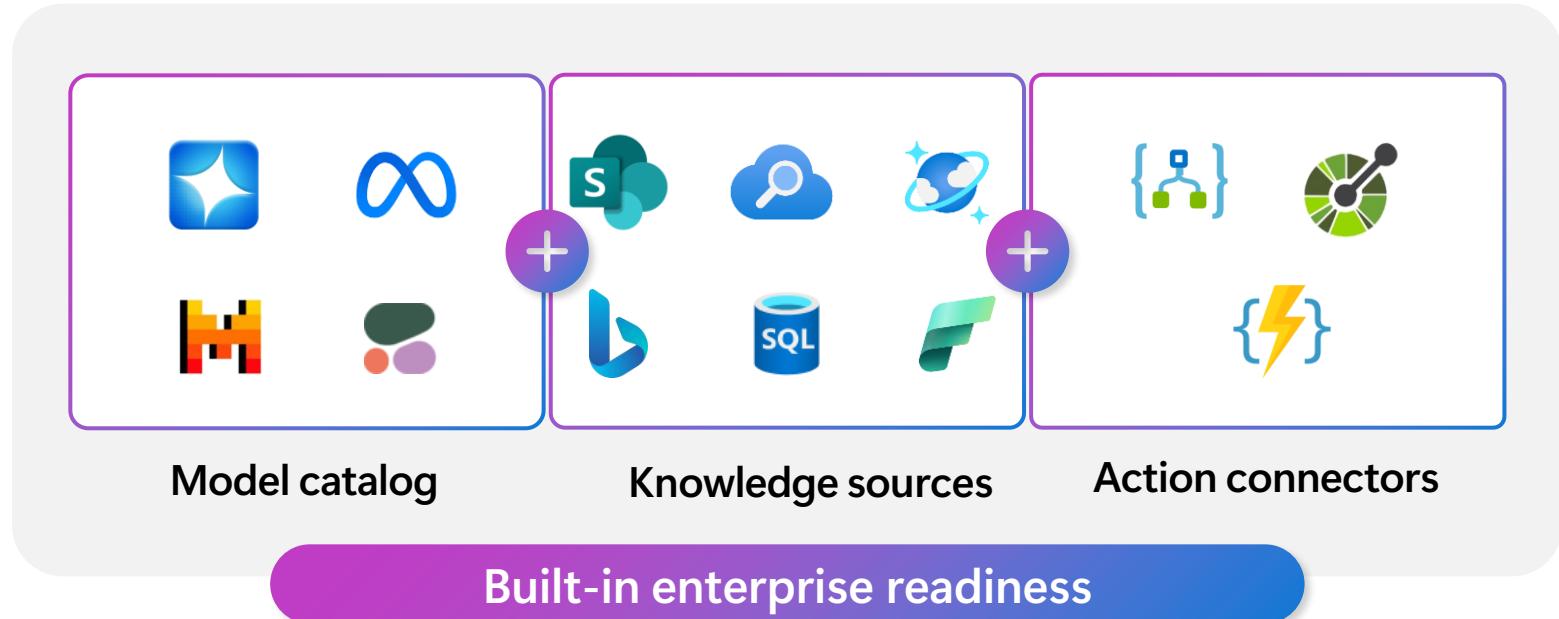


# Deploy and scale AI agents with ease

## Azure AI Foundry Agent Service:

- Enterprise-grade security
- Extensive data connections
- Flexible model selection
- Rapid development and automation

## Azure AI Foundry Agent Service



# The full enterprise package



## Azure AI Foundry Agent Service

### Trust



Customer control over data, networking, and security

- BYO-file storage
- BYO-search index
- BYO-virtual network
- BYO-thread storage
- Tracing/monitoring
- Evaluation

### Choice



Model choice and flexibility with the model catalog



#### Azure OpenAI Service

o-series & GPT-4.1 Models



#### Models-as-a-Service



Llama 3.1-405B-Instruct



Mistral Large, Small



Cohere-Command



DeepSeek v3

### Skills

Richest set of enterprise connectivity

#### Knowledge



#### Actions



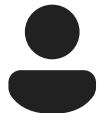
Logic Apps Azure Functions OpenAPI

Azure AI Foundry portal

Azure AI Foundry SDK

# How does your agent work?

"Help me book a trip to New York for a client meeting? I need to fly out next Monday and return on Friday."



User

"I've booked your trip to New York as requested. Here are details:..."



Travel Booking Agent



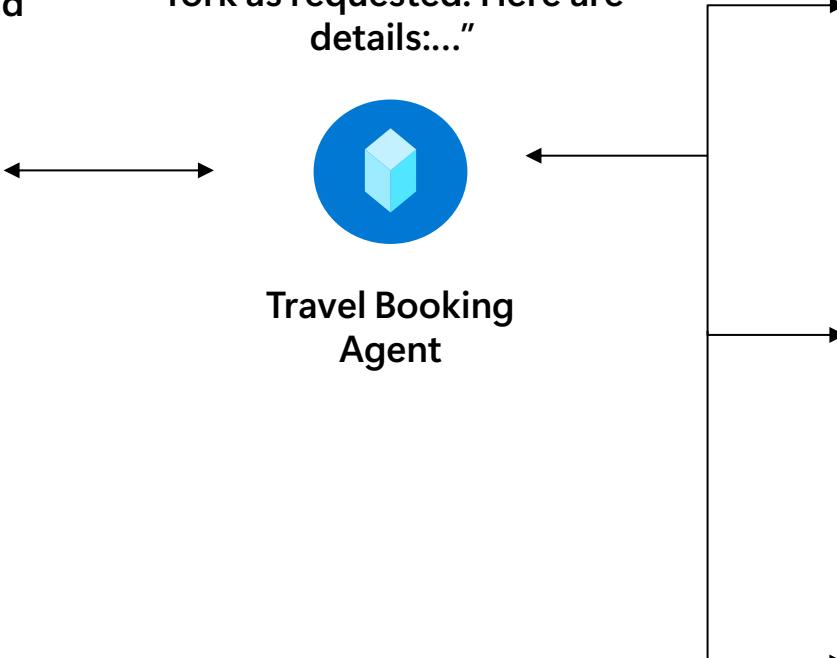
**Knowledge Sources**  
(search, files, databases, storage etc.)



