



# Transform with Azure AI, Data and Apps Platform

# Session 1a

## Azure AI Product Family

A Comprehensive, End to End Platform for AI Apps & Services



### Azure AI Studio

Comprehensive user interface (UI), SDK, CLI



**Azure AI Services**  
Intelligent services



**Azure AI Search**  
Modern retrieval



**Azure ML**  
Build and train models

Turing  
Florence

GPT-4 and GPT 3.5 Turbo  
Whisper

DALL-E  
Embeddings

Meta Llama 3  
Hugging Face

Azure infrastructure

# Session 1b



## API and Model Choice

Identify the best AI services and models for your use case

### **Build intelligent apps with industry-leading APIs\***

Develop multimodal, multi-lingual AI with out-of-the-box and customizable APIs and models.

### **Discover models for your use case**

Choose from the latest cutting-edge foundation and open models. Benchmark models by performance and key parameters and deploy what's right for your use case.

### **Deploy serverless models**

Get started quickly with ready-to-use APIs, without the need to provision GPUs.

# Session 2



## Complete AI Toolchain

Access collaborative, comprehensive tooling to support the development lifecycle and differentiate your apps

---

### **Ground models on your protected data**

Use your protected data to customize models with advanced fine-tuning and for contextually relevant retrieval augmented generation.

---

### **Orchestrate and debug AI workflows**

Streamline app development with easy-to-use prompt orchestration, tracing, and debugging via interactive visual and code-first workflows.

---

### **Streamline model and app evaluations**

Systematically measure and improve app performance with manual and automated evaluations.

# Session 3



## Responsible AI practices

### Design and safeguard applications

#### **Design apps responsibly**

Confidently build apps with technologies, templates, and best practices to help manage risk, improve accuracy, protect privacy, reinforce transparency, and simplify compliance.

#### **Safeguard with configurable filters and control**

Detect and filter harmful content, protect PII, and safeguard applications against prompt attacks.

# Session 4



## Enterprise-grade Production at Scale

Deploy AI innovations to Azure's managed infrastructure with continuous monitoring and governance across environments

### **Deploy to production**

Scale AI for use in websites, applications, and other production environments.

### **Operationalize and monitor workflows**

Continuously monitor AI safety, quality, and token consumption in production. Automate workflows and alerts for timely issue resolution.

### **Enable developer agility and enterprise governance at scale**

Provide easy project creation and resource management across the organization and enterprise controls for security, privacy, and compliance.

# Session 5

## Choose where to build your copilot

Three approaches to support the developer experience



**Copilot Studio**  
*Power Platform*

- Build your own copilot using visual building experiences
- Customize Microsoft Copilots with your own enterprise scenarios
- Leverage a connected, integrated platform
- Manage copilot experiences with full visibility into customizations



**Azure AI Studio**  
*Azure*

- Explore, build, test, deploy, and manage custom copilots using interactive visual and code-first workflows
- Access out-of-the-box and customizable APIs and models
- Systematically evaluate model and app responses and pinpoint fine-tuning opportunities
- Scale copilots for use in websites, applications, and other production environments



**AI Toolkit for Visual Studio Code**  
*Agnostic*

- Simplify app development by bringing together AI development tools and models
- Explore, try, and integrate AI models
- Use local and cloud compute to fine-tune and optimize small language models for app-specific use cases on the cloud and edge
- Package fine-tuned models as containers and deploy to Azure or the edge

# The Future of Work with AI



# The future of AI is here

Forbes

What ChatGPT And Generative AI Mean For Your Business?

COMPUTERWORLD

Microsoft's new Teams Premium tier integrates with OpenAI's GPT-3.5

MarketWatch

Microsoft's Nadella: AI is taking the computer age from 'the bicycle to the steam engine'

The Washington Post

Meet Windows Copilot, the AI coming to help you understand your PC

techradar.pro

Microsoft Fabric looks to offer the next generation of AI analytics for your business

TC TechCrunch

Microsoft's Azure AI Studio lets developers build their own AI 'copilots'

VentureBeat

Microsoft announces generative AI-powered Copilot 365 to 'change work as we know it'

ON BUSINESS.

Real estate agents say they can't imagine working without ChatGPT now

THE VERGE

Microsoft's AI-powered Copilot is getting plug-ins

# Organizations are ready to embrace the transformational potential of AI

**87%**

Of organizations believe AI will give them a competitive edge

**40%**

In a recent study, skilled workers using generative AI saw a 40% boost in the quality of their work

For every \$1 a company invests in AI, it is realizing an average return of

**\$3.5**

**14 months**

Average time it takes for organizations to realize a return on their AI investment

# What could slow down generative AI adoption?

## Getting started

The state of the art is evolving so quickly, it makes it difficult to decide what to use. Along with that, guidance and documentation is hard to find

## Development

Applications often require multiple cutting-edge products and frameworks which requires specialized expertise and new tools to stitch these components together

## Context

Generative AI doesn't know about your data

## Evaluation

It is hard to figure out which model to use and how to optimize for their use case

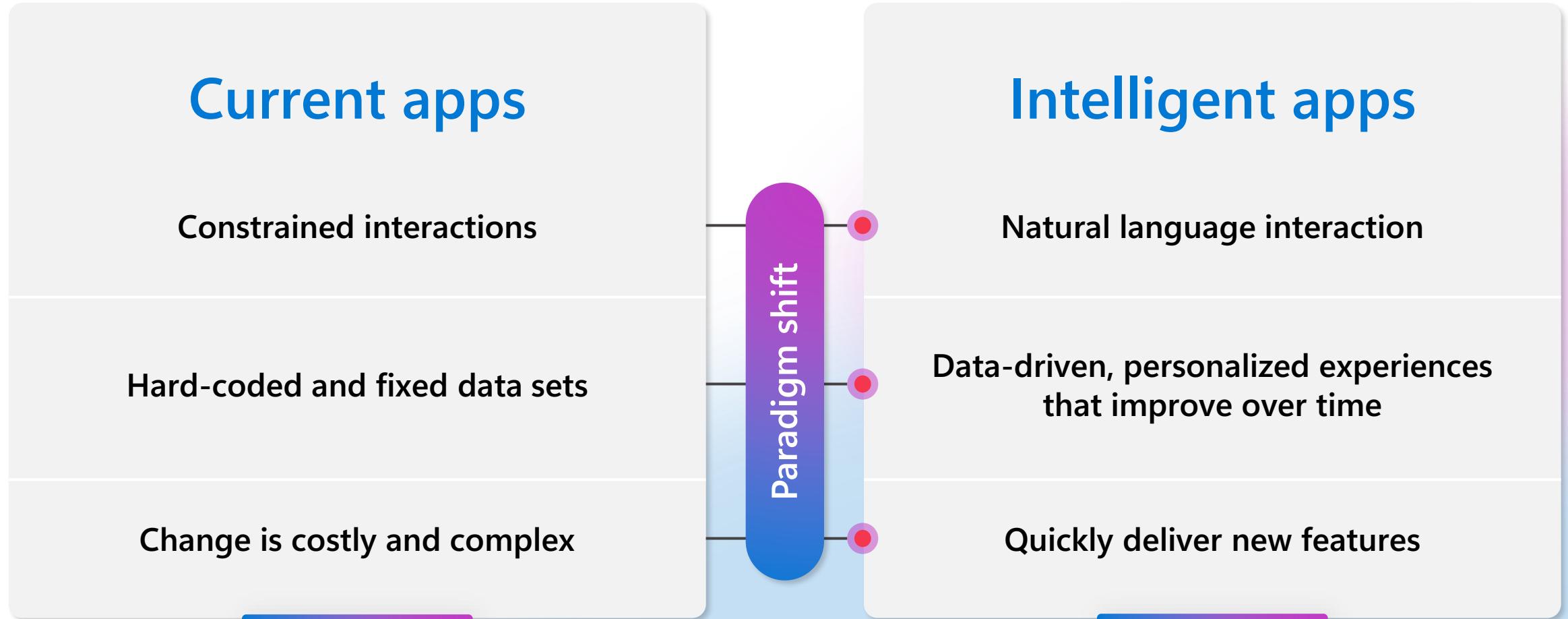
## Operationalization

Concerns around privacy, security, and grounding. Developers lack the experience and tools to evaluate, improve, and validate the solutions for their Proof of Concepts, and to scale and operate in production

## Experience

Less than 30% of surveyed executives say their organizations have the in-house expertise needed to adopt and scale generative AI.<sup>1</sup>

# Add generative AI to your apps to gain true intelligence



# Developers want...

To learn how to select the right tools and models for their use case

Any easy way to switch between UI and code as they build

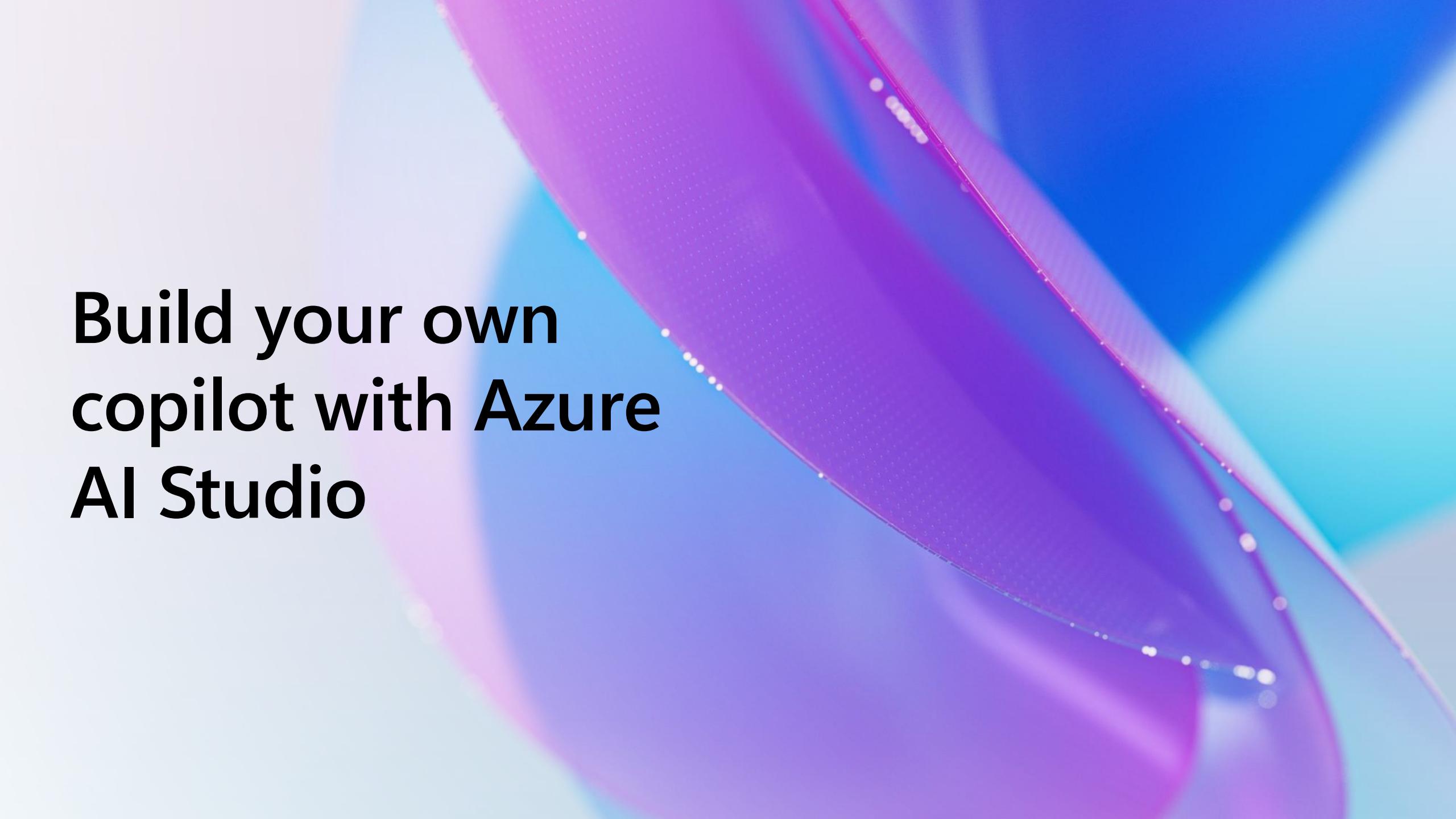
An orchestration framework to handle the complex mapping of functions and code

Helpers, building blocks, and app templates

To evaluate the model, prompt, and overall application

To scale their PoCs into production

To continuously monitor AI app performance



Build your own  
copilot with Azure  
AI Studio

# Azure AI Studio

## Generally available

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



The screenshot displays the Azure AI Studio interface. At the top, there's a navigation bar with links for Home, AI services, Model catalog, AI language, and AI forums. Below the navigation is a banner titled "Innovate with AI" with sub-sections for "Explore cutting-edge models", "Start using Azure OpenAI in AI Studio", and "Work in code with the SDK". A sidebar on the left lists "Jump to top tasks and tools" such as "Get up and running in AI Studio in just five minutes", "Experiment with prompts in the playground", and "Deploy large language models (LLMs)". The main content area includes sections for "Explore cutting-edge models" (listing Phi-3-midi-4x-instruct, Meta-Llama-3-70B-instruct, Cohere-commercial-instruct, and Llama-large), "Infuse your solutions with AI capabilities" (Speech, Language, Vision & Document, and Content safety), "Dive into documentation" (Deployment, Evaluation and monitoring, Pre-training, Deployments, and Manage flow), and "Take in a video tutorial" (with a link to a YouTube video). The bottom right shows a tablet displaying the AI Studio interface.

Generally available



# Azure AI Studio

A comprehensive platform to develop and deploy custom copilots

API and model  
choice

Complete AI  
toolchain

Responsible AI  
tools and practices

Enterprise-grade  
production at scale

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

# Get to know Azure AI

## Azure AI Infrastructure

State-of-the-art silicon and systems for AI workloads

High-Bandwidth Networking

Microfluidic Cooling

Azure Maia Silicon

## Azure AI Studio

One place for building and deploying AI solutions

API & Model Choice

Complete AI Toolchain

Responsible AI Tools & Practices

Enterprise-grade Production at Scale

## Cutting-Edge Models

Access to the latest foundation and open-source models

Model Catalog

Models As a Service

GPT Model Family

Open-Source Models

Small Language Models

## Azure AI Services

Pre-trained, turnkey solutions for intelligent applications

Azure OpenAI Service

Azure AI Search

Azure AI Speech

Azure AI Vision

Azure AI Content Safety

Azure AI Document Intelligence

Azure AI Language

Azure AI Translator

## Azure Machine Learning

Full-lifecycle tools for designing and managing responsible AI models

Prompt Flow Orchestration

Responsible Model Design

Model Fine-Tuning

Model Training

# Top use cases for

## Azure AI Studio



**Build your own copilot**  
Your data. Your apps. Your people

**Enterprise chat**  
Provide multi-modal knowledge mining

**Speech analytics**  
Improve interactions and service

**Content generation**  
Deliver new products and services

**Hyper-personalization**  
Support better sales and marketing

# Accessibility to empower developers of all abilities to thrive in the age of AI



Azure AI Studio was designed and built with feedback from developers with disabilities

A screenshot of the Azure AI Studio interface. The top navigation bar includes 'Azure AI Studio', 'Explore', 'Build', 'Manage', and 'Sign in'. Below the navigation is a banner with the text 'Create innovative AI solutions' and 'Build, evaluate, and deploy your AI solutions all within AI Studio'. There are three main sections: 'Get started with Azure AI Studio' (with a 'Discover the studio' button), 'Explore cutting-edge models' (listing 'mistral-large', 'llama-2', 'gpt-4-32k', and 'Ded-iDeLLM-7B-Instruct'), and 'Azure AI dev tools' (with a 'View available SDKs and documentation' button). The central area has sections for 'What's new and notable' (mentioning 'Antonius API available on Azure OpenAI service' and 'Model-7B-v1.1 Large Language Model (LLM) is a pretrained generative text model with 7 billion parameters'), 'Explore cutting-edge models' (with cards for each model), 'Learn about the features' (with cards for 'Copilot', 'Preferences and monitoring', 'Compose and invoke', and 'Compose code solutions'), and 'Watch a video' (with a thumbnail for a 'Deep dive demo' video).

# API and Model Choice





# API and Model Choice

Identify the best AI services and models for your use case

## Build intelligent apps with industry-leading APIs\*

Develop multimodal, multi-lingual AI with out-of-the-box and customizable APIs and models.

## Discover models for your use case

Choose from the latest cutting-edge foundation and open models. Benchmark models by performance and key parameters and deploy what's right for your use case.

## Deploy serverless models

Get started quickly with ready-to-use APIs, without the need to provision GPUs.