**Models**  
(Azure OpenAI Service, Models-as-a-Service)



**Actions**  
(Pre-built or custom tools to automate processes)



Orchestrate and debug workflows

## Azure AI Foundry Agent Service

### Built-in enterprise readiness

BYO-file storage  
(coming soon)

BYO-search index

OBO Authorization Support

Enhanced Observability

### Model Catalog



Azure OpenAI Service  
(GPT-4o, GPT-4o mini)



### Models-as-a-Service



Llama 3.1-405B-Instruct



Mistral Large



Cohere-Command-R-Plus

### Extensive Ecosystem of Tools

#### Knowledge

Microsoft Fabric

SharePoint

Grounding with Bing Search

Azure AI Search

Files (local or Azure Blob)

Your own licensed data

File Search

Code Interpreter  
**Action**

Azure Logic Apps

OpenAPI 3.0 Specified Tools

Azure Functions

# Chat Completions API vs Assistants API vs Agents

## Chat Completions API

- Lightweight and powerful
- Inherently stateless

vs

## Assistants API

- Build with OpenAI models
- Stateful (inbuilt conversation state management)
- Access persistent threads
- Automatic management of the model's context window
- Access files in several formats
- Utilized Microsoft-managed storage
- File Search (API handles chunking, embeddings storage and creation, and implementing vector search)
- Code Interpreter
- Function Calling

vs

## Azure AI Foundry Agent Service

- **Assistants API features, plus**
- Build with a model of your choice (OpenAI, Llama, Mistral, Cohere...)
- Real time web-grounding with Bing
- Secure grounding on enterprise data in SharePoint and Fabric
- Bring Your Own Licensed data (Tripadvisor)
- Connect to 1400+ data sources and services with Azure Logic Apps
- Long running, event driven actions with Azure Functions
- Standardized OpenAPI 3.0 tools
- Bring Your Own Storage
- Bring Your Own Private Network
- Bring Your Own AI Search Resource
- Limitless scaling with PTUs
- Open Telemetry based tracing

# How does your agent work?

**Step 1:**  
Create an Agent

**Step 2:**  
Create a Thread

**Step 3:**  
Run the Agent

**Step 5:**  
Check the Run  
status

**Step 6:**  
Display the  
Agent's Response

**Agent**  
Travel Planning Agent

**Instructions**  
You are a travel booking and expense management assistant designed to help employees plan, book, and manage business travel.

**Model**



**Your data (optional)**

Azure AI Search

Files (local or Azure Blob)

**Tools (optional)**

File Search  
Code Interpreter  
Function Calling  
Bing Search  
Microsoft SharePoint (coming soon)  
Microsoft Fabric (coming soon)  
Azure Logic Apps (coming soon)  
Azure Functions  
OpenAPI 3.0 specified tools

**Thread**  
Travel Planning

**User's message**  
I need to book a hotel in New York for 2 stays.

**Agent's message**  
Here are some suggestions:

**Run 1**

1 Use Tripadvisor API to search the nearest hotel  
2 Create message

**User's message**  
What's the daily meal allowance for the business trip?

**Agent's message**  
The daily allowance for your business trip is \$75, as per company policy.

**Run 2**

1 Use Microsoft SharePoint to query the company travel policy  
2 Create message

# Trustworthy AI



# Safeguard with Trustworthy AI



Secure your data, AI apps, and models

Build end-to-end verifiable privacy into  
your AI apps

Utilize built-in AI safety tooling

The screenshot shows the Azure AI Foundry interface for creating a new evaluation. The left sidebar includes sections like Home, Model catalog, Playgrounds, AI Services, Build and customize, Code, Prompt flow, Fine-tuning, Monitor performance, Tracing, Evaluation (which is selected), and Content filters. Under 'Evaluation', there are sub-sections for My assets (Models + endpoints, Prompts, Data + indexes, Web apps) and Management center.

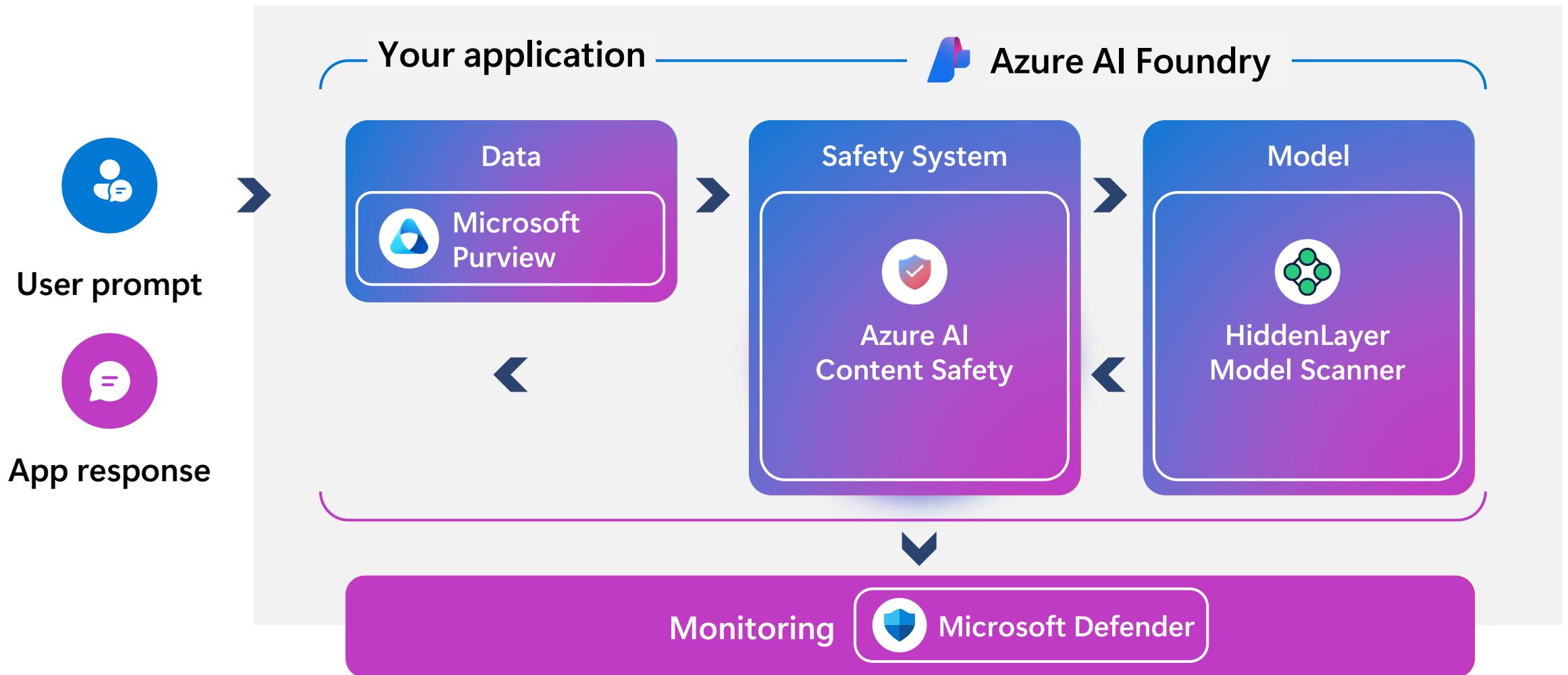
The main area is titled 'Create a new evaluation' and shows a step-by-step process: Basic information (done), Configure test data (done), Select metrics (in progress, indicated by a blue dot), and Review and finish (not yet started).

Under 'Select metrics', there are two sections: 'Performance and quality metrics curated by Microsoft' and 'Risk and safety metrics curated by Microsoft'. Each section contains several checkboxes for different metrics, each with a brief description.

- Performance and quality metrics curated by Microsoft:**
  - Groundedness: Measures how well the generative AI application's generated answers align with information from the input source.
  - Relevance: Measures the extent to which the generative AI application's generated responses are pertinent and directly related to the given questions.
  - Coherence: Measures how well the generative AI application can produce output that flows smoothly, reads naturally, and resembles human-like language.
  - Fluency: Measures the language proficiency of a generative AI application's predicted answer.
  - GPT similarity: Measures the similarity between a source data (ground truth) sentence and the generated response by a generative AI application.
  - F1 score: Measures the ratio of the number of shared words between the generative AI application prediction and the source data (ground truth).
- Risk and safety metrics curated by Microsoft:**
  - Self-harm-related content: Measures the predisposition of the generative AI application toward producing self-harm-related content.
  - Hateful and unfair content: Measures the predisposition of the generative AI application toward producing hateful and unfair content.
  - Violent content: Measures the predisposition of the generative AI application toward producing violent content.
  - Sexual content: Measures the predisposition of the generative AI application toward producing sexual content.
  - Protected material: Detects whether protected material is present in the AI content.
  - Indirect attack: Detects whether an indirect attack for XPLA is used in the content.

A help section on the right provides links to supported metrics, connection and deployment needs, and dataset mapping instructions.

# Built-in security and safety system



# Responsible by design

Because AI principles are not self-executing, we share our learnings and embed data-driven guardrails, guidance and best practices into Azure AI Foundry to help you operationalize trustworthy AI.

- Microsoft had early access to OpenAI models and gained valuable experience launching enterprise GenAI apps in the past two years—all built on Azure AI
- Microsoft has nearly 350 employees specializing in responsible AI at Microsoft, and we are investing to expand this number further
- Microsoft is a recognized leader in cloud platform services with highly secure, state-of-the-art Azure datacenters across 60+ announced regions
- Microsoft has committed to investing \$20 billion in cybersecurity over five years and we employ more than 8,500 security and threat intelligence experts
- Azure has one of the largest compliance certification portfolios in the industry and deep experience helping regulated industries and governments take advantage of AI technologies responsibly

## Microsoft's Responsible AI principles



Fairness



Reliability & Safety



Privacy & Security



Inclusiveness



Transparency



Accountability



Microsoft first adopted our six AI principles in 2018, and they continue to drive our policy, research, and engineering investments.

# Safety Models

## Update content filter

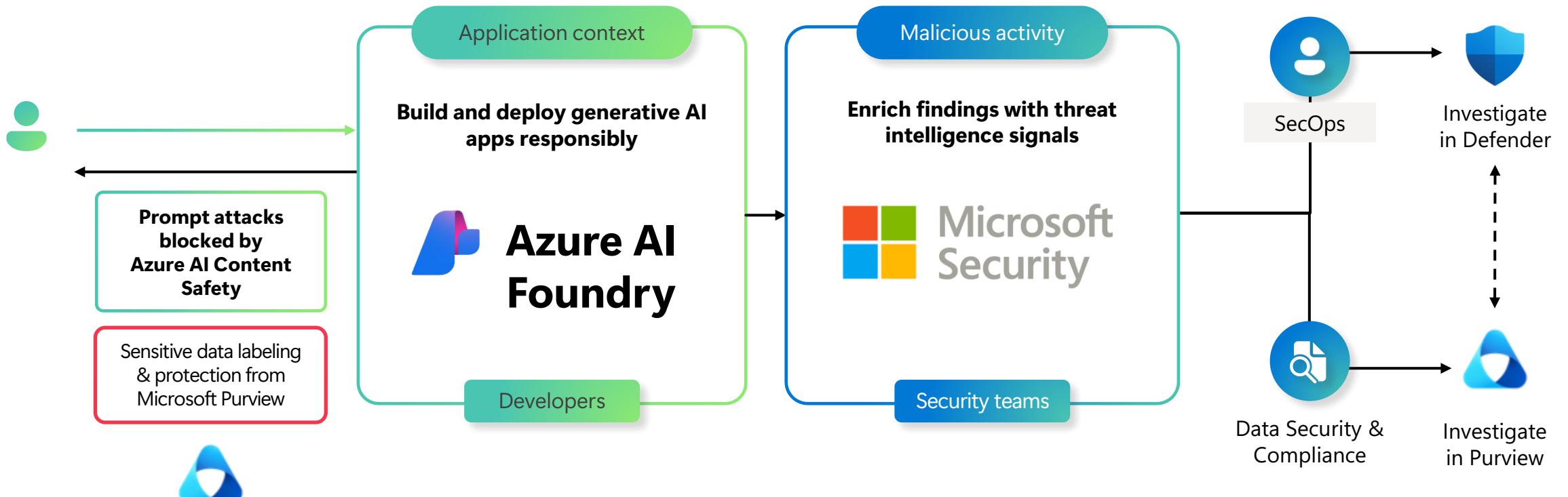
- Configure filters
- Additional models (Optional) - Preview
- Add blocklist (Optional) - Preview
- Streaming mode (Optional) - Preview
- Review and finish