# Build 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 (Whisper)
- 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

- Customized translation for industry-specific terminology
- Seamless translation of technical documents and manuals
- Supports multilingual comprehension for diverse audiences

### Azure AI Translator

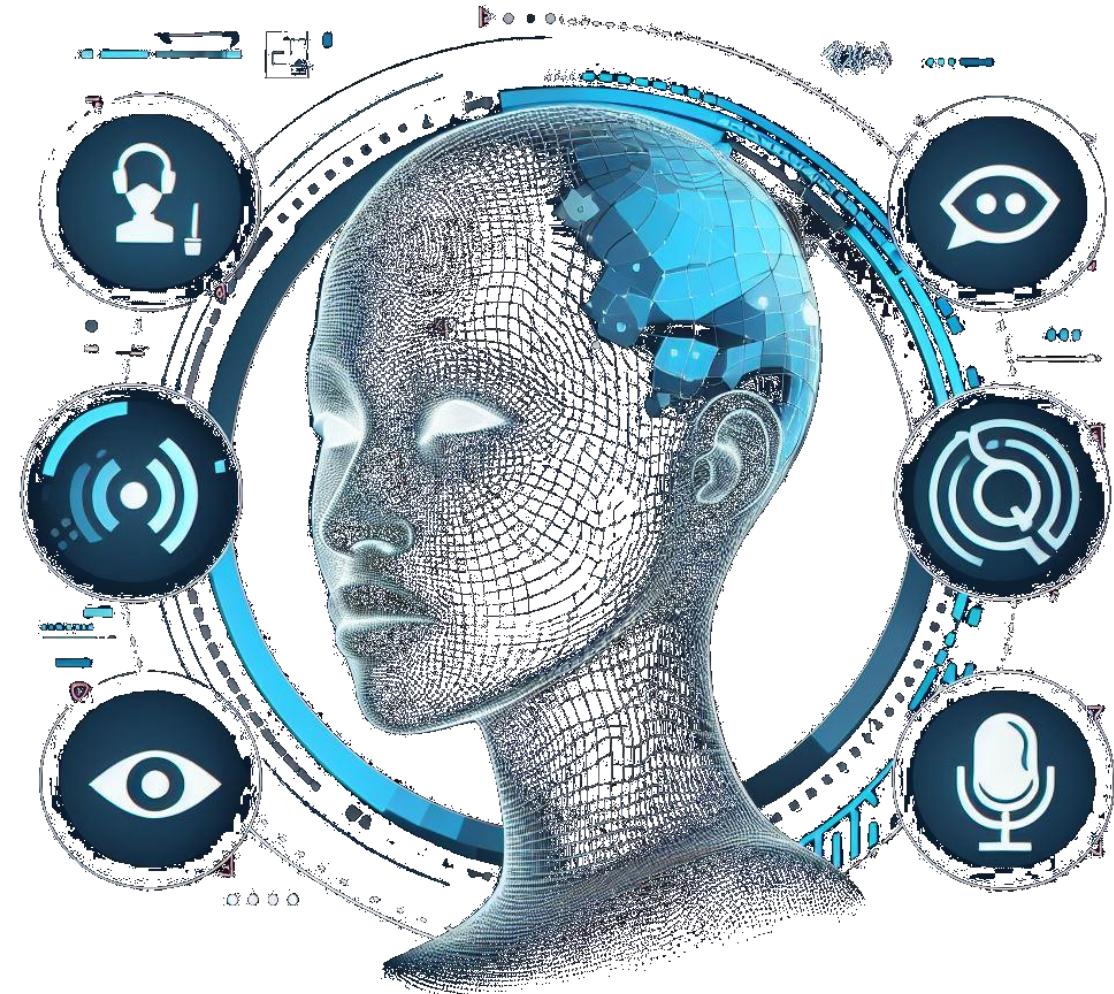
- Multilingual text and speech translation
- Synchronous and asynchronous translation request support
- Utilizes machine learning for highly-accurate translations

# Multimodality is here

Language

Speech

Vision

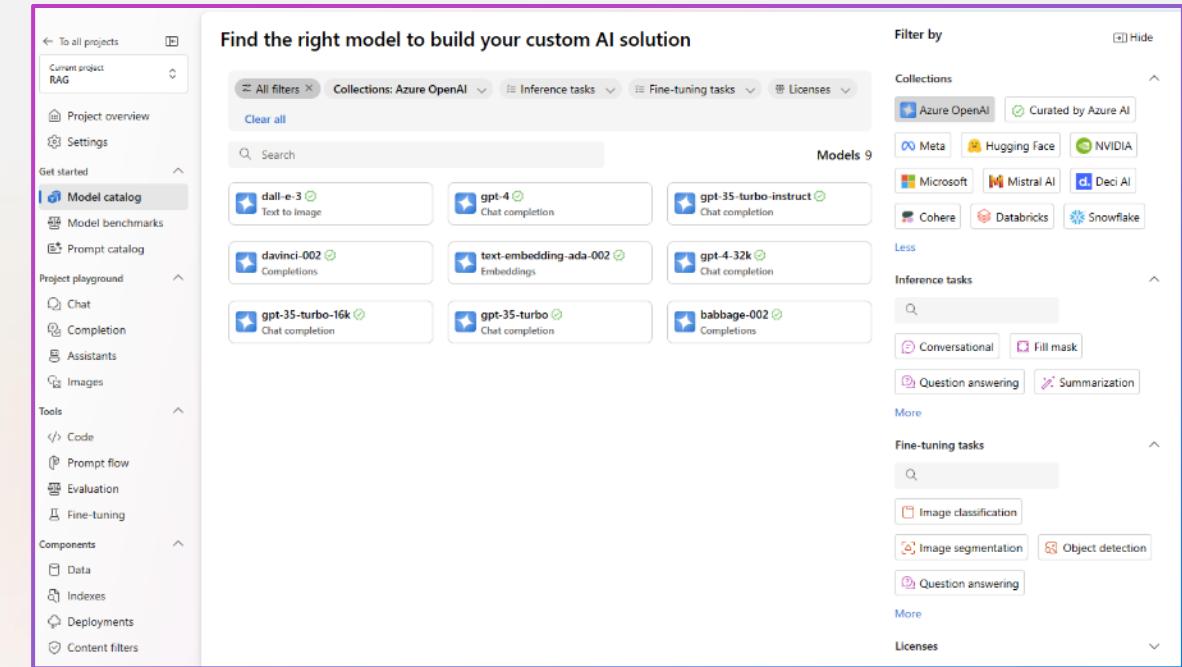


# Azure OpenAI Service (AOAI)

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.

- Access advanced AI models like GPT for diverse applications and tasks
- Generate high-quality imagery with DALL-E 3 through textual descriptions and text-prompt
- Transcribe and translate audio with Whisper, converting 30-second segments into text across 57 languages
- Integrate seamlessly with Azure AI services, enhancing AI capabilities and workflows within existing applications
- Customize and fine-tune AI models to suit specific use cases and requirements effectively
- Ensure compliance and security standards, adhering to responsible AI practices and regulations for trustworthy deployments

Build intelligent apps with industry leading AI



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.

# Create your branded AI voice



Adapt across languages  
and speaking styles

Deploy one-of-a-kind  
to text speech solutions

**Welcome to the custom neural voice**

Custom neural voice (CNV) lets you create a natural-sounding synthetic voice that is trained on human voice recordings. Your custom voice can adapt across languages and speaking styles, and is perfect for adding a one-of-a-kind voice to your Text-to-Speech solutions.

Learn more about custom neural voice. [View documentation](#) [Use REST API](#)

**See and hear how it works**

Voice recordings and transcripts → Train by custom neural voice → Synthetic voice for your brand

[Hear how custom neural voice works](#)

[View the script](#)

**Different options for creating a custom neural voice**

With custom neural voice (CNV), you can create two types of voices, *Lite* and *Pro*. The following table summarizes key differences between the CNV lite and CNV pro voice types.

| Voice type                               | <i>Lite</i> <a href="#">PREVIEW</a>  | <i>Pro</i>  |
|--|--|---|
| Best for                                 | Create a synthetic voice of your own in just under an hour; ideal for testing and evaluation | Design and create a best-in-class synthetic voice for your brand based on professionally recorded samples; ideal for real world scenarios |
| Voice quality                            | Moderate quality   | Highly natural-sounding<br>Resembles the voice actors' accent and intonation  |
| Voice samples<br>(Default neutral style) | Lite, Male, English (UK)<br>Trained with 40 voice samples recorded online                    | Pro, Male, English (UK)<br>Trained with 300 professional studio recorded samples  |
|  | <a href="#">Original voice recording</a>   | <a href="#">Original voice recording</a>  |
|  | <a href="#">Train with CNV pro</a>   | <a href="#">Train with CNV pro</a>  |

Image may not reflect actual user interface.

**On this page**

- Introduction
- See how it works
- Voice types
- Responsible use
- Key steps
- How to create a voice
- Featured customers

# Azure AI Vision

GPT-4 TURBO ENHANCEMENTS

Video Prompting

Object Grounding

Dense Text Accuracy

HIGH QUALITY PRE-TRAINED MODELS

OCR

FACE

IMAGE ANALYSIS

SPATIAL ANALYSIS

Printed Text Extraction

Detection

Tagging & Captioning

Person Tracking

Handwritten Extraction

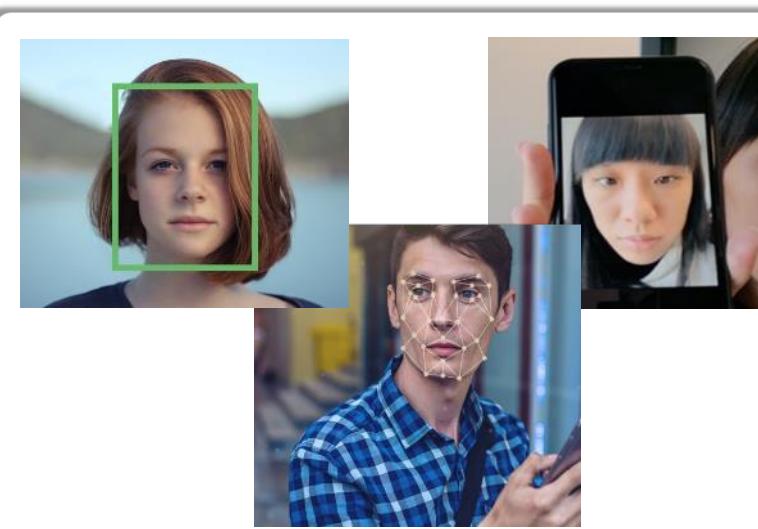
Recognition & Liveness

Content Moderation

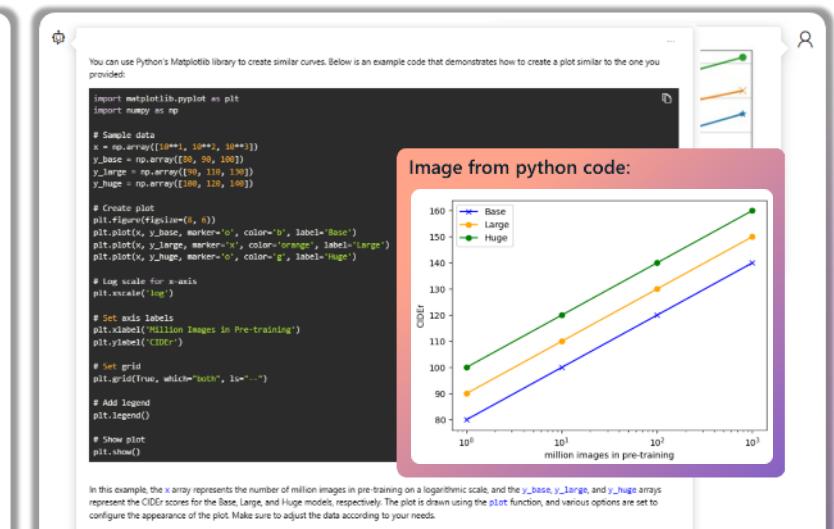
Object Tracking



Analyze objects in images and videos and provide detailed descriptions



Detect and recognize facial liveness



Convert images into code

# What can Speech Analytics do for you?

## Multiple Use Cases

Flexible service catering multiple use cases, e.g.:

- Summarize conversations or videos and generate additional meta-data
- Generate Key Insights and extract meaningful metrics
- Generate additional results like PII redacted recordings, video captions, dubbed videos and more

## Integrated Service

Significant reduction of effort to implement and maintain based on an Azure AI project

Pre-built integration of AI capabilities optimized for processing of audio content

## From PoC to production

Easy to setup and get started with analytics scenario templates for common use cases

Scales effortlessly from proof-of-concept to productive workloads

## Flexibility

Start with pre-built analytics allowing you to easily make changes to cater your business needs

Simple experience for basic changes (e.g., parameters, prompts)

Allow rich customization for advanced needs (e.g., flows)

# Azure AI Model Breadth

Offering a wide collection of foundation and open models

## A Azure OpenAI Service

- GPT-4-Turbo
- GPT-4
- GPT-4V
- Text-embedding-ada-002
- GPT-3.5-Turbo

## ∞ Meta

- Llama-2-70b/70b-chat
- Llama-2-13b/13b-chat
- Llama-2-7b/7b-chat
- CodeLlama

## Mistral AI

- Mistral Large
- Mistral 7b
- Mixtal 7b\*8 – Mixture of Experts

## cohere

- Cohere R+\*
- Cohere R\*
- Embed v3-Multilingual\*
- Embed v3-English\*

\*Available as pay-as-you-go

## 🤗 Hugging Face

- Falcon/TII
- Stable Diffusion/Stability AI
- Dolly/Databricks
- CLIP/OpenAI

## 👁️ nvidia

- Nemotron-3-8B-4k
- Nemotron-3-8B-Chat-SFT/ RLHF/ SteerLM
- Nemotron-3-8B-QA

## databricks

- Databricks/dbrx-base
- Databricks/dbrx-instruct

## Small language models

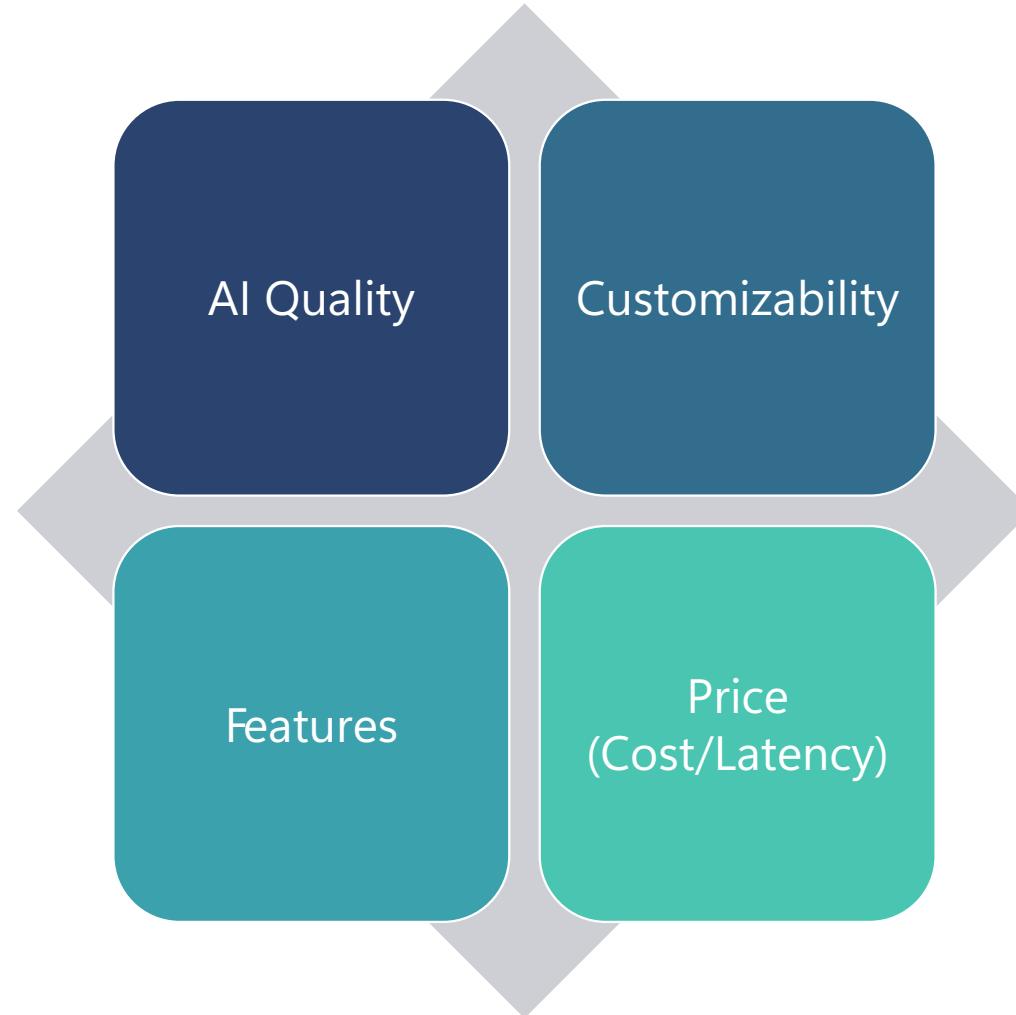
### Phi

- Phi-1
- Phi-1.5
- Phi-2

### Orca

- Orca 1
- Orca 2

# Considerations while selecting a model



## Model Comparison

(Cost/Latency)

AI Quality

Features

Customizability



—●— Gemma-7b

—●— Mistral-7b

—●— Mixral-8x7b

—●— Llama-3-8B-In

—●— GPT-4

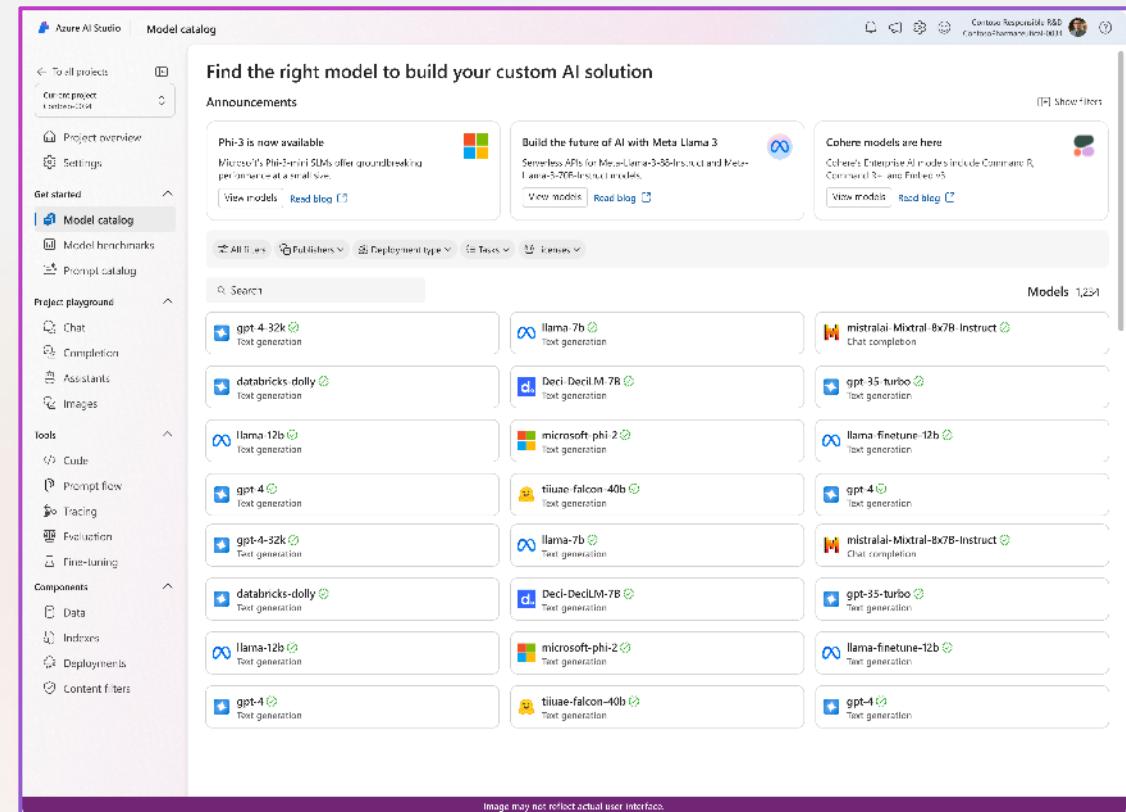
—●— Phi-3-Small-8K-In

# Model catalog

API and Model Choice

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

- Models from OpenAI, Hugging Face, and Meta, including Llama-2-7b, GPT-4V, and GPT-35-turbo
- Quickly try out any pre-trained model using the Sample Inference widget on the model card, providing your own sample input to test the result
- Filter model catalog by collection, model name, or task to find the model that best suits your needs
- Compare models by task using open-source datasets
- Leverage ready-to-use, curated fine-tuning pipelines, code-based inferencing, and evaluation of the mode



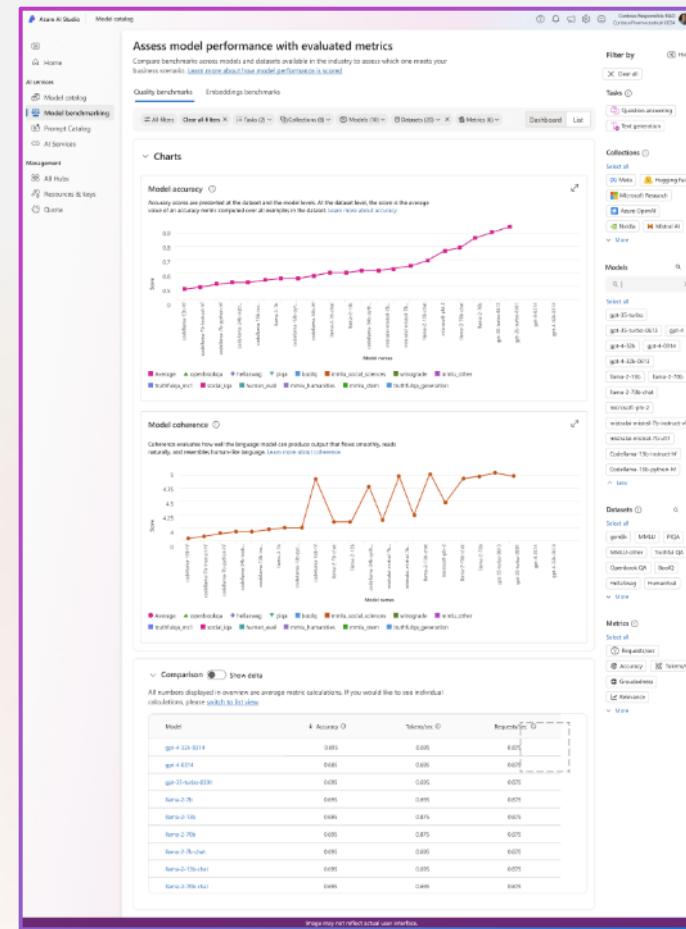
Easily find the foundation model that best suits your needs while improving business outcomes, reducing operational expenses, and enhancing competitive advantage.