### Additional models (Optional) - Preview

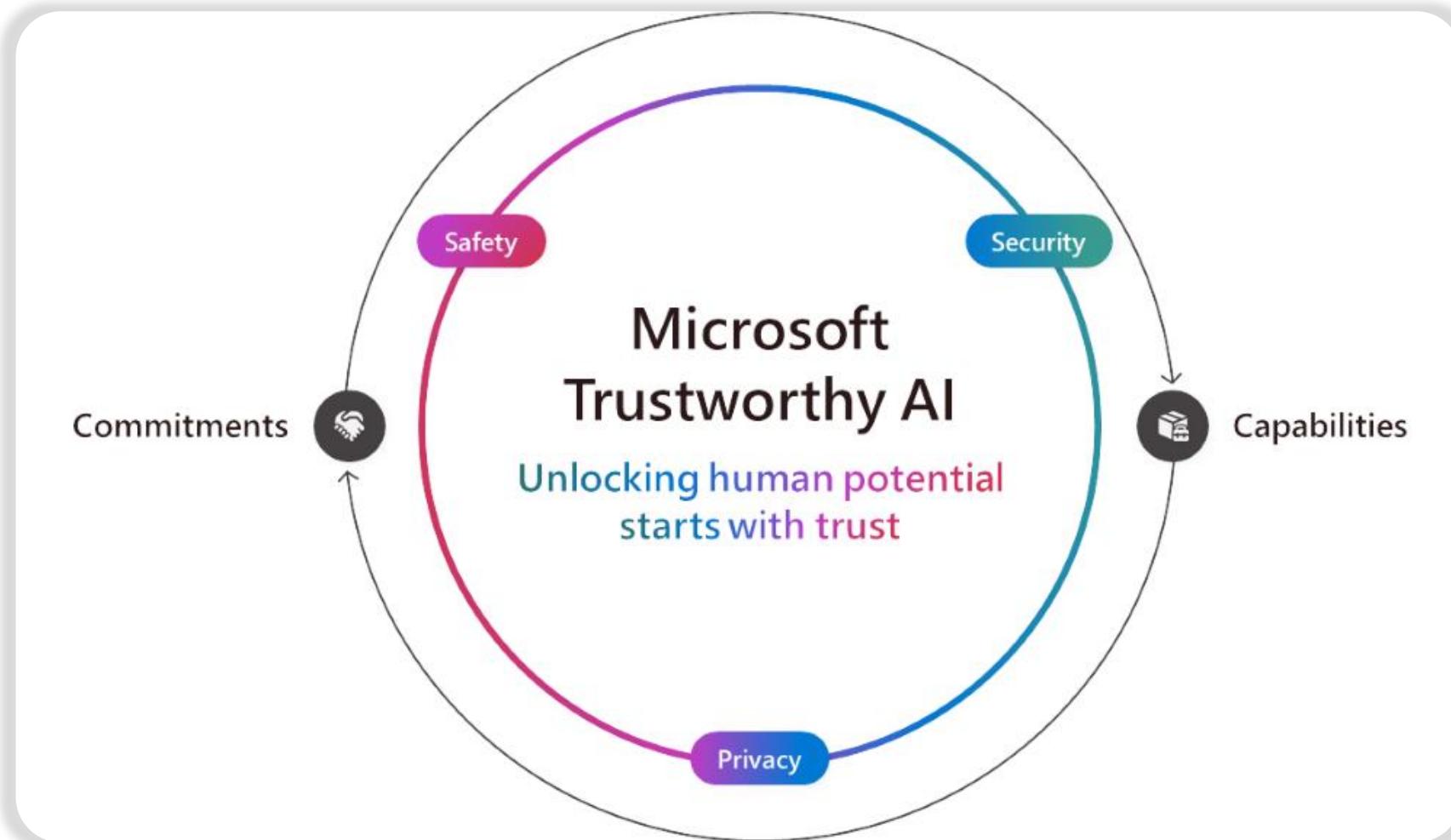
Enable additional content safety models that can be run on top of the prompts or completions (DALL-E, GPT-4 Turbo with Vision). [Learn more](#)

Enable/Annotate	Filter	Model
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> On	Prompt Shield for jailbreak attacks
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> On	Prompt Shield for indirect attacks
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> On	Protected material text
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> On	Protected material code

# Securely govern your AI applications



# Microsoft Trustworthy AI



# Security

Evaluations in Azure AI Studio to support proactive risk assessments.

Microsoft 365 Copilot will provide transparency into web queries to help admins and users better understand how web search enhances the Copilot response. Coming soon.

# Safety

Correction capability in Microsoft Azure AI Content Safety's Groundedness detection feature - helps fix hallucination issues in real time before users see them.