# Model benchmarking

API and Model Choice

Model Benchmarking enables you to review and compare the performance of various AI models, simplifying the selection process and allowing users to make confident choices with their modeling needs.

- Gain quality metrics for Azure OpenAI Service, Llama 2 family, Code Llama, and Mistral models
- Access pre-built metrics and benchmark comparison models within the same build, train, deploy environment
- Compute accuracy scores at both the task (dataset) and model levels by utilizing public datasets, yielding a model score for each dataset
- Compare scores of multiple models across datasets and tasks
- Benchmark results originate from public datasets; Azure AI evaluation pipelines download data from original sources, extract prompts from each row, generate model responses, and then compute relevant accuracy metrics



Enable the comparison of models based on accuracy and empower users to make data-driven decisions, ensuring their AI solutions are optimized for the best performance.

# Models as a Service (MaaS)

## *Offering within the model catalog*

API and Model Choice

Models as a Service (MaaS) lets you fine-tune models without provisioning compute, 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 'Llama-2-70b' model page in the Azure ML Model Registry. At the top, there are tabs for 'Overview', 'Versions', and 'Artifacts'. Below the tabs are buttons for 'Task: Text generation', 'Fine-tuning task: text-classification', 'Fine-tuning task: text-generation', 'Languages: EN', and 'License: custom'. There are also buttons for 'Refresh', 'Fine-tune', 'Deploy', and 'View license'. The main content area has sections for 'Description', 'Model Details', and 'Pay-as-you-go Inference APIs and Hosted Fine-tuning'. The 'Model Details' section includes a note about the Meta license and a detailed description of the Llama 2 family of models. The 'Pay-as-you-go Inference APIs and Hosted Fine-tuning' section provides pricing information for inference and fine-tuning. A table below lists training data, parameters, content length, GQA, tokens, and LR for three variations of the Llama 2 model. At the bottom, there is a 'Deploy or fine-tune model' button and a 'Model ID' field containing the URL 'azurerm://registries/azureml-meta/models/Llama-2-70b/versions/22'.

| Training Data  | Params | Content Length | GQA | Tokens | LR                      |
|--|--------|----------------|-----|--------|-------------------------|
| Llama 2<br>A new mix of publicly available online data | 7B     | 4k             | X   | 2.0T   | 3.0<br>10 <sup>-4</sup> |
| Llama 2<br>A new mix of publicly available online data | 13B    | 4k             | X   | 2.0T   | 3.0<br>10 <sup>-4</sup> |
| Llama 2<br>A new mix of publicly available online data | 70B    | 4k             | ✓   | 2.0T   | 1.5<br>10 <sup>-4</sup> |



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

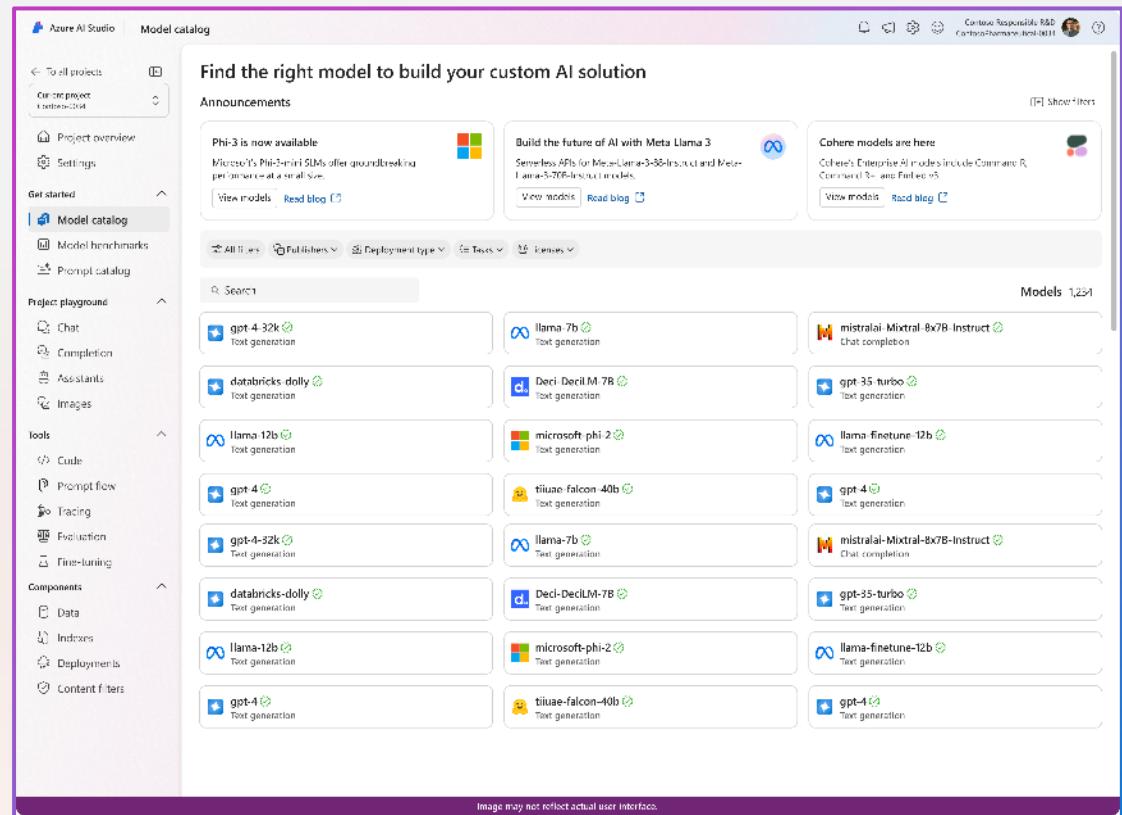
# Models as a Platform (MaaP)

## *Offering within the model catalog*

API and Model Choice

Models as a Platform (MaaP) lets you innovate with models from vetted providers using a generative AI model store to explore large foundation models curated by Microsoft, Hugging Face, Meta, and other open-source contributors.

- Explore thousands of large foundation models (LLMs and SLMs) packaged for out-of-the-box usage and optimized for Azure AI Studio
- Manage your own compute resources with a self-managed GPU infrastructure
- Enhance model performance with highly customized fine-tuning
- Strengthen data security and privacy by controlling the network environment with managed virtual network scenarios
- Access shared compute resources and temporary endpoints for testing, available for seven days and intended for use in testing scenarios



MaaP offers scalable, self-managed hosting for greater control and customization, giving developers flexibly to choose from the most comprehensive selection of open-source generative AI models.

# **DEMO – Session 1 Lab**

# **Session 2 -**

# **Complete AI**

# **Toolchain**



# Complete AI Toolchain

Access collaborative, comprehensive tooling to support the development lifecycle and differentiate your apps

## Ground models on your protected data

Use your protected data to customize models with advanced fine-tuning and for contextually relevant retrieval augmented generation.

## Orchestrate and debug AI workflows

Streamline app development with easy-to-use prompt orchestration, tracing, and debugging via interactive visual and code-first workflows.

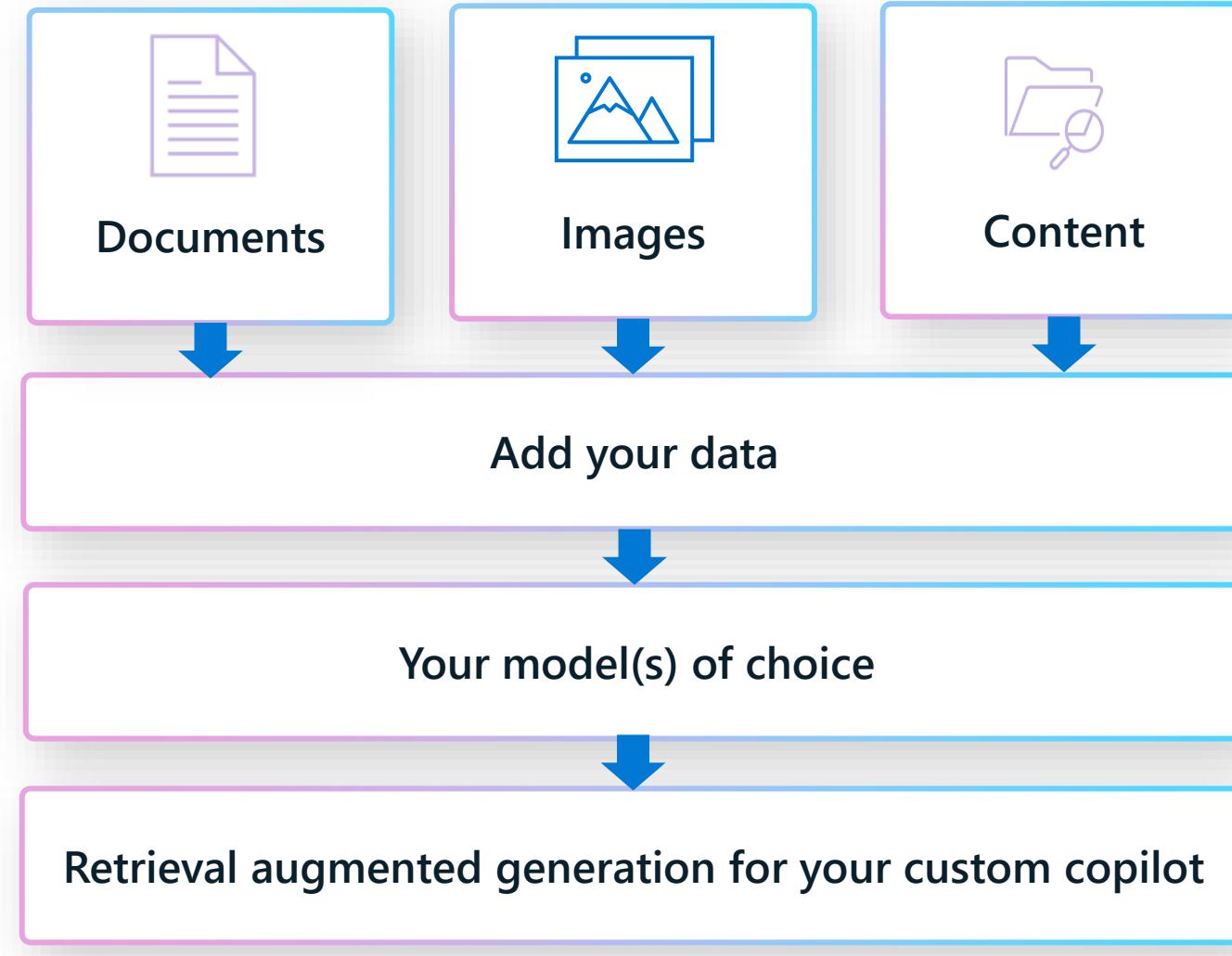
## Streamline model and app evaluations

Systematically measure and improve app performance with manual and automated evaluations.

# Designing your orchestration flow

|               | RAG   | Function/API                                   | Agent   |
|---------------|---|--|---|
| Scenario      | One well-defined task (Q&A)<br>Known outcomes (answers) | Multiple tasks<br>Controlled outcomes (calls)  | Complex task<br>Open outcomes                     |
| example       | <i>looks for answers in product documentation</i>       | <i>query an API to validate account number</i> | <i>chain multiple APIs to solve a problem</i>     |
| Models        | Question answering<br>(understanding/summarization)     | Intent detection<br>and planning               | Multiple  |
| Orchestration | Systematic workflow                                     | Flexible<br>Resilient<br>(code is the trigger) | Dynamic<br>Self-healing<br>(agent is the trigger) |

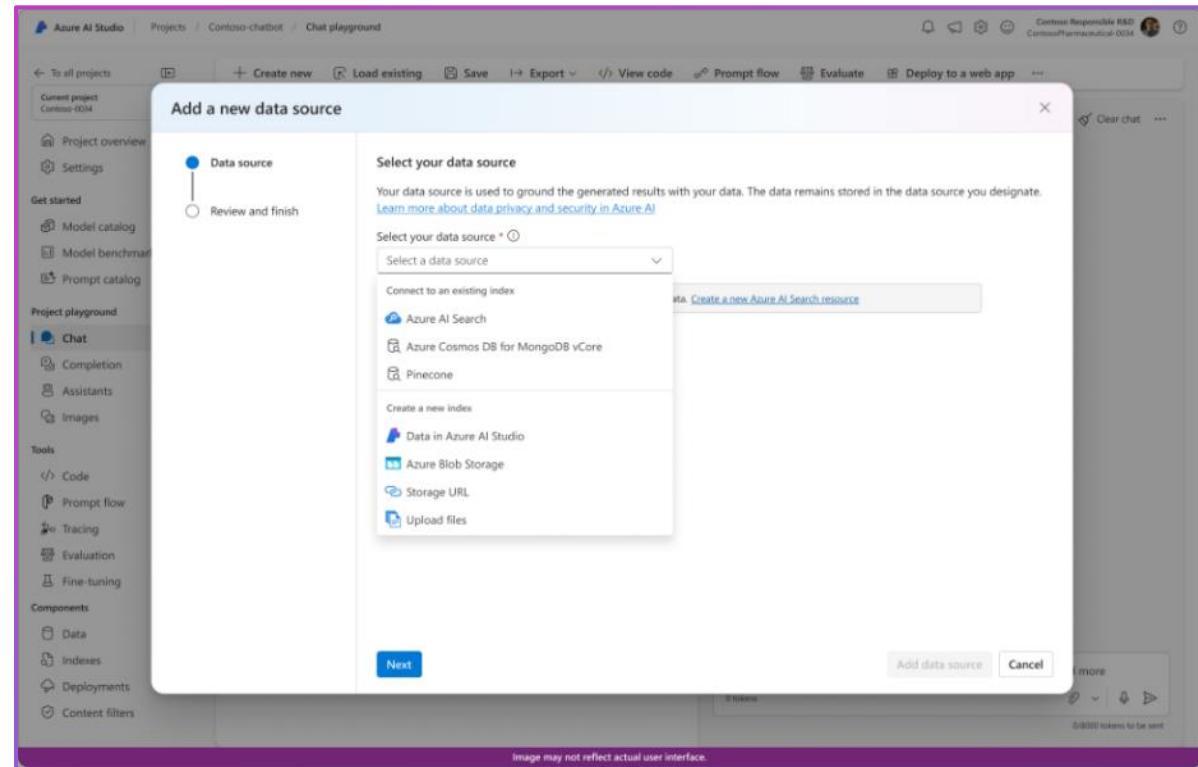
# Retrieval Augmented Generation



# Integrate Structured & Unstructured Data



- Azure AI Search
- Blob storage
- Local files/folders
- Storage URLs including OneLake in Microsoft Fabric
  - Azure Databricks
  - S3 buckets via Amazon S3 Shortcuts



# Azure AI Search

Azure AI Search is an enterprise-ready retrieval system with a robust set of advanced search technology, enabling superior retrieval augmented generation (RAG) so your app delivers the best experience for every user, at any scale.