Embedded Content Safety - allows customers to embed Azure AI Content Safety on devices.

New evaluations in Azure AI Foundry - help customers assess the quality and relevancy of outputs and how often their AI application outputs protected material.

Protected Material Detection for Code (preview) - help detect pre-existing content and code

# Privacy

Confidential inferencing (preview) - in Azure OpenAI Service Whisper model, so customers can develop generative AI applications that support verifiable end-to-end privacy.

Azure OpenAI Data Zones for the EU and US - coming soon and build on the existing data residency provided by Azure OpenAI Service by making it easier to manage the data processing and storage of generative AI applications.

# Hands-on Labs



# Architecting Success with Multi-Agent AI Systems

⌚ Day 2 (5 hours)



⌚ 2 hours

## Module 1: Innovate with Microsoft 365 Copilot and agents

- M365 Copilot and agents
- How Copilot works
- Semantic index for Copilot
- M365 Copilot Chat
- Agents Use cases
- New agents in Microsoft 365
- Unlock more value with SharePoint agents

## Module 2: Extend Microsoft 365 Copilot with Agents

- Microsoft 365 Extensibility Planning and approach
- Declarative agents and agent tooling
- Explore Copilot Studio Agent Builder
- Build declarative agents with Microsoft 365 Agents Toolkit
- Build custom agents with Copilot Studio
- Autonomous agents overview
- Agent Governance - Overview
- Gen-AI decision guide – when to build, buy or extend



⌚ 3 hours

## Hands on labs

- Explore Copilot Studio Agent Builder
- Build HR Assistant Agent with Copilot Studio
- Incorporate actions in HR Agent
- Enable Autonomous Capabilities in Microsoft Copilot Studio for HR Activities



⌚ 2 hours

## Module 3: Customize Agents with Gen AI in Copilot Studio

- Customizing your agents – Orchestrator, UI, Knowledge, Actions, Autonomy
- Copilot Studio implementation guidance for architects
- Generative AI in Copilot Studio
- Copilot Studio + Power Platform
- Building voice-enabled agents
- AI Foundry integration
- Developing agents using Microsoft 365 Agents SDK



⌚ 3 hours

## Hands on labs

- Setup AI Project and perform Chat Completion from VS Code
- Build a simple AI Agent
- Develop a multi-agent system



## Solution Architect Assessment Purpose:

Evaluate the Solution Architect's competency in rapidly designing, architecting, and clearly communicating technical solutions leveraging Microsoft Copilot, Azure AI Foundry, Semantic Kernel, AutoGen, and related technologies.

Live- Instructor led

On demand

# Hands-on Lab



Please note that the labs are only open to a limited number of Microsoft partner participants and are offered on a **first-come-first-served basis**.

**Step 1** : Click on **Launch** in the event curriculum to launch your labs

Date/Time (PDT)	Title	Track	ICS   Training Bag
On-Demand	<a href="#">Spektra Hands-on lab: Agentic AI Accelerator   AMER (PDT)</a>	AMER (PDT)	<a href="#">Launch</a>

**Step 2** : Join the **Live session for Lab Support and Q&A Session**

<a href="#">Lab Support and Q&amp;A</a>	<a href="#">Lab Support and Q&amp;A</a>		<a href="#">Sign Up</a>	
<p><b>Abstract:</b> We invite you to join this session to begin your labs. The Trainer will provide you with instructions, demonstrate live scenarios and respond to your questions.</p>				

\*For all Lab logistical support issues, please go to: <http://cloudlabs.ai/ms-support/> and contact via chat

# Sign-up and launch Hands-on Labs

Note: Please register and launch the lab only when the hands-on lab session is about to begin (not at the start of the day)



- Click the link to complete your registration -  
<https://bit.ly/3Hfoc0>



- Enter the Email id and **Activation code** (based on your time zone).

IST	AI2084IST
BST	AI2084BST
PDT	AI2084PDT



- Select the preferred language, accept the **Usage policy** and select **Submit** to complete registration.



- Once registration is completed, click on **Launch Lab** to activate the workshop.

Register Now

Email\*

Activation Code\*

Choose one lab based on your preferred language. Selection is final; other labs will be locked.\*

Azure Agentic AI Accelerator Workshop - English

Azure Agentic AI Accelerator Workshop - Spanish

Azure Agentic AI Accelerator Workshop - Japanese

Azure Agentic AI Accelerator Workshop - Korean

Azure Agentic AI Accelerator Workshop - Simplified Chinese

Azure Agentic AI Accelerator Workshop - Traditional Chinese

Azure Agentic AI Accelerator Workshop - Portuguese

Microsoft or training providers may use your contact information to provide updates and special offers about Microsoft Azure and other Microsoft products and services. You can unsubscribe at any time. To learn more, you can read the [Privacy Policy](#)

**Submit**

Please click on the 'Launch Lab' button to activate your lab environment.

**Launch Lab**

# FAQs - Lab Experience

## Lab Guide (First Tab)

Provides step-by-step instructions for performing the lab exercises.

The screenshot shows the 'Guide' tab selected. It displays the 'Azure Agentic AI Workshop Overview' which describes the workshop series. Below it, the 'Objectives' section lists several learning goals:

- Learn to design and deploy AI agents with Microsoft Copilot Studio and Azure AI services.
- Gain hands-on experience in multi-agent orchestration using the Azure AI Agent Service SDK and Semantic Kernel.
- Build custom Retrieval-Augmented Generation (RAG) applications with Azure AI Foundry and integrate Semantic Kernel plugins.
- Evaluate, fine-tune, and deploy AI models using Prompt Flow for real-world use cases.
- Construct intelligent escalation systems using conversational interfaces, event-driven architecture, and AI-powered workflows.

## Environment (Second Tab)

Lists the credentials required to access the environment and the lab

The screenshot shows the 'Environment' tab selected. It contains a message about logging into Microsoft Azure. Below it is a table for 'Auth Fields' with two entries: 'Username' and 'Password'. A red box highlights these two entries. Further down, it shows 'Resource Group : lab-vm' with a table containing 'Deployment ID' and 'Labvm Admin Username'.