- Ingest and search across all your data, no matter the type, volume or source
- Streamlined data preparation and indexing with automatic data ingestion, chunking, extraction and enrichment.
- Feature-rich vector database with support for Exhaustive KNN search and ANN search
- Enable hybrid search (vector + keyword) results using RRF (Reciprocal Rank Fusion) and multi-vector retrieval for multi-modal search.
- Improve relevance immediately with semantic ranker, a transformer-based reranking step that runs on GPUs to deliver speed and quality in results.
- Tailor search experiences by utilizing customizable features such as granular access control and relevance tuning

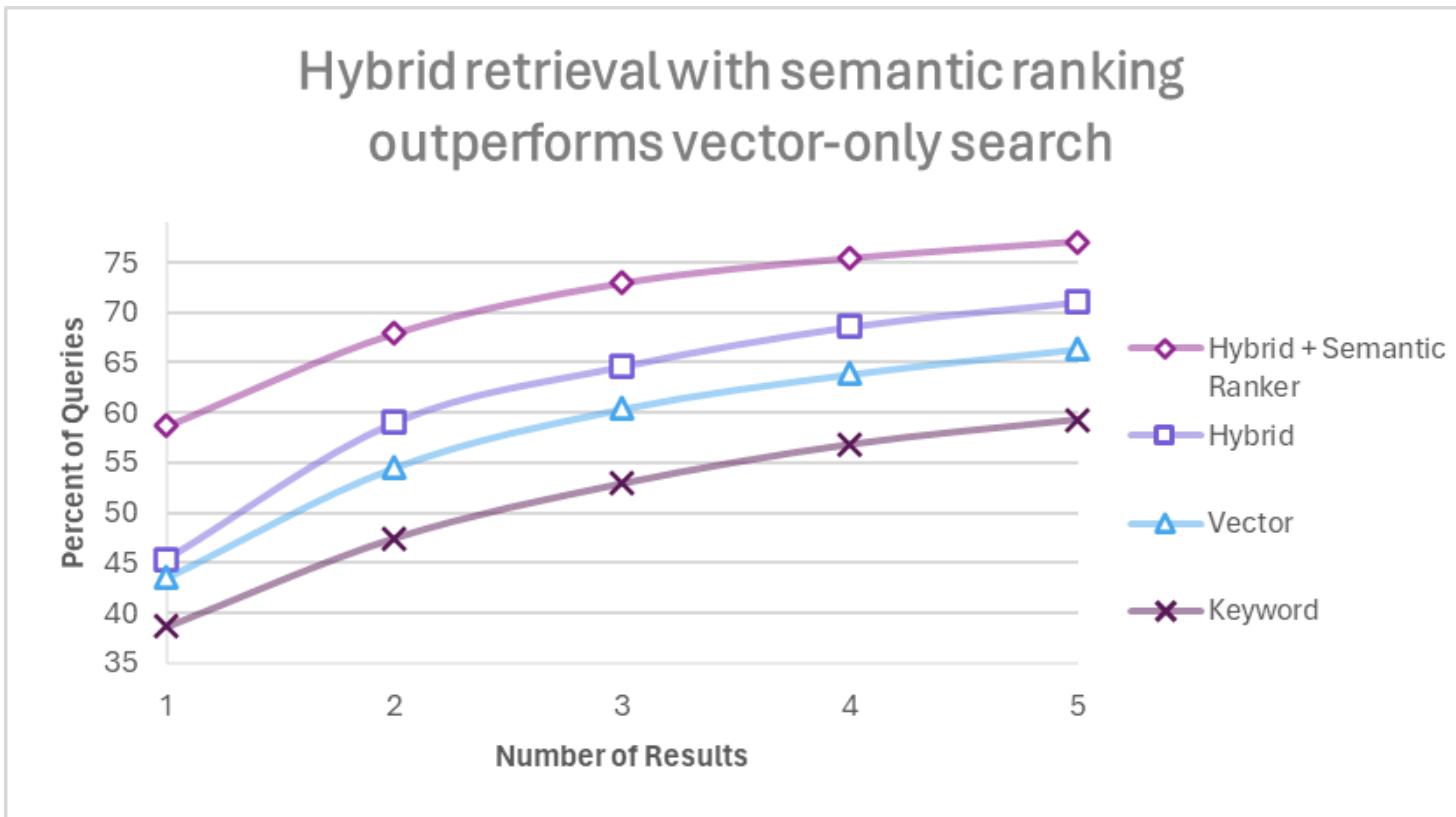
Ground your models on your protected data

The screenshot shows the Azure AI Search configuration interface. On the left, a vertical navigation bar lists steps: Data source (checked), Data field mapping (checked), Data management (checked), and Review and finish (unchecked). To the right, the 'Data management' section is displayed. It includes a sub-section for 'Search type' with a dropdown menu set to 'Hybrid + semantic'. Below it is a dropdown for 'Select an existing semantic search configuration' with 'default' selected. A note states: 'Using semantic search will incur usage to your Azure AI Search account. [View Pricing](#)' and 'Adding vector embeddings will incur usage to your account. [View Pricing](#)'. At the bottom of the interface is a dark purple footer bar with the text 'Image may not reflect actual user interface.'



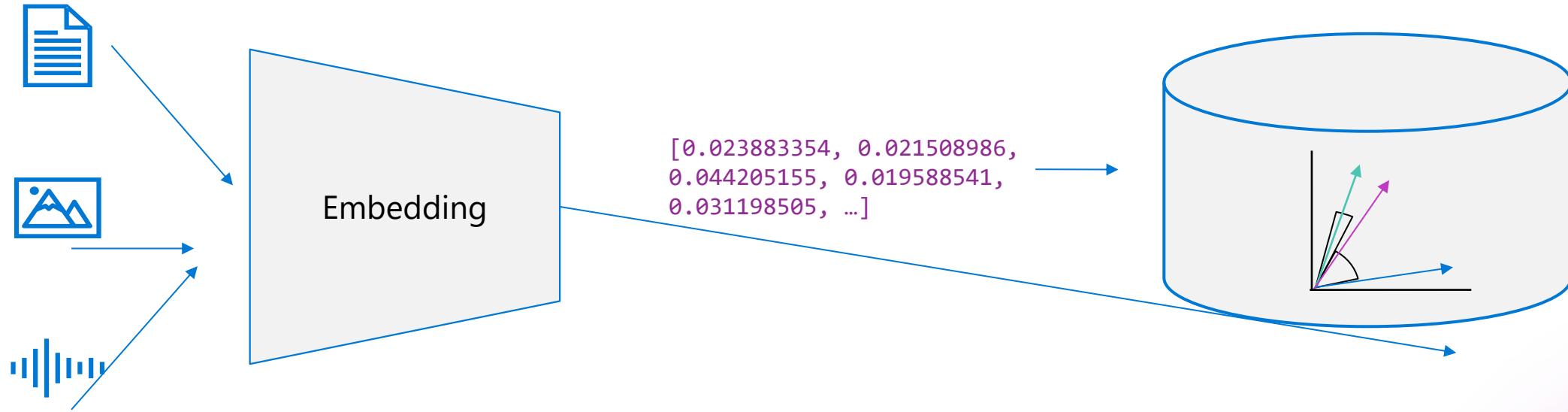
Easily enable search best practices through the UI, backed by extensive product tests and benchmarks, like being able to select hybrid search + reranking in one click.

# Retrieve Using Semantic Similarity and Hybrid Search



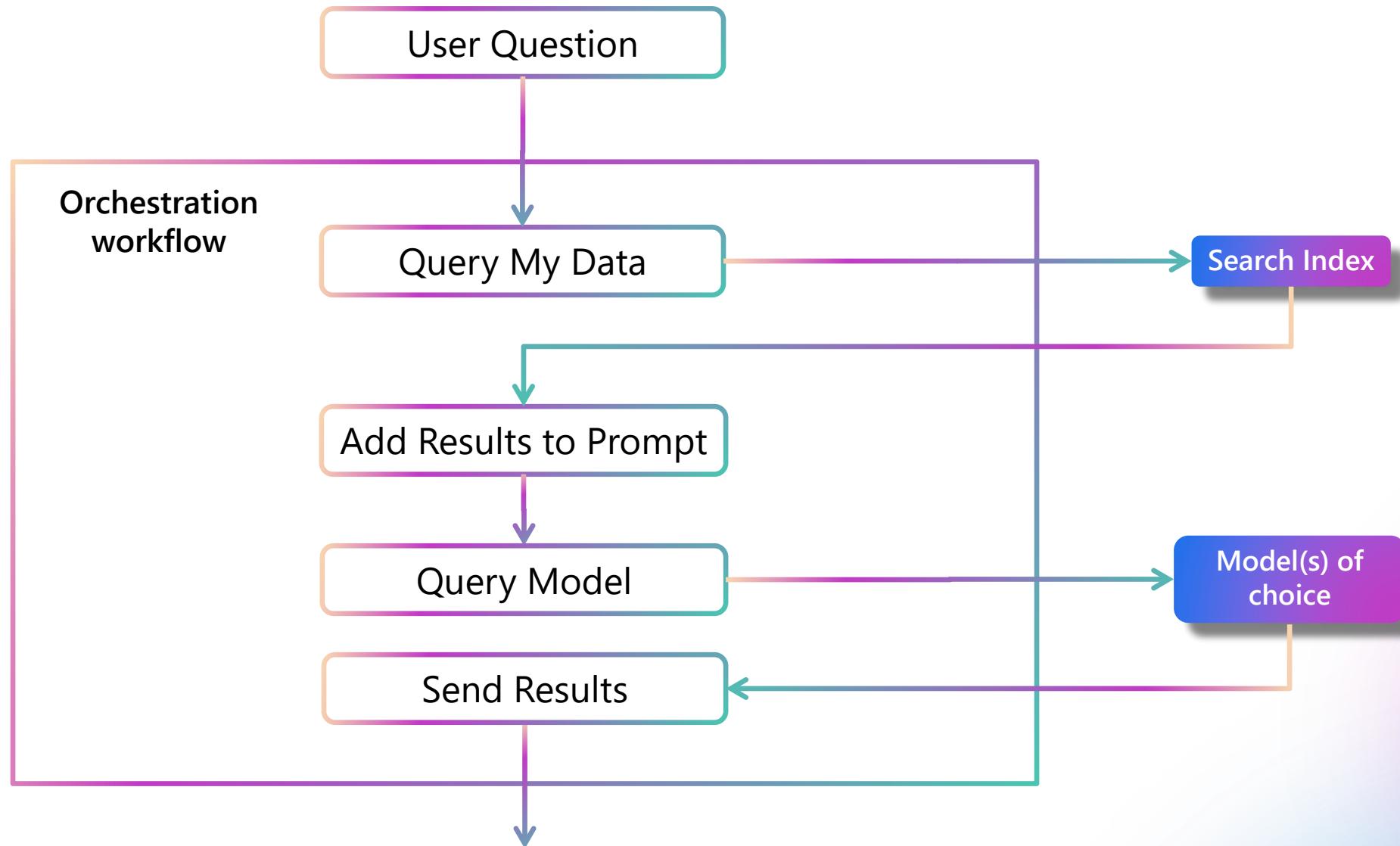
Find the most relevant information in large datasets using hybrid search with keywords, vectors, and semantic ranker

# Retrieve Using Semantic Similarity and Hybrid Search



Find the most relevant information in large datasets using hybrid search with keywords, vectors, and semantic ranker

# Retrieval Augmented Generation Orchestration



## Agents

AI tools that respond to events and can operate independently of user input (after initial and potentially ongoing user authorization) to complete complex objectives that span multiple applications, data sources, and services.

- Build sophisticated stateful agents
- Augmented your copilot to access multiple APIs
- Perform complex computations and data analysis
- Safely act on a user's behalf
- Retrieve useful data in multiple formats

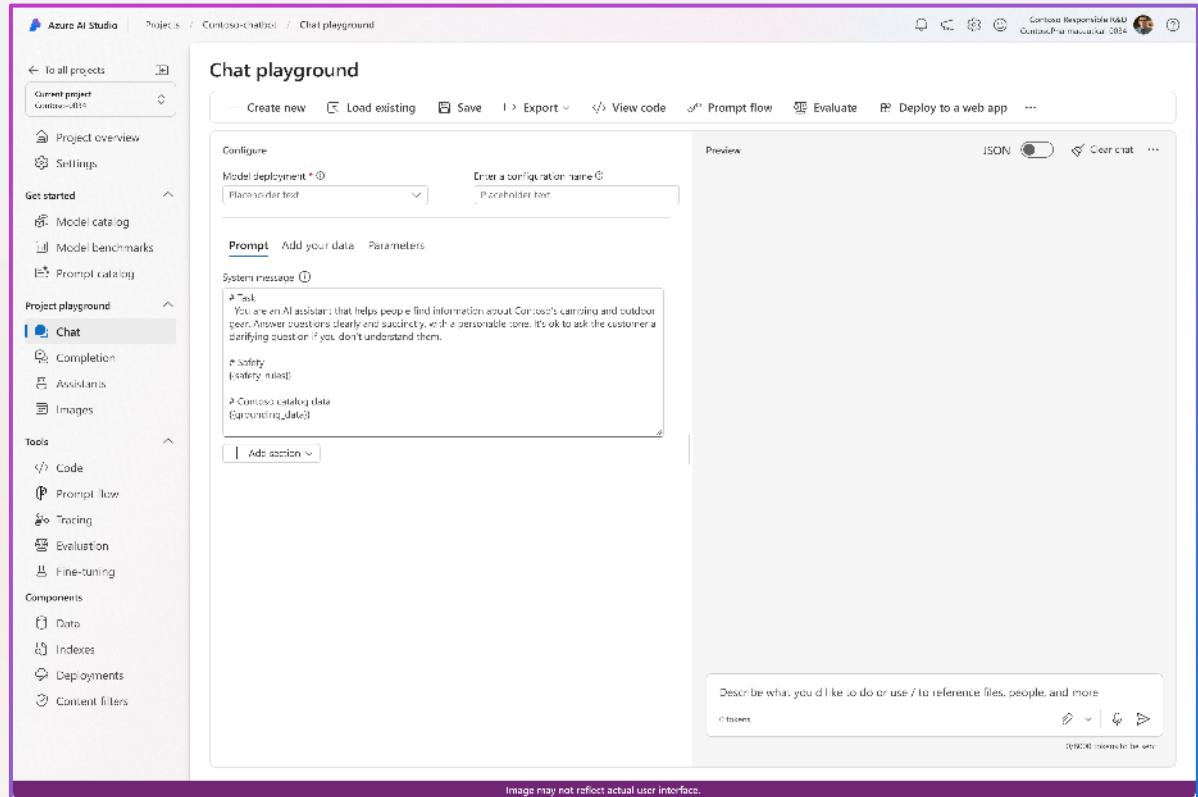
# Chat vs. Agent frameworks

| Chat                     | Agents  |
|--------------------------|---|
| Lightweight and powerful | Stateful (inbuilt conversation state management)  |
| Inherently stateless     | Access persistent threads   |
|                          | Access files in several formats. API handles chunking, embeddings, storage and creation, and implementing vector search |
|                          | Automatic management of the model's context window  |
|                          | Access multiple tools in parallel (up to 128 tools per assistant) including code interpreter                            |
|                          | Build your own function calling   |

# Azure AI Studio playground

The playground in Azure AI Studio offers an environment for developers to experiment with AI models, algorithms, and data.

- Construct system instruction prompts to create your own AI agents and copilots
- Easily connect with indexed data sources to implement retrieval-augmented generation (RAG)
- Leverage prompt samples and templates to eliminate writing long prompts from scratch
- Easily import prompt samples from local or export for local development
- Pair models with Azure AI services APIs to enable richer interactions
- Click "Customize in prompt flow" to further enrich and manage large language model (LLM) workflows



 Give developers the freedom to experiment, innovate, and learn without constraints of production environments.

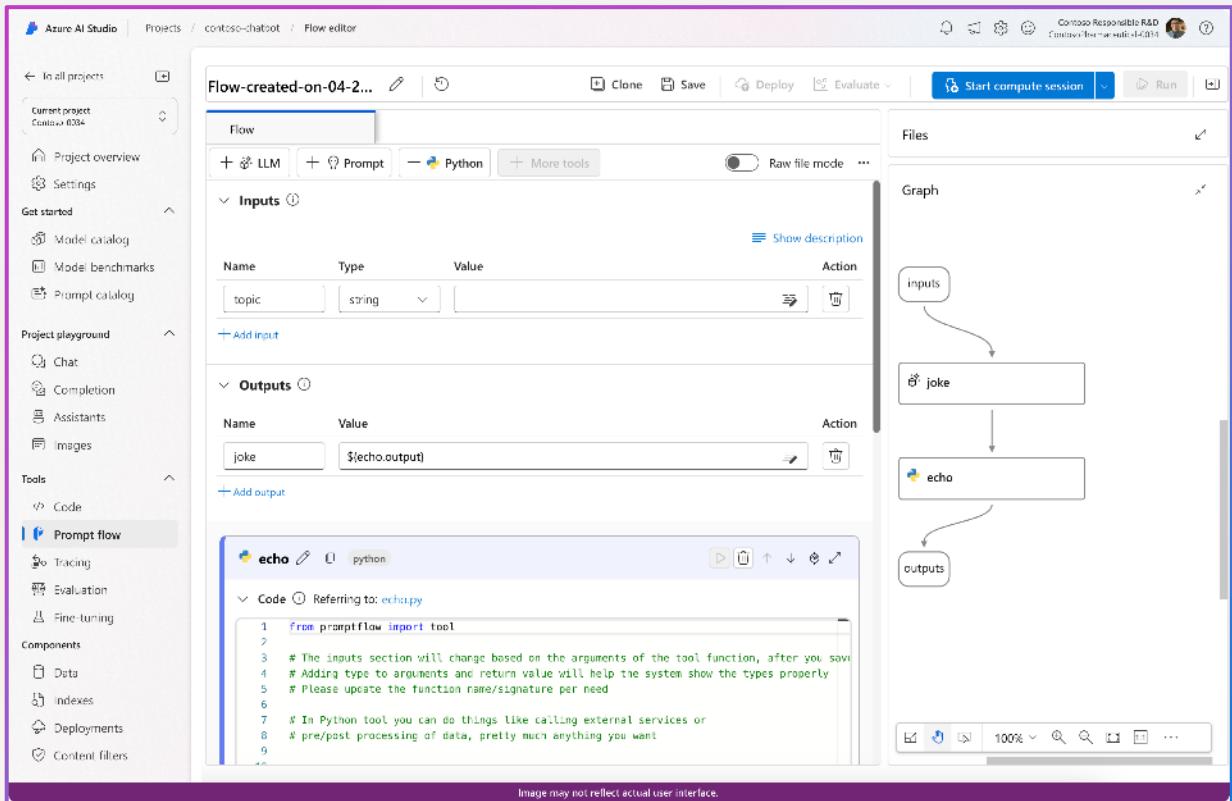
Ground your models on your protected data

# Prompt flow

Prompt flow orchestrates app development by facilitating iterative experimentation and the practical deployment of flows.

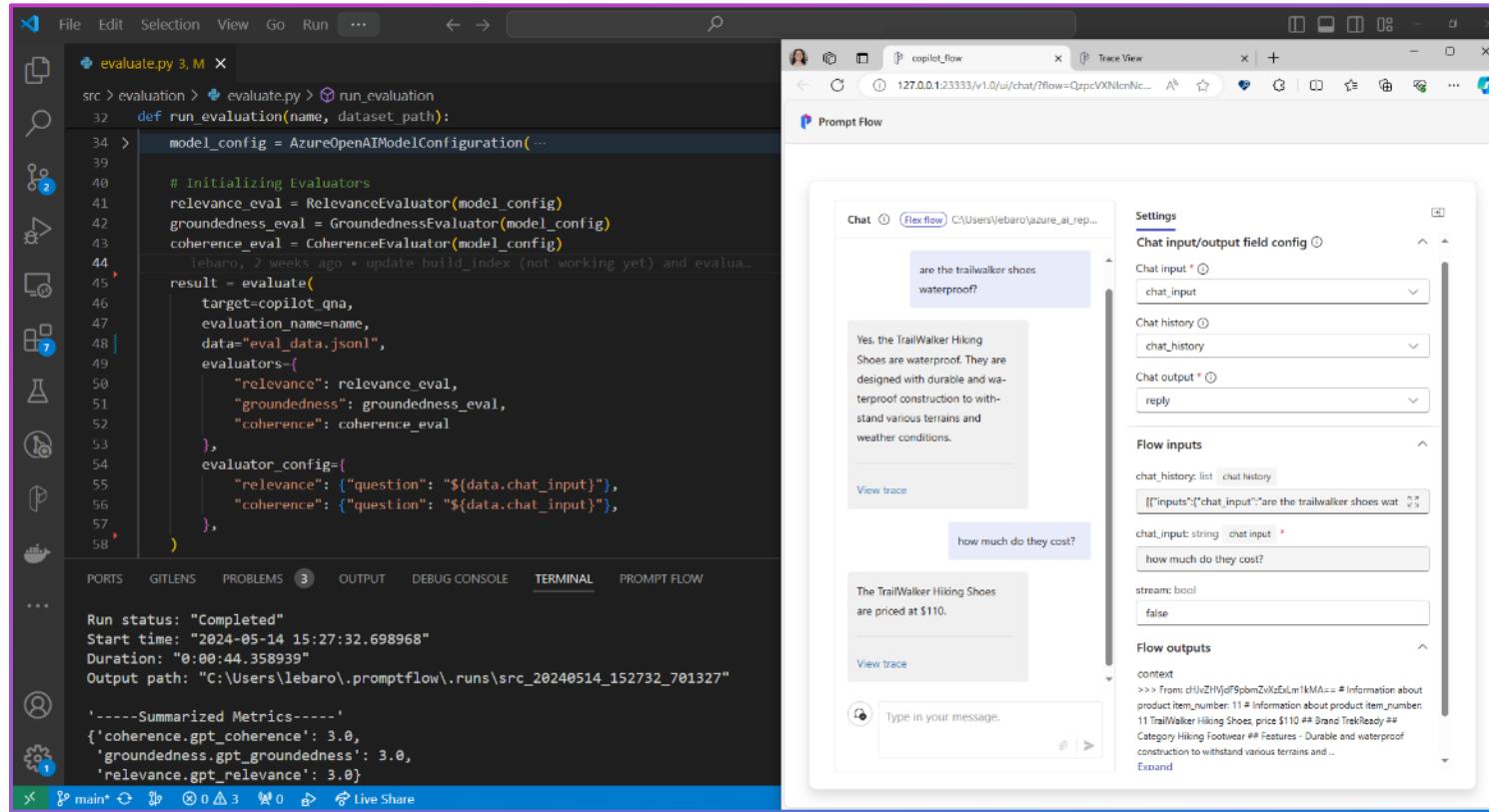
- Develop and manage flows that connect to a variety of models, databases, APIs, prompts, and tools
- Efficiently evaluate prompt flows with large datasets while assessing scenarios for optimized performance. Flow evaluations can be done with different metrics (quality, groundedness, safety, etc.)
- Deploy out-of-the-box prompt flow solutions and evaluations
- Manage prompts and tune with variants and versions to fit your needs
- Compare results across experiments to deliver optimization and uncover efficiencies

Orchestrate and debug AI workflows



Streamline efficiencies with one platform to design, construct, tune, evaluate, test, and deploy LLM workflows with a rich set of pre-built metrics and safety systems.

# Move seamlessly between UI and code



The Azure AI SDK and VS Code extension provides a first-class local developer experience

# Quick start your AI development



Deploy applications from the [AI Project Template Library](#)

Starter repos provide an end-to-end development template for Azure AI Studio with prompt orchestration

GitHub Codespaces support helps developers try sample code without having to set up a local environment

The screenshot shows the homepage of the [awesome-azd](#) website, specifically the **Template Library** section. The page has a light blue and pink gradient header with the title "Template Library" and a subtitle "A community-contributed template gallery built to work with the Azure Developer CLI (azd).". A search bar is present at the top right. On the left, there are filter options for "Community Authored", "Microsoft Authored", "New", and "Popular" templates, and a dropdown for "Language" with options like ".NET/C#", "Java", "JavaScript", "Node.js", "PHP", and "Python". The main content area displays two template cards:

- Azure AI Studio Starter** by Azure Dev: "Bicep template that deploys everything you need to get started with Azure AI Studio. Includes AI Hub with dependent resources, AI..."
- Azure AI Starter** by Azure Dev: "Bicep template that deploys Azure AI services with configured machine learning models."

At the bottom of each card is a command line snippet: "azd init -t azd-aistudio-starter" and "azd init -t azd-ai-starter".

# Tracing/debugging

Effortlessly enable large language model (LLM) application tracing to gain comprehensive insights into the workflow, facilitating quick identification and resolution of performance bottlenecks while streamlining the debugging process.

- Offer out-of-the-box visualization in Azure AI Studio, making traces durable and comparable between different app versions
- Make debugging easier by allowing users to load and run individual test cases against the function or flow
- Enable tracing by adding 2 lines of code with SDK
- Framework agnostic: Prompt flow, Langchain, Semantic Kernel, AutoGen
- Support both local and cloud runs
- Group traces by sessions/applications
- Detail every LLM call including message, prompt, parameters, token, and latency

Orchestrate and debug AI workflows

The screenshot displays the Azure AI Studio interface with several windows open. The main window shows a 'Trace group: myLangChain' with a table of trace details. Below it is a 'Trace details: Main' pane showing a hierarchical tree of traced events, including 'openai.resource.chat.completions.create' and 'openai.chat.completions.create'. A code editor window at the bottom contains Python code for setting up tracing with the OpenAI API. Another window in the background shows a 'Conversation' interface with a message history.

```
# start the trace
from promptflow import start_trace, trace
start_trace(session="my_agent")

# load LLM model config
from autogen import config_list_from_json
config_list = config_list_from_json(env_or_file="OAI_CONFIG_LIST")

# create a teachable agent
from autogen import ConversableAgent
my_agent = ConversableAgent("my_oai_agent", llm_config="temperature")
```

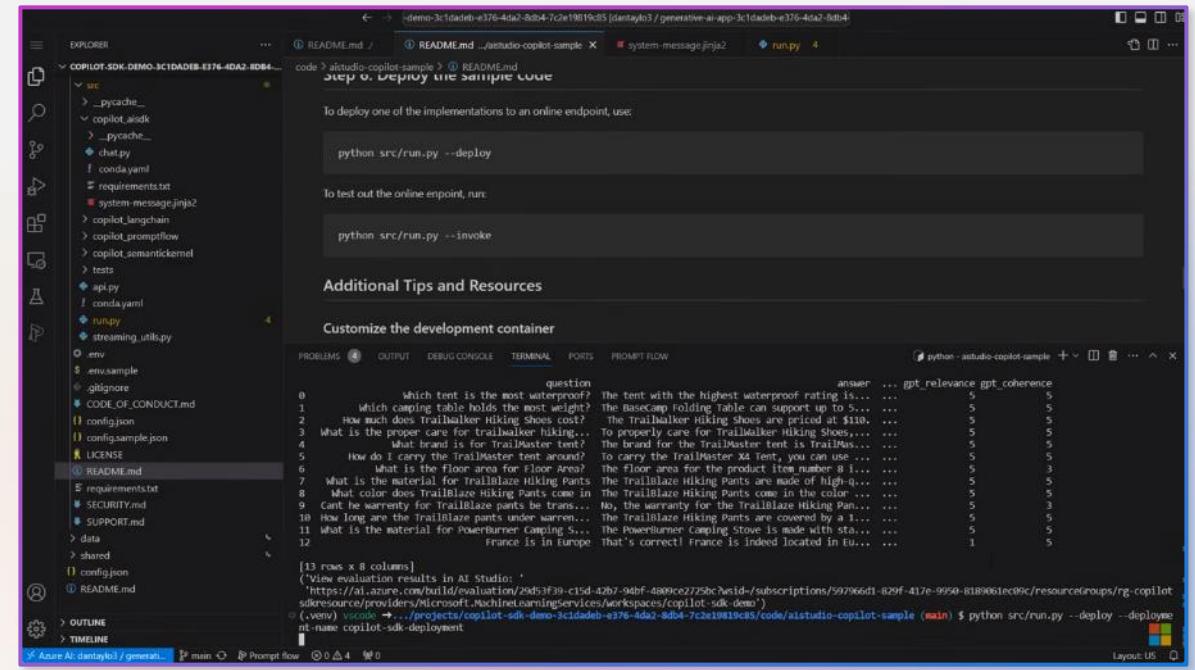
Enable detailed application tracing with just two lines of code, unveiling insights into each step of your application's journey.

# Azure AI SDK

Orchestrate and debug AI workflows

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

- Offer packages for managing Azure AI resources, simplifying tasks like creating, configuring, and monitoring AI services
- Utilize generative packages locally to build an index, run an evaluation, and deploy chat functions and prompt flows
- Interactively execute commands or script larger processes for automation, allowing control-plane and data-plane operations without writing any code
- Streamline workflow management by providing a consistent interface across Azure AI services
- Incremental Azure Building Block app templates beyond SDKs, including templates for copilot scenarios, hosted in web, container, function app, and more

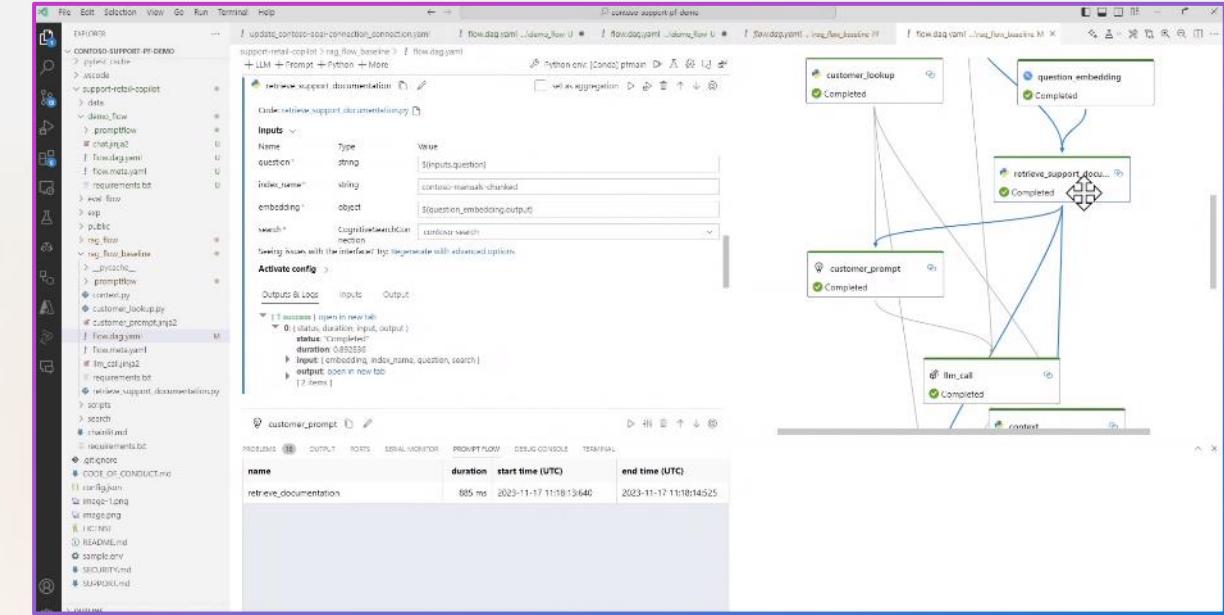


Make life simpler with a single SDK and CLI , reducing complexity of resource creation and management across services.

# Extended tools and frameworks

Prompt flow offers a developer-friendly and easy-to-use code first experiences through SDK, CLI, and VS code extension, allowing seamless integration with open-source frameworks and CI/CD pipelines for automation.

- Use your favorite frameworks and editors that allow you to work in your preferred code environments, whether simplified UI or Visual Studio Code/IDE
- Initially build flows with open-source frameworks like LangChain or Semantic Kernel, and use prompt flow to scale the experiments
- Define flows in YAML format, which can stay aligned with the references source files in a folder structure for flow versioning in code repository
- Easily integrate with existing CI/CD pipelines through GitHub Actions and Azure DevOps for LLM Ops
- Smoothly transition from local to cloud by exporting flow folders to local, or uploading folders to the cloud for further authoring, testing, and deployment



Streamlines the development lifecycle of generative AI app with version-controlling, CI/CD integration, and enhanced collaboration among team members.

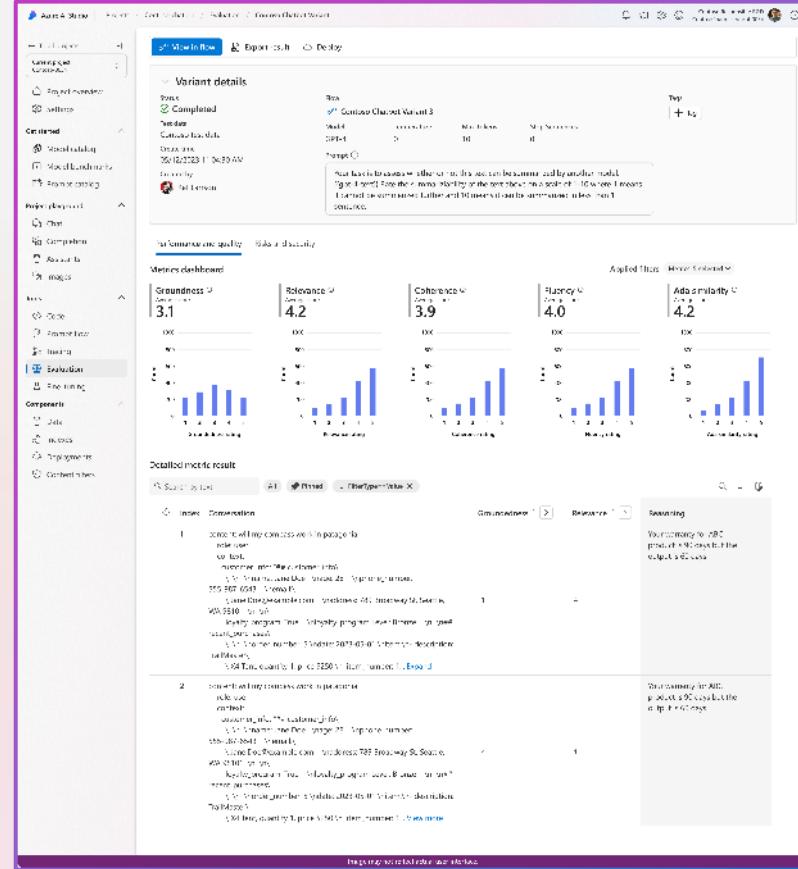
Orchestrate and debug AI workflows

# Automated evaluation

Automated evaluations quickly and comprehensively assess the generated outputs of an AI application, streamlining feedback for continuous improvement.

- Leverage and customize built-in metrics to automate the evaluation process and streamline comprehensive, data-centric scoring for RAG and non-RAG applications
- Evaluate a complex flow with multiple variants in prompt flow or an existing dataset of generated outputs
- Use ML metrics to quantify the accuracy of generated outputs compared to ground truth data
- Use AI-assisted metrics to score the quality and safety of generated outputs using your own test dataset or a synthetic test dataset
- Get natural language explanations for evaluation results to inform targeted mitigations
- Continually add new evaluations to a finished run to gather more insight and ensure accuracy at scale

Streamline model and app evaluations



Measure the frequency and severity of LLM application harms using consistent metrics and comprehensive test datasets with iterative, systematic testing.

# Manual evaluation

Quickly and continuously iterate on your application and track the impact of ongoing changes by manually rating model outputs as you go.

- Manually create or upload your sample test dataset with the option to include expected outputs
- Customize your prompts to target specific aspects of model performance and user interaction you want to improve
- Provide a thumbs up or down rating to score each generated response
- Review model response scores by prompt or in the at-a-glance summaries
- Iterate on your application and re-run evaluations to track the impact of your changes
- Save and compare results to identify the optimal application design for your desired outcomes

The screenshot shows the Azure AI Studio interface with the 'Contoso-crabot' project selected. The left sidebar includes 'Project overview', 'Settings', 'Get started', 'Model catalog', 'Model benchmarks', and 'Prompt catalog'. Under 'Tools', 'Evaluation' is highlighted. The main area displays 'Assistant setup' with a 'Prompt' tab selected, showing a system message about being an AI assistant for Contoso's camping gear. It also shows configurations for a 'Free Trial GPT 3.5 Turbo' model, including 'Max response' (800) and 'Temperature' (0.7). Below this is the 'Manual evaluation result' section, which includes a 'Data rated' summary (0/3 data volume, 30% thumbs up, 70% thumbs down), a table of inputs, expected responses, and outputs, and a note about the speed of light in a vacuum. At the bottom, there's a note about missing data columns for inference.

Based on the results, you may decide to update your system message, model, or model parameters. Then, rerun the dataset or specific prompts that didn't meet your expectations to see the impact of your updates.

# Evaluation datasets

| Dataset           | Description                                    | Metrics / Goals   |
|-------------------|--|---|
| Qualified Answers | Small number of examples obtained from experts | Maximize response quality (groundedness, relevance, coherence, ...)   |
| Synthetic         | Large number of synthetic samples              | Maximize response quality<br>Maximize retrieval metrics (recall, precision)<br>Maximize tool selection metrics (accuracy) |
| Adversarial       | Jailbreak attempts, harmful content            | Minimize unsafe responses   |
| OOD               | Out of domain questions                        | Minimize relevance ;-)  |
| Thumbs down       | Dataset of bad answers                         | Eye-ball, use to discuss  |
| PROD              | Scrubbed questions from opt-in users           | Maximize response quality<br>Ensure production satisfaction   |

# Evaluation best practices



Multiply  
your  
datasets  
and metrics



Evaluate  
systematically  
using LLMOps



Review results  
as a team, with  
multiple points  
of view

# **DEMO – Session 2 Lab**

# **Session 3 -**

# **Responsible AI**

# **Tools & Practices**





# Responsible AI practices

## Design and safeguard applications

### Design apps responsibly

Confidently build apps with technologies, templates, and best practices to help manage risk, improve accuracy, protect privacy, reinforce transparency, and simplify compliance.

### Safeguard with configurable filters and control

Detect and filter harmful content, protect PII, and safeguard applications against prompt attacks.