Auth Fields	Value	Action
Username	odl_user_1734975@msa	copy
Password	xtnt26BED*9K	copy

Key	Value	Action
Deployment ID	1734975	copy
Labvm Admin Username	demouser	copy

## Resources (Third Tab)

Allows you to manage resources (e.g., Start, Restart, or Stop VMs.)

The screenshot shows the 'Resources' tab selected. It displays a table for 'Virtual Machines' with one entry: 'labvm-1734975' which is 'VM running'. The 'Actions' column for this entry is highlighted with a red box.

Name	Status	Actions
labvm-1734975	VM running	[Start] [Stop] [Restart] [Delete]

# FAQs – Lab Experience

## Help (Fourth Tab)

Help section to troubleshoot basic or known issues related to the workshop

The screenshot shows the 'Help' tab selected in a navigation bar. Below it, a section titled 'Troubleshooting known issue' lists a common issue: 'Unable to copy paste'. The 'Recommended quick fix' is to click on the SSL certificate symbol, open the pop-up, and change the clipboard dropdown to allow. It also notes that once clipboard access is enabled, users can use specific keyboard shortcuts. A table at the bottom provides these shortcuts for Windows OS.

Attendee OS	VM OS	Copy Shortcut	Paste Shortcut
Windows	Windows	Ctrl + C	Ctrl + V

## VM native Clipboard

Enable clipboard which helps to copy paste the values from Lab Guide to VM

The screenshot shows the 'VM Native Clipboard' toggle switch being turned on. Below it, the same troubleshooting section and clipboard shortcut table are visible.

Attendee OS	VM OS	Copy Shortcut	Paste Shortcut
Windows	Windows	Ctrl + C	Ctrl + V

## Access Lab Now – Alternate method

Access the lab in-case of accidental closure of the browser tab

This page contains details for accessing the lab via an alternate method. It includes a table for 'On Demand Lab: GitHub Copilot Innovation Workshop' with fields for Name and Value, and another for 'JumpVM-RG-1487427' with fields for Name and Value. It also lists 'Licenses' with their types and values. A prominent blue button labeled 'Access Lab Now' is at the bottom, which is also highlighted with a red box.

On Demand Lab: GitHub Copilot Innovation Workshop

Please use the below details for future use in your labs:

JumpVM-RG-1487427 :

Name	Value
DeploymentId	[REDACTED]
vm Admin Password	[REDACTED]
windows VM DNS Name	[REDACTED]
gitHub User Email	[REDACTED]
gitHub User Password	[REDACTED]

Please use the below Licenses details if required during the lab:

Licenses Type	Licenses
GCW - GitHub UserName	[REDACTED]
GCW - GitHub Password	[REDACTED]

**Access Lab Now**

If you have any questions, please contact us at [cloudlabs-support@spektrasyystems.com](mailto:cloudlabs-support@spektrasyystems.com)

This email is sent by Spektra Systems LLC, on behalf of Microsoft.  
You are receiving this message as you have registered for On Demand Lab at <https://experience.cloudlabs.ai>.

# Troubleshooting common issues

Get guidance to troubleshoot some of the most common errors you may face while performing your labs.

<https://docs.cloudlabs.ai/troubleshooting/Overview>

The screenshot shows a web browser displaying the 'Common Troubleshooting' section of the CloudLabs documentation. The left sidebar has a navigation menu with links like 'Introduction', 'Learner', 'Instructor', 'Admin', 'Lab Requester', 'MCT Pass Program', 'Common Troubleshooting' (which is expanded to show 'Overview', 'RDP: Known Functionality Issues', 'Copy Paste', 'Unable to launch Cloud Shell', 'Add Safe Sender', 'Login with Azure AD - Consent Experience', 'General FAQs', 'Contact Support', and 'Databricks Lab Environments'), and 'Databricks Lab Environments'. The main content area has a breadcrumb trail: Home > Common Troubleshooting > Overview. The title 'Overview' is displayed in large bold letters. Below the title is a paragraph of text: 'Here are some of the most common errors you may face while performing your labs. Please perform the given steps to troubleshoot the issue. For any other additional queries, reach out to your Instructor or program contact.' A table titled 'Common Issues' lists various troubleshooting topics with 'Start Here' links. At the bottom, there are 'Previous' and 'Next' navigation links.

Common Issues	Guide
Troubleshooting Checklist	<a href="#">Start Here</a>
Fix "Websocket connection blocked" to access HTTPS over RDP	<a href="#">Start Here</a>
Copy & Paste functionality not working in lab environment	<a href="#">Start Here</a>
Unable to launch Cloud Shell	<a href="#">Start Here</a>
Add Safe Sender	<a href="#">Start Here</a>
Login with Azure AD - Consent Experience	<a href="#">Start Here</a>

Previous: [« Getting Started](#)

Next: [RDP: Known Functionality Issues »](#)



# Architecting Success with Multi-Agent AI Systems

⌚ Day 2 (5 hours)



⌚ 2 hours

## Module 1: Innovate with Microsoft 365 Copilot and agents

- M365 Copilot and agents
- How Copilot works
- Semantic index for Copilot
- M365 Copilot Chat
- Agents Use cases
- New agents in Microsoft 365
- Unlock more value with SharePoint agents

## Module 2: Extend Microsoft 365 Copilot with Agents

- Microsoft 365 Extensibility Planning and approach
- Declarative agents and agent tooling
- Explore Copilot Studio Agent Builder
- Build declarative agents with Microsoft 365 Agents Toolkit
- Build custom agents with Copilot Studio
- Autonomous agents overview
- Agent Governance - Overview
- Gen-AI decision guide – when to build, buy or extend



⌚ 3 hours

## Hands on labs

- Explore Copilot Studio Agent Builder
- Build HR Assistant Agent with Copilot Studio
- Incorporate actions in HR Agent
- Enable Autonomous Capabilities in Microsoft Copilot Studio for HR Activities