# Foundation models introduce new harms



Ungrounded outputs & errors



Jailbreaks & prompt injection attacks



Harmful content & code



Copyright infringement



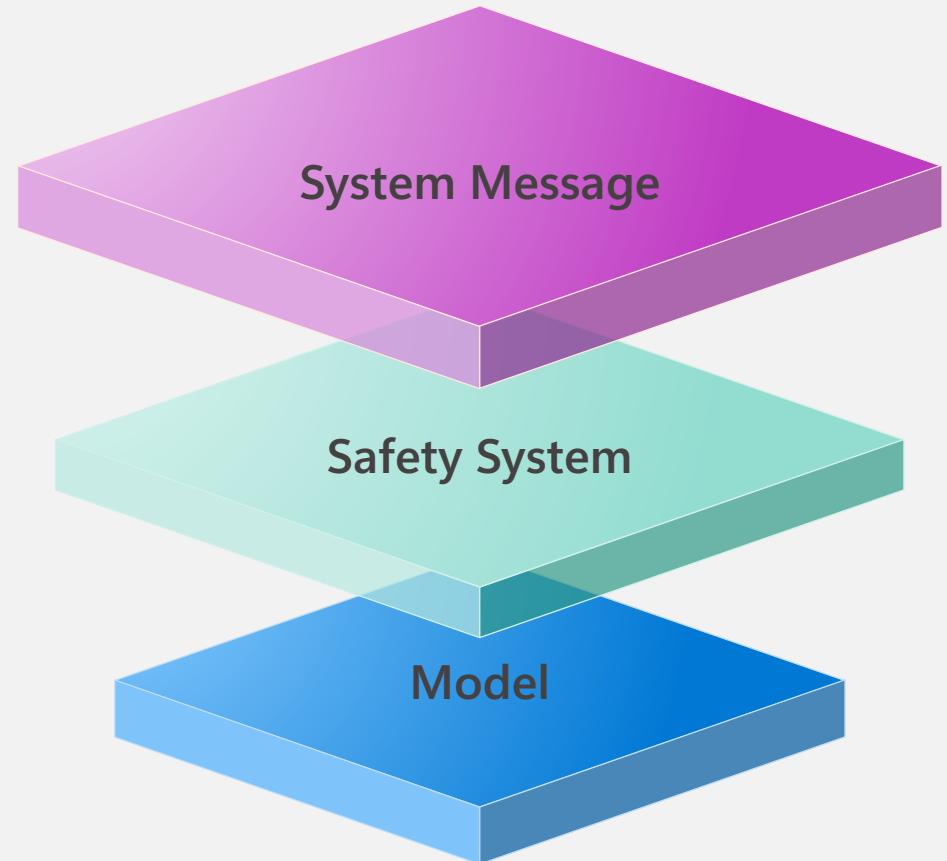
Manipulation and human-like behavior

# Mitigation layers in Azure AI Studio

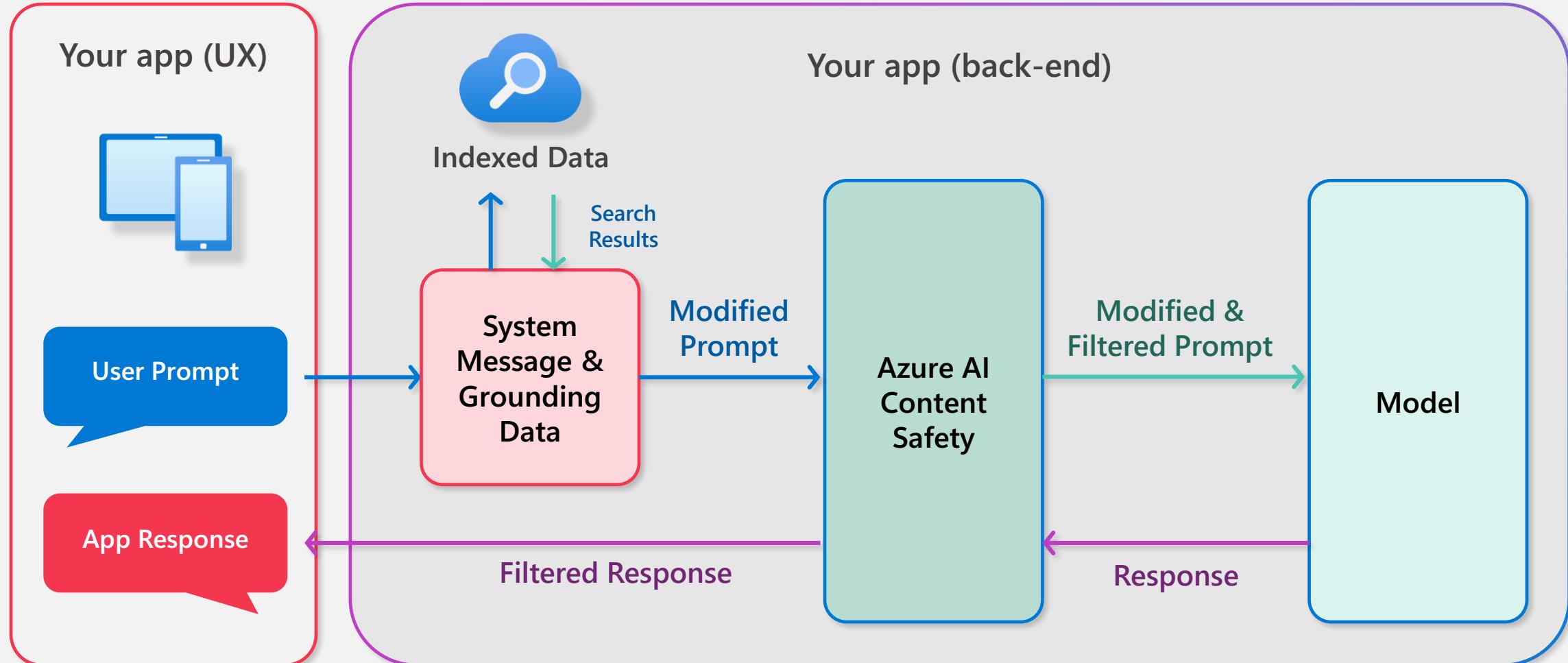
The **system message layer** provides hidden instructions to your model with every user prompt, so you can guide the model's behavior and data retrieval to generate higher quality responses by default.

The **safety system layer** acts as a shield or firewall around your foundation model, by detecting and filtering/blocking risky prompts before they reach your model and risky responses before they reach your end user, for high quality experiences that stay on-brand.

The **model layer** is your foundational model, including any fine-tuning performed by you or the model developer to improve performance and safety.



# How these mitigations happen in real-time



# 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 Studio 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

Design apps responsibly

## Microsoft's Responsible AI principles



Fairness



Reliability & Safety



Privacy & Security



Inclusiveness



Transparency



Accountability



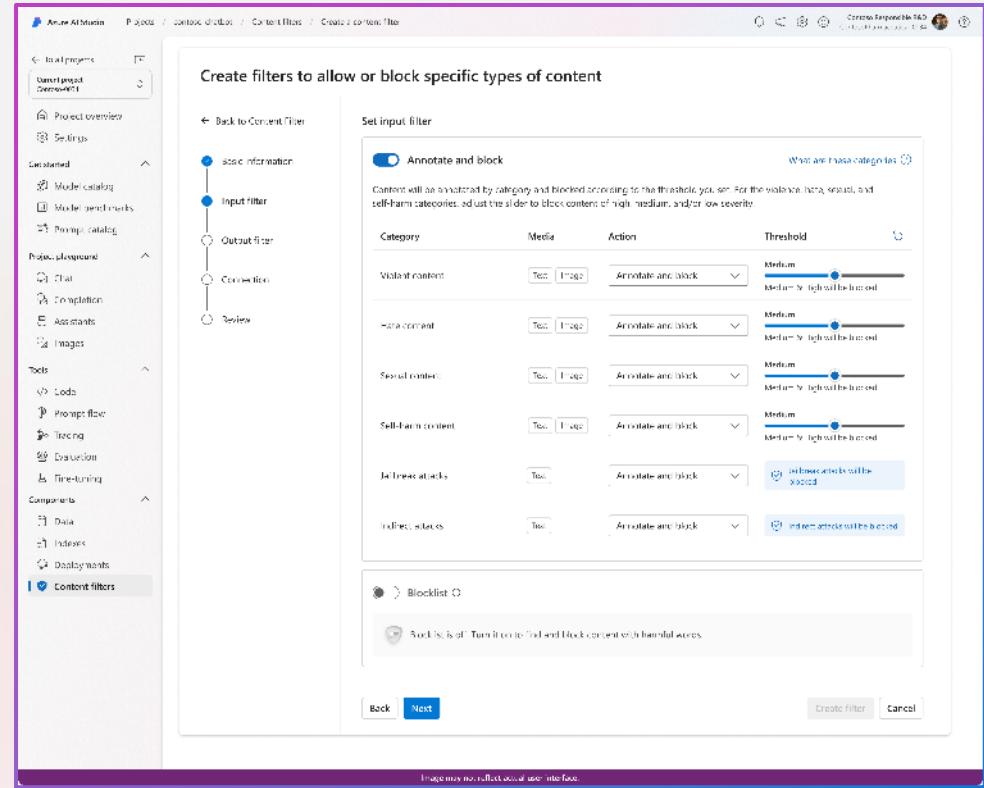
Microsoft first adopted our six AI principles in 2018, and they continue to drive our policy, research, and engineering investments.

# Azure AI Content Safety

Azure AI Content Safety is a content moderation platform that uses AI to keep your content safe, creating better online experiences for everyone with powerful models that detect inappropriate content.

- Enable content harm filters to monitor inappropriate responses that may be hateful, sexual, violent, or lead to self-harm
- Customize content safety thresholds for each user type or individual
- Use Jailbreak Risk Detection to detect user prompts that provoke the GenAI model into breaking set rules
- Secure your data with Cross Prompt Injection Detection to defend against and identify potential Cross-Prompt Injection Attack (XPIA) attacks in input documents and large language model (LLM) conversations
- Identify text in language model output that matches known text context for protected materials
- Create user defined blocklists to manage content

Safeguard with configurable filters and controls



Enhance user safety and data security with Azure AI Content Safety, providing tailored protection, real-time threat detection, and customizable controls.

# **DEMO – Session 3 Lab**

# **Session 4 -**

# **Enterprise-grade**

# **Production at Scale**



# Enterprise-grade Production at Scale

Deploy AI innovations to Azure's managed infrastructure with continuous monitoring and governance across environments

## Deploy to production

Scale AI for use in websites, applications, and other production environments.

## Operationalize and monitor workflows

Continuously monitor AI safety, quality, and token consumption in production. Automate workflows and alerts for timely issue resolution.

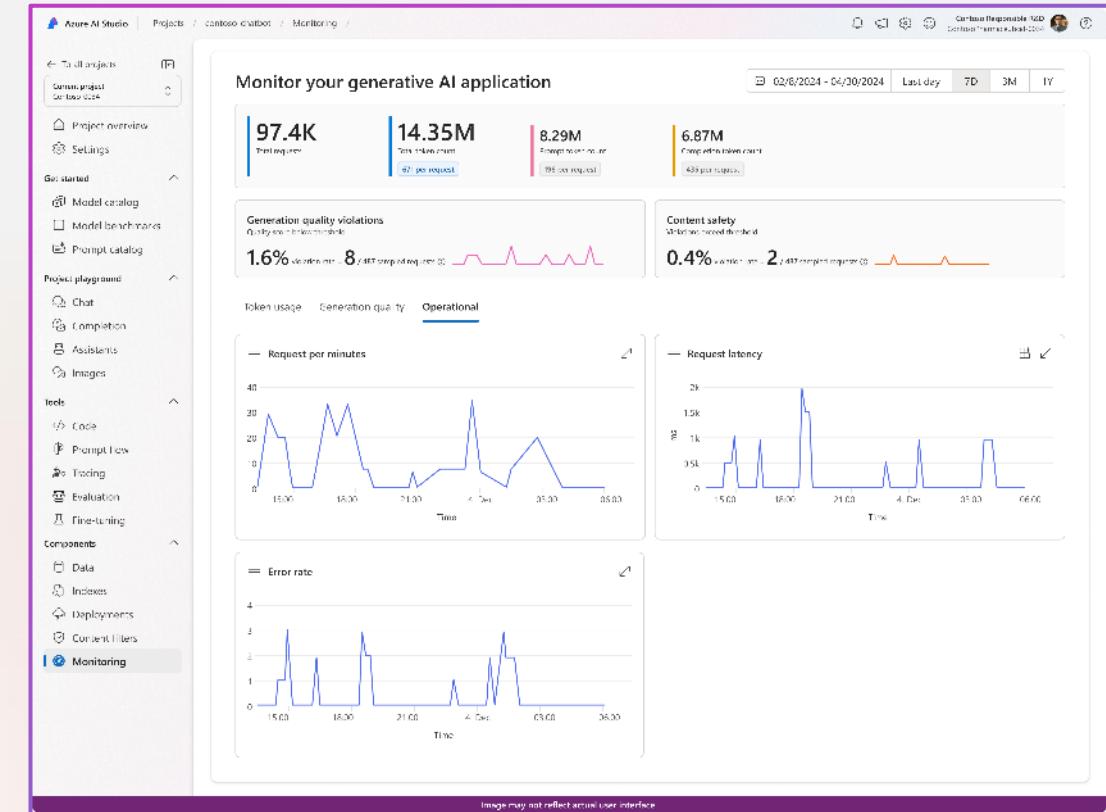
## Enable developer agility and enterprise governance at scale

Provide easy project creation and resource management across the organization and enterprise controls for security, privacy, and compliance.

# Monitoring & observability

Monitoring for generative AI applications makes it easier for developers to monitor the quality, safety, and operational metrics to ensure it's delivering maximum business impact in a safe and compliant manner.

- Evaluate the quality of workflows with rich set of pre-built metrics like groundedness, coherence, fluency, relevance, and similarity
- Enable safety metrics including self-harm, violence, and sexual to monitor harmful contents from the model output
- Keep track of operational metrics such as token usage and latency to ensure optimal system performance, cost-efficiency, and user satisfaction
- Configure alerts for violations based on organizational targets and run monitoring on a recurring basis



 Simplify processes and enhance developer workflows by consolidating design, tuning, testing, and deployment into one platform

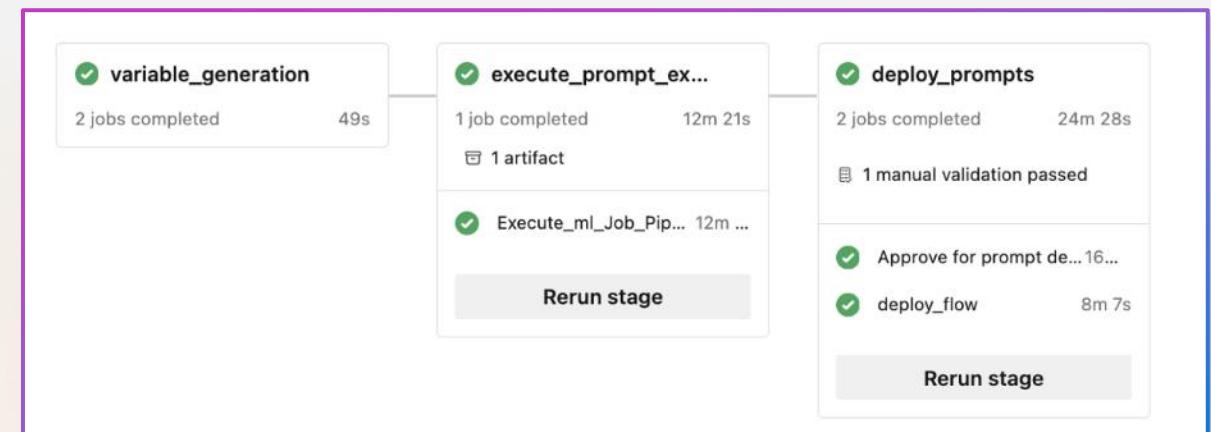
Operationalize and monitor workflows

# Automate workflows

Integrate prompt flows with GitHub Actions/Azure DevOps to automate the LLM-powered application development lifecycle.

- Trigger flow runs, including evaluation runs and batch testing, serving as an automated workflow in Continuous Integration (CI) pipelines
- Deploy flow as an online endpoint in Azure AI Studio for Continuous Deployment (CD), allowing developers to integrate flows into their application
- Deploy flows to multiple targets including Azure App Service, Azure Kubernetes, and managed endpoints, ensuring flows can scale as needed.
- Seamlessly implement A/B deployments, enabling developers to compare different flow versions

Operationalize and monitor workflows



Simplify processes and enhance developer workflows by consolidating design, tuning, testing, and deployment into one platform.

# IT governance at scale

Govern at scale

Enterprise governance provides a balance of SaaS-like self-serve experiences for dev and business teams, customizable configurations for managers, and policy management and oversight by IT teams for enhanced agility, security, and compliance.

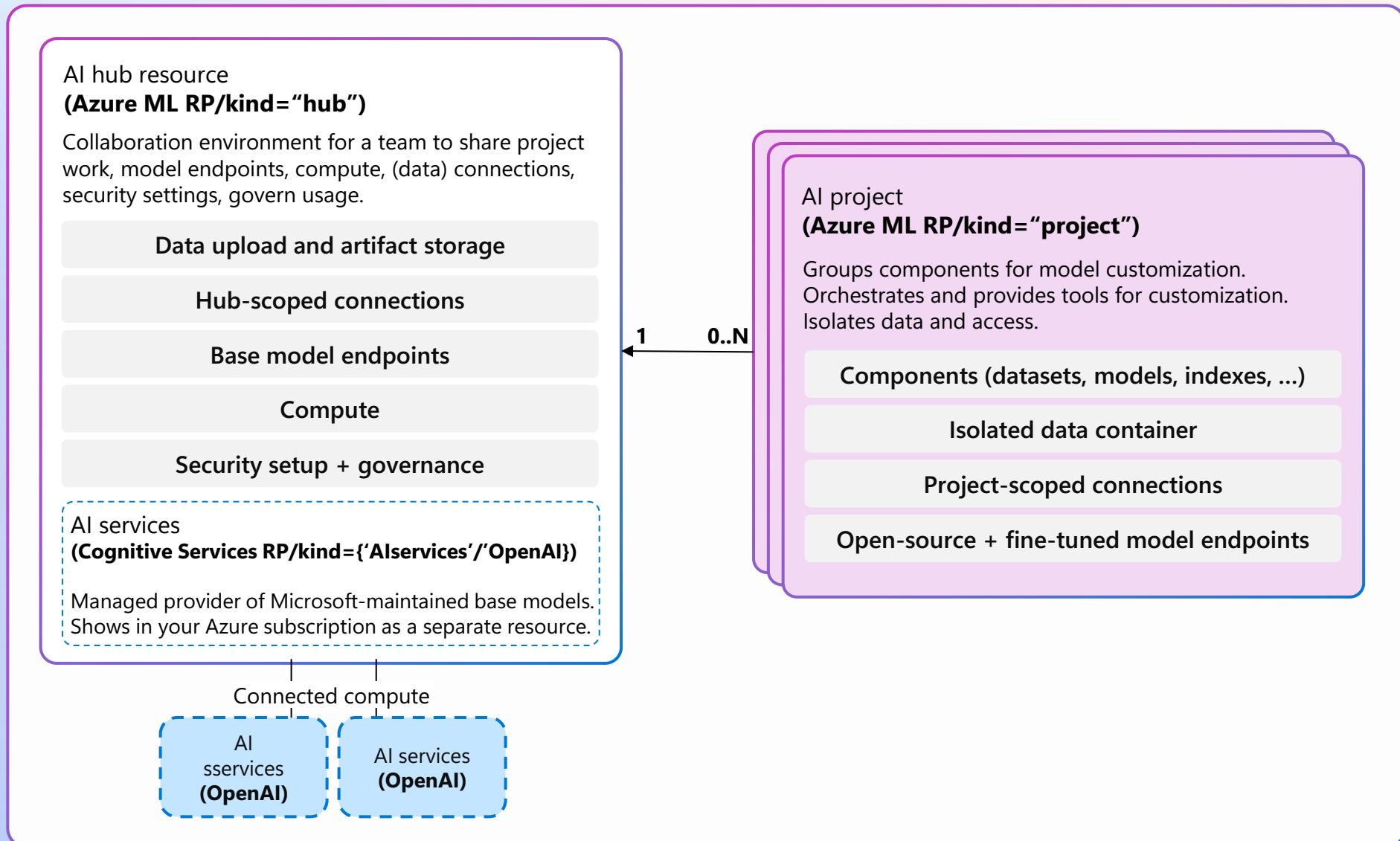
- Self-serve development for a team:** Centrally govern security and shared resource connectivity for a team using hubs, reducing IT bottlenecks.
- Identity & access management:** Leverage built-in roles to manage access for cross-functional teams within collaborative development environments
- Network security:** Streamline your network isolation experience, speed up your workspace setup, and free IT from the hassles of VNET management
- Data protection & encryption:** Secure data in transit and at rest by default and use provided encryption keys or your own customer-managed keys
- Plan deployments:** Get Infra as code templates and landing zone reference implementations
- Costs and quotas:** Centrally plan and manage costs, quotas, and autoscaling for cost efficiency

| Key   | Authentication type | Access                 | Owner                |
|-------|---------------------|------------------------|----------------------|
| ..... | API key             | Shared to all projects | contoso-support      |
| ..... | API key             | Shared to all projects | contoso-support      |
| --    | Microsoft Entra ID  | Shared to all projects | contoso-support      |
| --    | Account key         | customer-support-bot   | customer-support-bot |
| --    | SAS                 | customer-support-bot   | customer-support-bot |
| --    | SAS                 | customer-support-bot   | customer-support-bot |

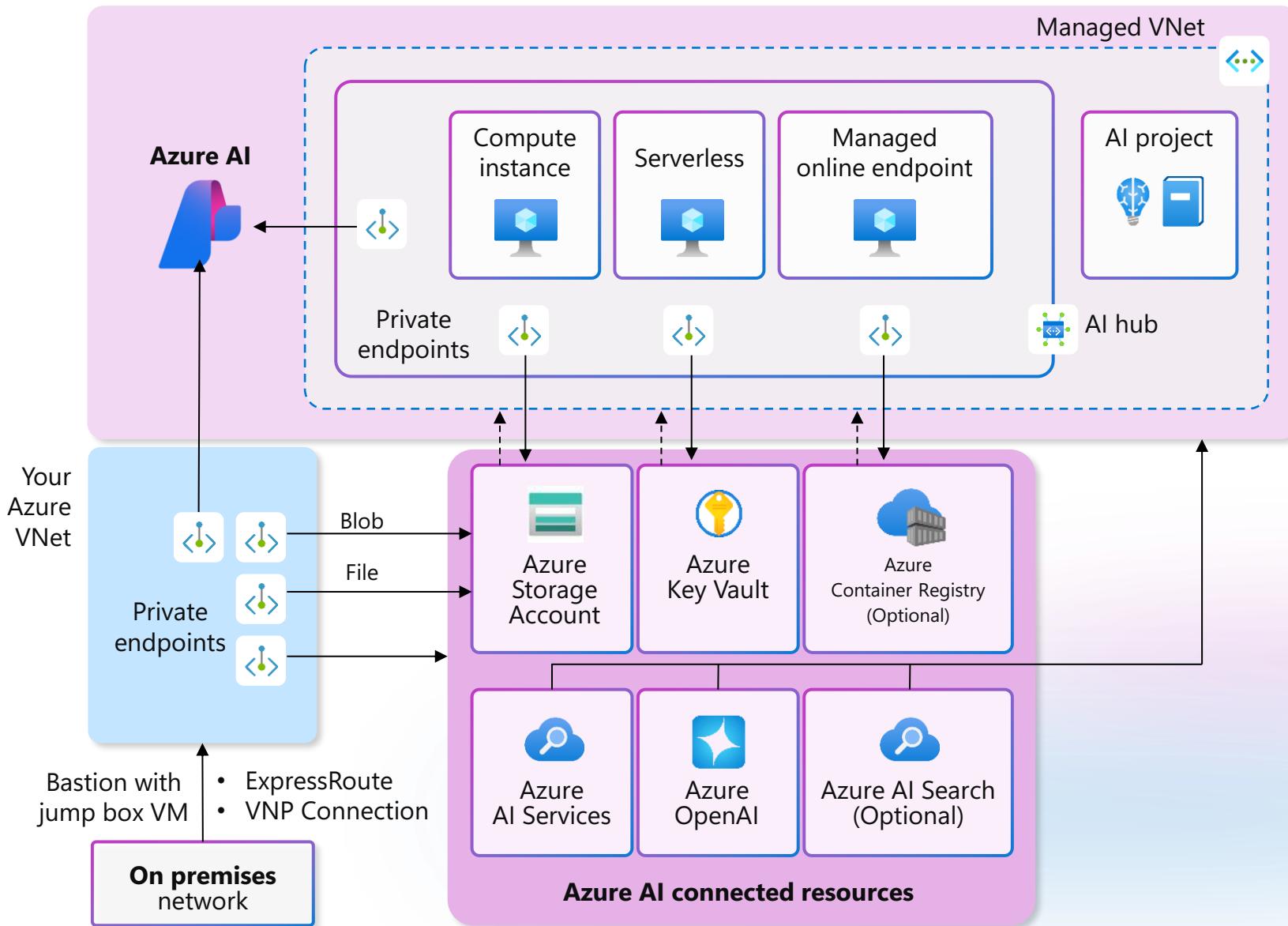


Get started with pre-built RBAC roles for Owner, Contributor, Reader, AI Developers, and AI Inference Deployment Operator roles at the Azure AI hub or Azure AI project level.

# Azure AI Studio Hubs



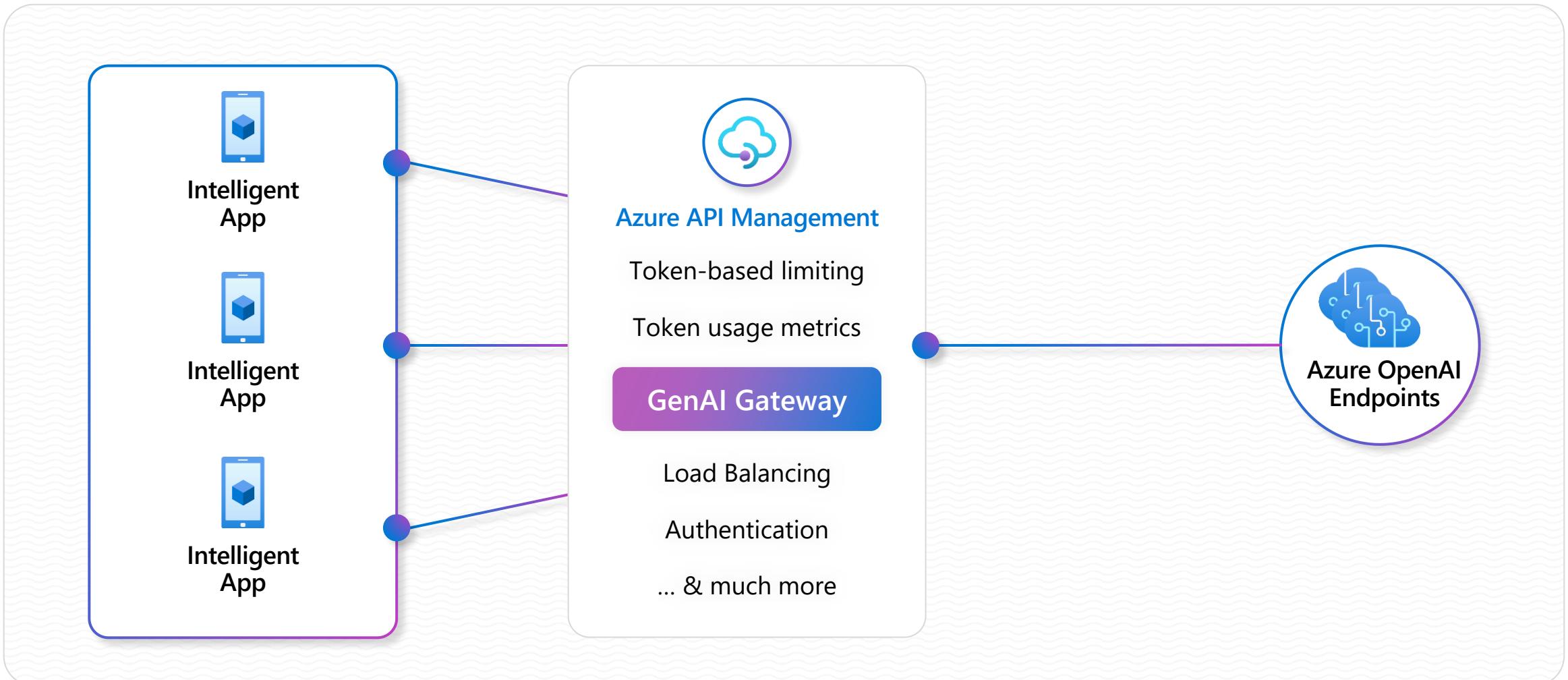
# Azure AI Studio private links



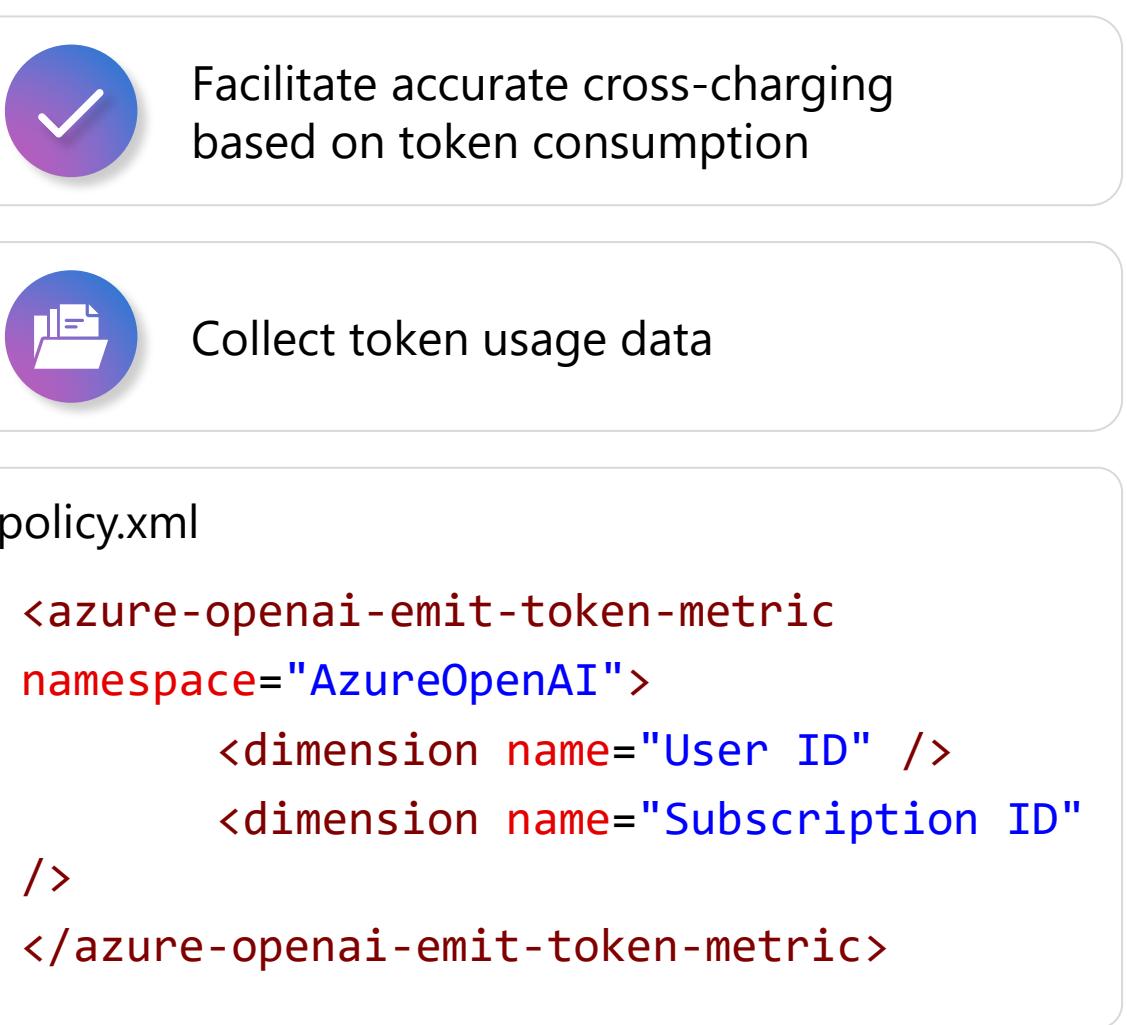
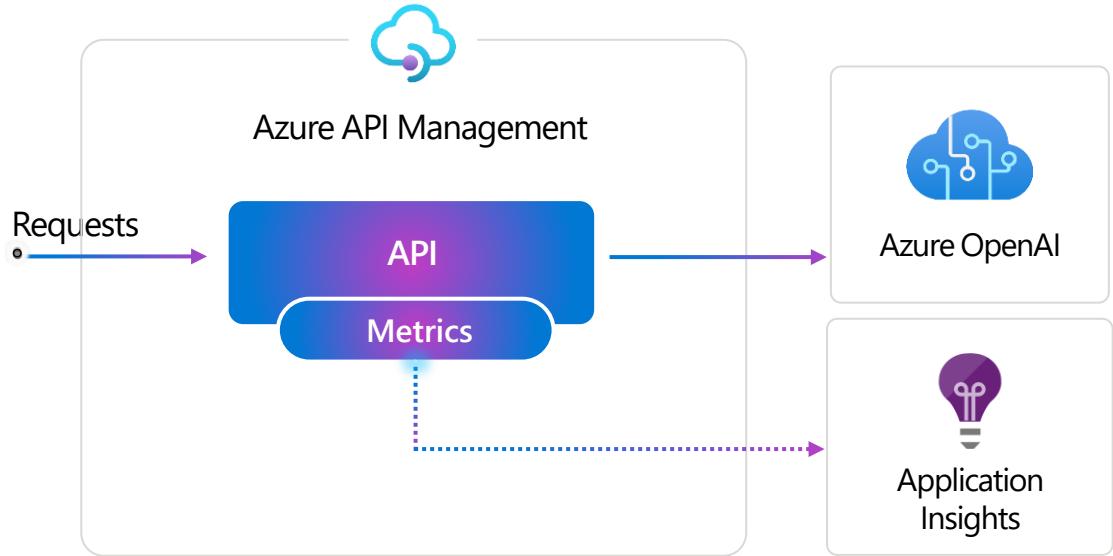


Scaling up using  
Azure API  
Management

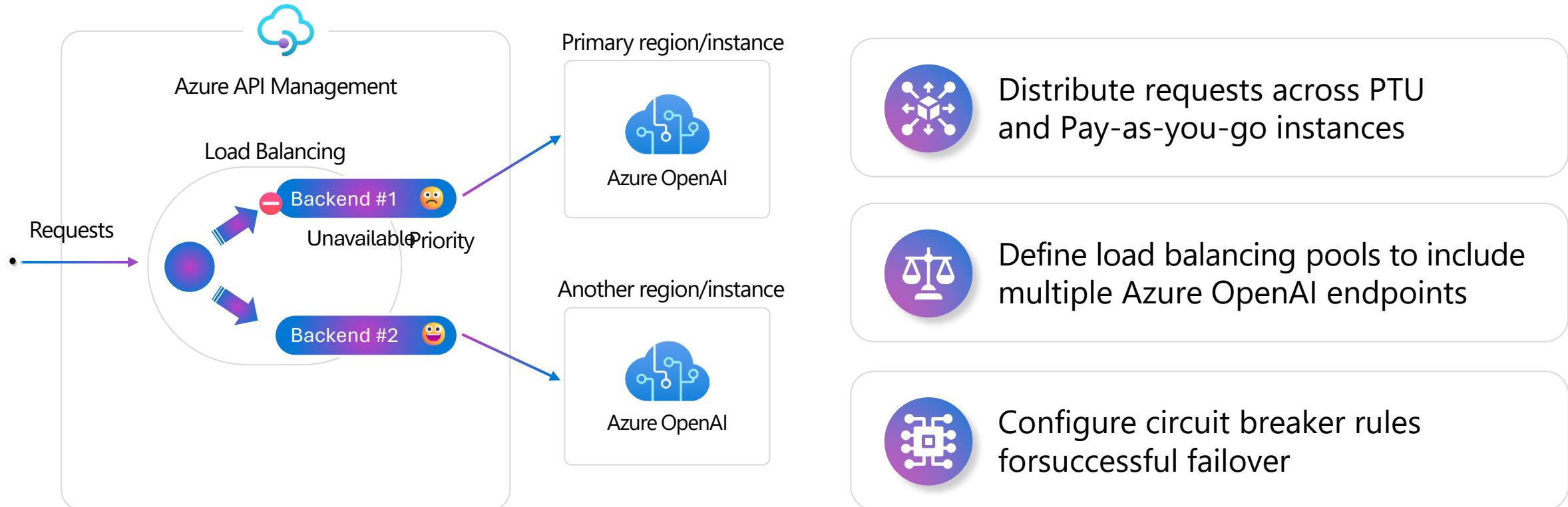
# Scaling Up: Multiple Apps, Multiple OpenAI Endpoints