⌚ 2 hours

## Module 3: Customize Agents with Gen AI in Copilot Studio

- Customizing your agents – Orchestrator, UI, Knowledge, Actions, Autonomy
- Copilot Studio implementation guidance for architects
- Generative AI in Copilot Studio
- Copilot Studio + Power Platform
- Building voice-enabled agents
- AI Foundry integration
- Developing agents using Microsoft 365 Agents SDK



⌚ 3 hours

## Hands on labs

- Setup AI Project and perform Chat Completion from VS Code
- Build a simple AI Agent
- Develop a multi-agent system



## Solution Architect Assessment Purpose:

Evaluate the Solution Architect's competency in rapidly designing, architecting, and clearly communicating technical solutions leveraging Microsoft Copilot, Azure AI Foundry, Semantic Kernel, AutoGen, and related technologies.

Live- Instructor led

On demand

# Architecting Success with Multi-Agent AI Systems

Solution Architecture

## Solution Architecture Assessment

⌚ 4hr



### Purpose:

Evaluate the Solution Architect's competency in rapidly designing, architecting, and clearly communicating technical solutions leveraging Microsoft Copilot, Azure AI Foundry, Semantic Kernel, AutoGen, and related technologies.



### Assessment Components:

This assessment will consist of **two parts**:

- **Part A:** Scenario-Based Knowledge Assessment (Multiple-choice/Short-answer questions)
- **Part B:** Solution Architecture Design Evaluation



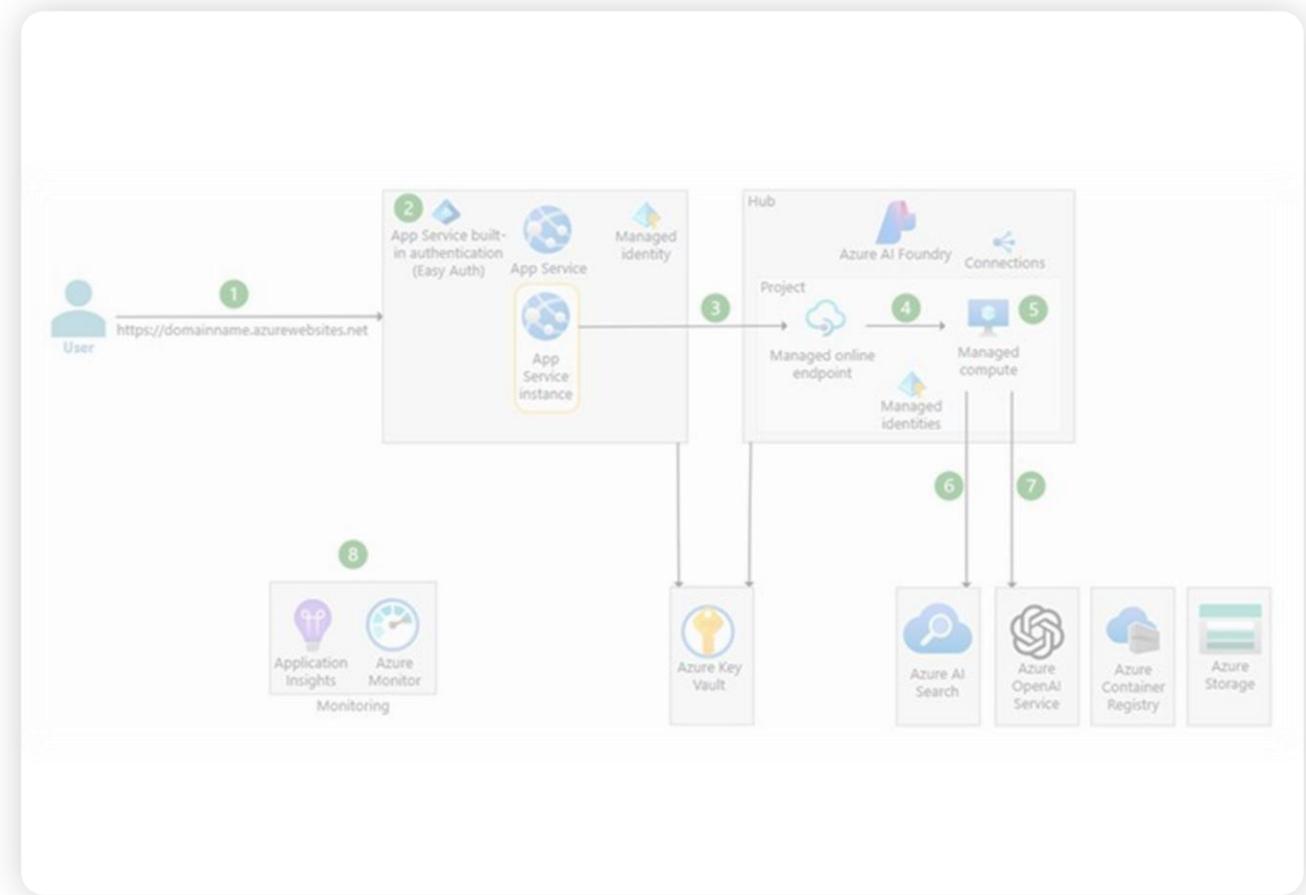
### Success Criteria for Assessment:

#### Theoretical Assessment (Part A):

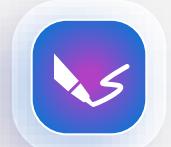
- Demonstrate clear understanding of Azure AI services, Semantic Kernel, AutoGen, and Copilot.
- Accurate and relevant responses aligned with industry best practices.

#### Solution Architecture Design Evaluation(Part B):

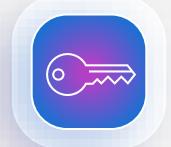
- Logical and robust solution design clearly addressing provided scenario.
- Effective integration of Microsoft 365 Copilot, Azure AI Foundry, and other relevant services.
- Strong articulation of security, compliance, and governance considerations.
- Clear, persuasive communication of solution value and benefits.



How to take  
**Solution Architect Assessment**



Register on LevelUp  
[skillupwithlevelup.com](https://skillupwithlevelup.com)



Use Access code – **OKDV-KHFL**



Take the Solution Architect assessment



Receive the Agentic AI Solution Architecture  
completion badge\*

\*badges will be awarded if the qualification criteria is met



# Get Agentic AI “Solution Architect Champion” badge

## Badge Qualification criteria



Attend “Agentic AI Accelerator” live session

- **Day 1**-Showcasing AI Potential with Agentic AI
- **Day 2**- Architecting Success with Multi-Agent AI Systems



Execute and complete Day 1 & Day 2 Hands-on lab exercises.



Take the Solution Architect assessment and pass with a minimum score of 70%.



Receive the Agentic AI “Solution Architect Champion” badge.



Receive the Agentic AI “Solution Architect Champion” badge by **30<sup>th</sup> June 2025**.

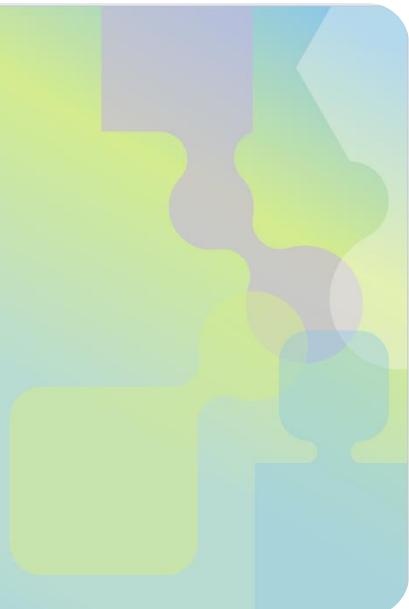
# Continue your learning deal-readiness journey with more ...



Microsoft Partner Project Ready

## Azure OpenAI Workshop

Duration 4 days (4 hrs/day)



Microsoft Partner Project Ready

**Build and Extend your own agents using pro-code capabilities**

Duration: 3 days (4hrs/day)



Explore and Register on LevelUp  
[skillupwithlevelup.com](http://skillupwithlevelup.com)

# Coming up next...

## Day 1

### Showcasing AI Potential with Agentic AI

#### Module 1: Innovate with Microsoft 365 Copilot and agents

- M365 Copilot and agents
- How Copilot works
- Semantic index for Copilot
- M365 Copilot Chat
- Agents Use cases
- New agents in Microsoft 365
- Unlock more value with SharePoint agents

#### Module 2: Extend Microsoft 365 Copilot with Agents

- Microsoft 365 Extensibility Planning and approach
- Declarative agents and agent tooling
- Explore Copilot Studio Agent Builder
- Build declarative agents with Microsoft 365 Agents Toolkit
- Build custom agents with Copilot Studio
- Autonomous agents overview
- Agent Governance - Overview
- Gen-AI decision guide – when to build, buy or extend

#### Hands-on Labs

- Explore Copilot Studio Agent Builder
- Build HR Assistant Agent with Copilot Studio
- Incorporate actions in HR Agent
- Enable Autonomous Capabilities in Microsoft Copilot Studio for HR Activities

## Day 2

### Architecting Success with Multi-Agent AI Systems

#### Module 3: Customize Agents with Gen AI in Copilot Studio

- Customizing your agents – Orchestrator, UI, Knowledge, Actions, Autonomy
- Copilot Studio implementation guidance for architects
- Generative AI in Copilot Studio
- Copilot Studio + Power Platform
- Building voice-enabled agents
- AI Foundry integration
- Developing agents using Microsoft 365 Agents SDK

#### Module 4: Innovate with Azure AI Platform

- How language models work
- AI Foundry and SDK introduction
- AI Foundry Model Catalog
- Azure AI Services
- Azure OpenAI Service and model guidance
- Models-as-a-Service
- Azure AI Foundry Agent Service
- Safeguard with Trustworthy AI

#### Hands-on Labs

- Setup AI Project and perform Chat Completion from VS Code
- Build a simple AI Agent
- Develop a multi-agent system

## Day 3

### Multi-Agent AI: Advanced Agent Dev in Azure AI Foundry

#### Module 5: Customize, orchestrate and experiment with Azure AI Foundry

- Retrieval Augmented Generation (RAG)
- Customizing models – Fine tuning, distillation
- Responses API (preview)
- Azure AI Foundry Agent Service - Orchestrate and debug AI workflows

#### Module 6: Build your own multi agents with Semantic Kernel or AutoGen

- Multi-agent applications
- Understanding Semantic Kernel
- Understanding AutoGen Agents Framework
- Multi-Agent Collaboration & Orchestration with AutoGen / Semantic Kernel

#### Hands-on Labs

- Set Up Azure AI Foundry SDK and Provision Resources
- Build a Retrieval-Augmented Generation(RAG) Pipeline
- Evaluate and Optimize RAG Performance
- Semantic Kernel Fundamentals
- Semantic Kernel Plugins

## Day 4

### Enterprise Grade: Optimization and production at scale

#### Module 7: Enterprise grade production at scale

- Scaling challenges and agent controls
- Manage AI performance in production
- Observability Tools
- Enabling Enterprise governance and management
- Enterprise grade security and data protection
- Monitoring and observability

#### Module 8: Advanced AI risk evaluation and mitigation

- Identifying risks
- Azure AI Content Safety
- Evaluation and GenAIops
- Identity and access management
- Network Security for AI apps
- Continuous security for AI

#### Hands-on Labs

- Understanding the Lifecycle of Flow Development
- Building and Customizing Prompt Flows
- Evaluation Flow Setup
- Fine-Tuning Prompts for Optimal Performance
- Implementing Chat Flow and Tool Integration
- Ensuring Responsible AI Practices with Content Safety



# Thank You