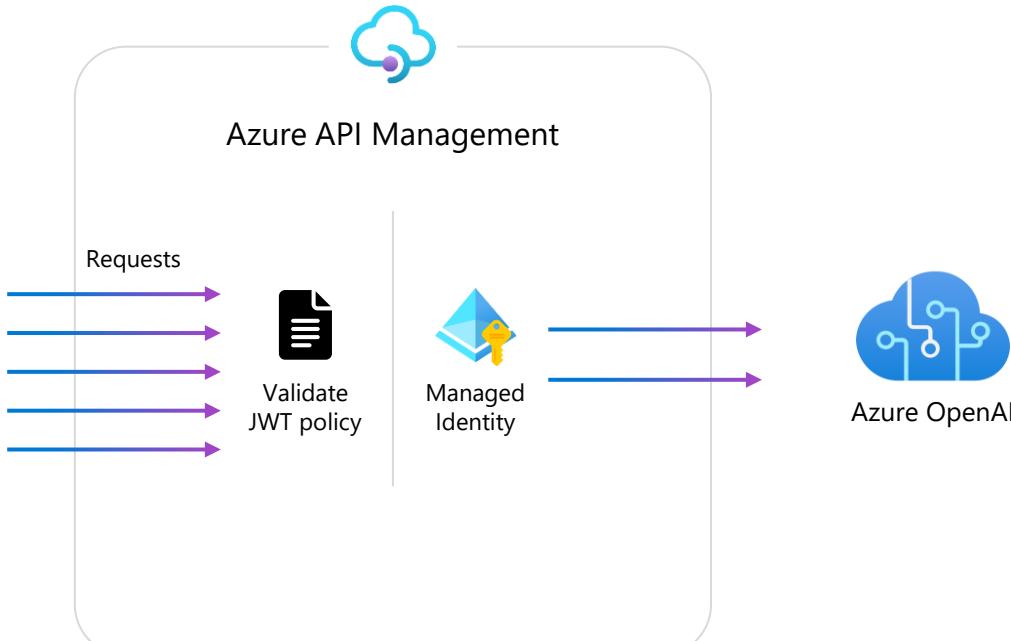
# [GA] Azure OpenAI Token Metric policy



# [GA] Load Balancer and Circuit Breaker



# Authentication and Authorization

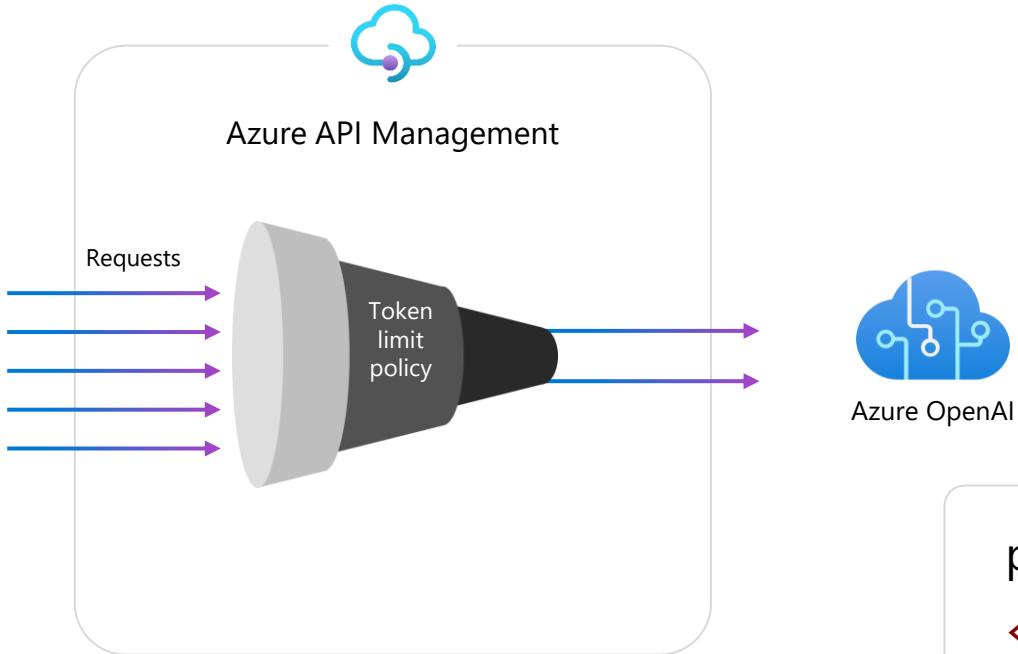


Configure managed identity authentication

Validate claims in JWT to manage access to OpenAI endpoints

Authenticate API consumers using subscription keys

# [GA] Azure OpenAI Token Limit policy



Configure tokens per minute (TPM) limits based on counter keys

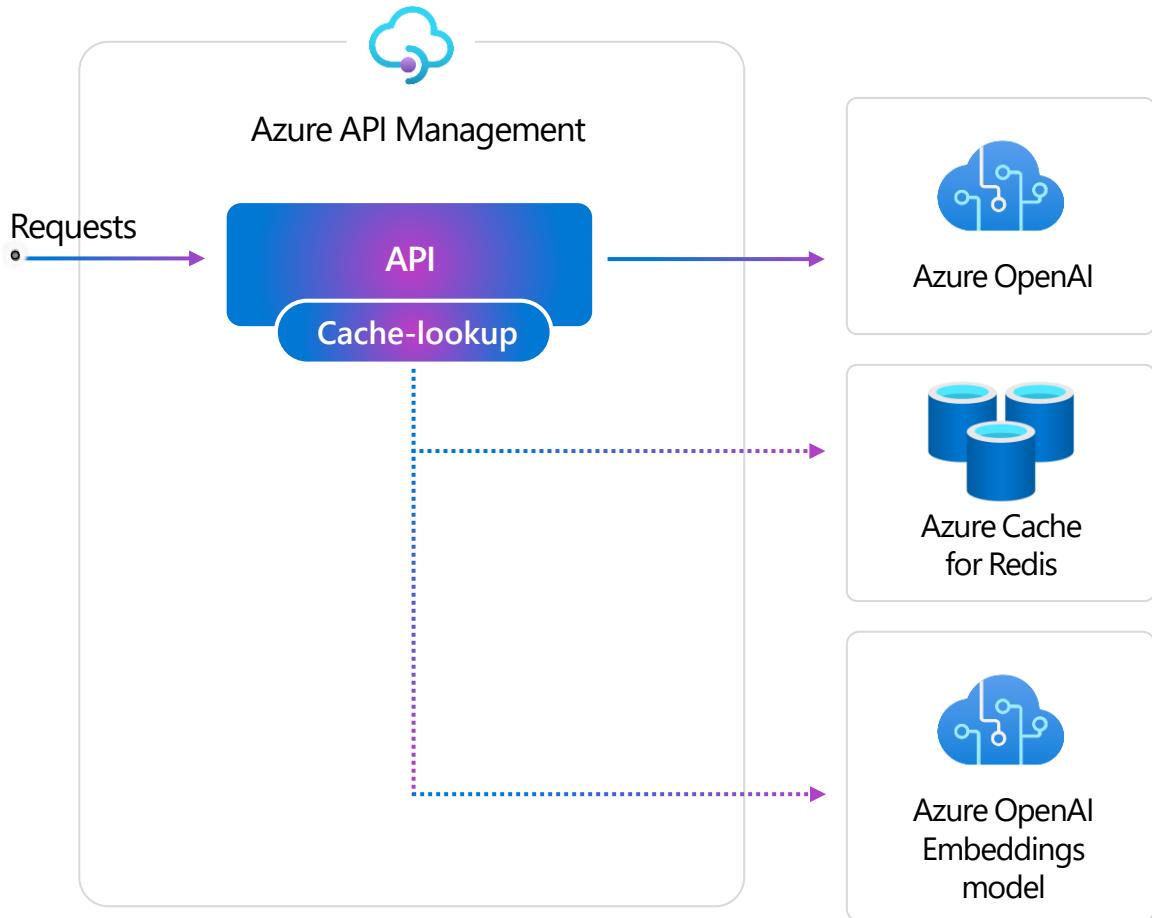


Define policy behavior for throttling

policy.xml

```
<azure-openai-token-limit  
    counter-key="@({context.Subscription.Id})"  
    tokens-per-minute="1000"  
    estimate-prompt-tokens="false" />
```

# [Preview] Azure OpenAI Semantic Caching policy



Configure semantic caching for all API consumers



Define similarity score threshold for caching

policy.xml

```
<azure-openai-semantic-cache-lookup  
    score-threshold="0.05"  
    embeddings-backend-id="azure-openai-backend">  
    <vary-by>@(context.Subscription.Id)"</vary-by>  
</azure-openai-semantic-cache-lookup>
```

# **Session 5 - AI Innovation Across landscapes**

# Choose where to build your copilot

Three approaches to support the developer experience



## Copilot Studio *Power Platform*

- Build your own copilot using visual building experiences
- Customize Microsoft Copilots with your own enterprise scenarios
- Leverage a connected, integrated platform
- Manage copilot experiences with full visibility into customizations



## Azure AI Studio *Azure*

- Explore, build, test, deploy, and manage custom copilots using interactive visual and code-first workflows
- Access out-of-the-box and customizable APIs and models
- Systematically evaluate model and app responses and pinpoint fine-tuning opportunities
- Scale copilots for use in websites, applications, and other production environments



## AI Toolkit for Visual Studio Code *Agnostic*

- Simplify app development by bringing together AI development tools and models
- Explore, try, and integrate AI models
- Use local and cloud compute to fine-tune and optimize small language models for app-specific use cases on the cloud and edge
- Package fine-tuned models as containers and deploy to Azure or the edge

# Capabilities Across Experiences



|  | Copilot Studio | Azure AI Studio | AI Toolkit for Visual Studio Code |
|--|----------------|-----------------|-----------------------------------|
| Extend Microsoft's copilot capabilities                              | ✓              |                 | ✓                                 |
| Build your own copilot   | ✓              | ✓               | ✓                                 |
| API and Model Choice   |                |                 |                                   |
| Access models from Azure OpenAI Service                              | ✓              | ✓               | ✓                                 |
| Access additional foundation and open models                         |                | ✓               | ✓                                 |
| Open-source models scanned for security and integrity                | ✓              |                 |                                   |
| Deploy multiple models   |                | ✓               | ✓                                 |
| Deploy out-of-the-box, customizable, interoperable APIs              | ✓              | ✓               | ✓                                 |
| Build your own GPTs / Deploy Assistant API                           | ✓              | ✓               | ✓                                 |
| Incorporate multimodality  | ✓              | ✓               |                                   |
| Complete AI Toolchain  |                |                 |                                   |
| Build using a code-first approach                                    |                | ✓               | ✓                                 |
| Ground models using Microsoft Graph data                             | ✓              |                 |                                   |
| Ground models using web data   | ✓              | ✓               | ✓                                 |
| Ground models using your own data                                    | ✓              | ✓               | ✓                                 |
| Fine-tune models   |                | ✓               | ✓                                 |
| Orchestrate data, models, and prompt instructions                    | ✓              | ✓               | ✓                                 |
| Integrate with OSS frameworks (e.g. SemanticKernel, LangChain, etc)  |                | ✓               | ✓                                 |
| Debug models and prompt instructions                                 |                | ✓               | ✓                                 |
| Test your application  | ✓              | ✓               | ✓                                 |
| Run systematic manual and automated evaluations over models and apps |                | ✓               | ✓                                 |
| Responsible AI Tools & Practices                                     |                |                 |                                   |
| Apply and configure content safety filters                           |                | ✓               | ✓                                 |
| Access built-in tooling for responsible AI practices                 | ✓              | ✓               | ✓                                 |
| Enterprise-grade Production at Scale                                 |                |                 |                                   |
| Deploy on cloud  | ✓              | ✓               | ✓                                 |
| Deploy on edge   |                | ✓               | ✓                                 |
| Manage application lifecycle   | ✓              | ✓               | ✓                                 |
| Deploy with Customer Copyright Commitment for Azure OpenAI Service   |                | ✓               | ✓                                 |
| Deploy with Azure security and compliance                            |                | ✓               | ✓                                 |

# Azure AI + Microsoft Fabric = boundless possibilities



Your gateway to AI apps and custom copilot experiences

- Enhance data insights**  
Build enterprise chat applications to uncover insights using natural language from structured, unstructured, and real-time data
- Ground your AI on unstructured data**  
Ensure your AI is using the latest data in its interactions with your employees, partners, and customers
- Analyze customer interactions**  
Build speech analytics applications to enhance customer service, tailor support responses, and make data-driven decisions
- Customize machine learning models**  
Train, deploy, and orchestrate machine learning models tailored to specific business needs—from predictive maintenance to customer sentiment analysis



Your unified solution for analytics



OneLake serves as the connective tissue to build generative AI apps powered by your data



# Microsoft Fabric

## The data platform for the era of AI

Intelligent data foundation



Data  
Factory



Data  
Engineering



Data  
Warehouse



Data  
Science



Real Time  
Analytics



Power BI



Data  
Activator



Powered by AI with Copilot in Microsoft Fabric



OneLake

UNIFIED

SaaS product experience

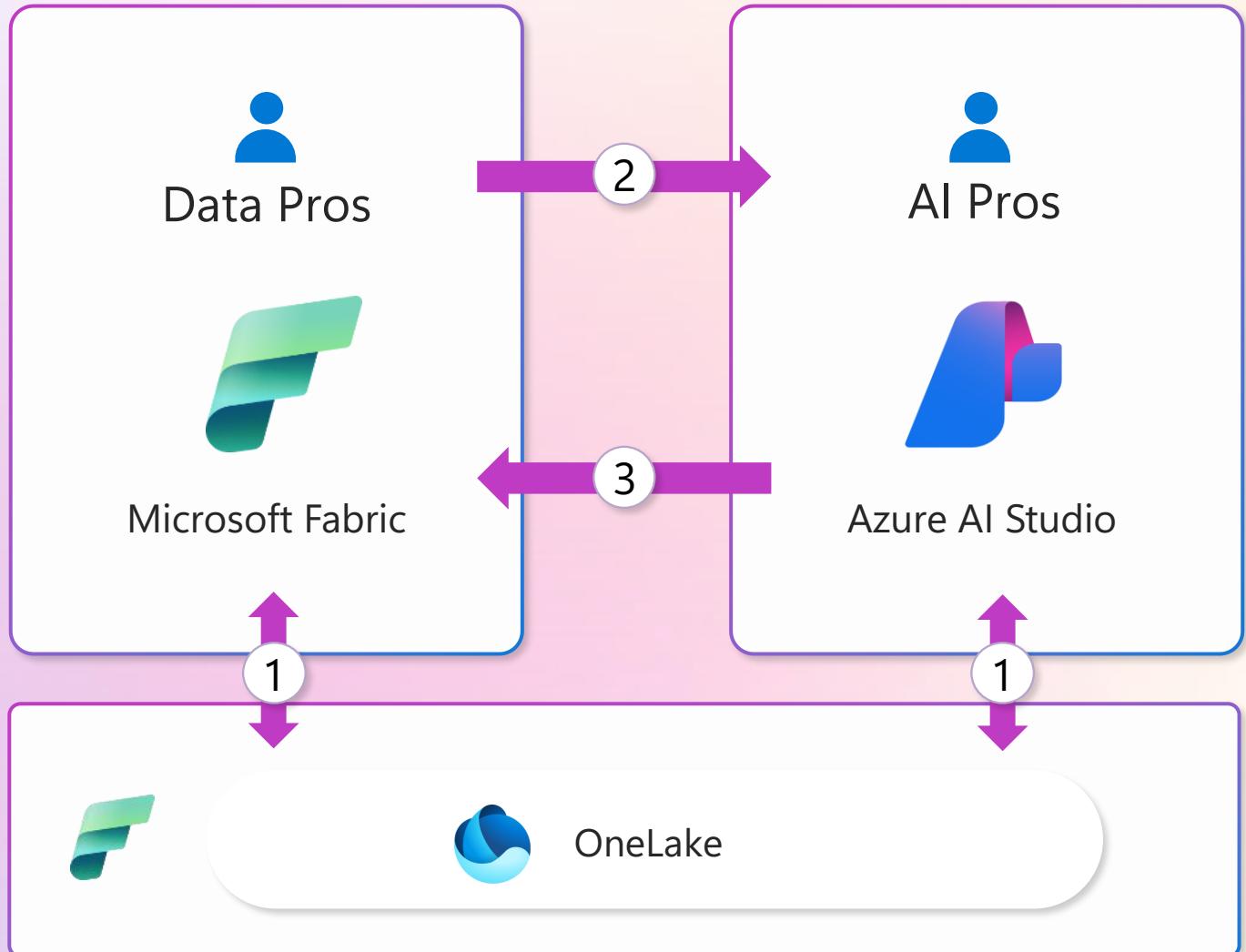
Security and governance

Compute and storage

Business model

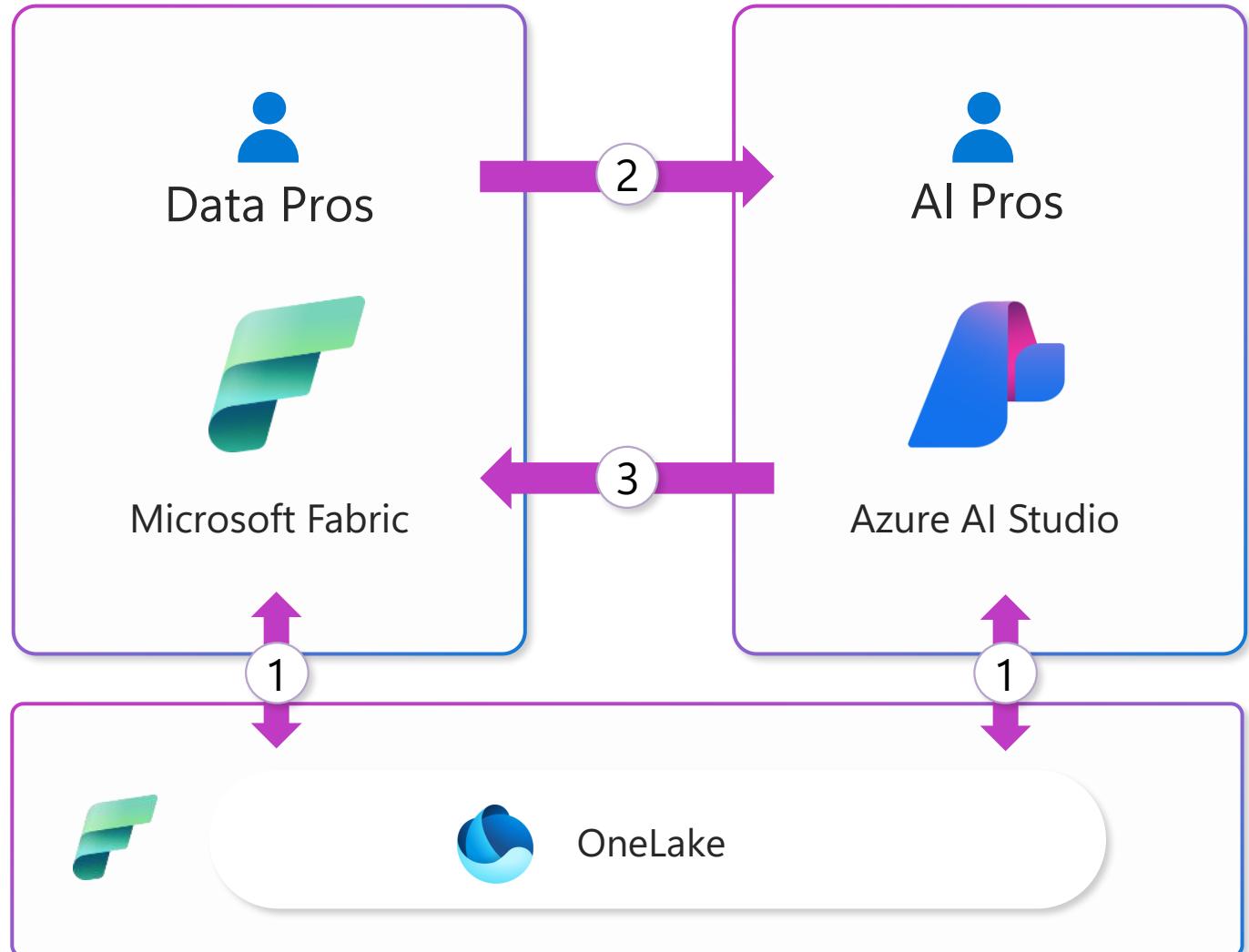
# Microsoft Fabric and Azure AI Studio integration

- 1 Data and AI Pros access, use and process data across the AI lifecycle with OneLake as the connective tissue.
- 2 Data Pros can pre-process and share AI-ready data with AI Pros. Once in Azure AI Studio, AI Pros can create their own generative AI applications and copilot experiences.
- 3 AI Pros can make their models and generative AI apps accessible in Microsoft Fabric for Data Pros to enrich their analytics workflows in their lakehouse and serve through Power BI.

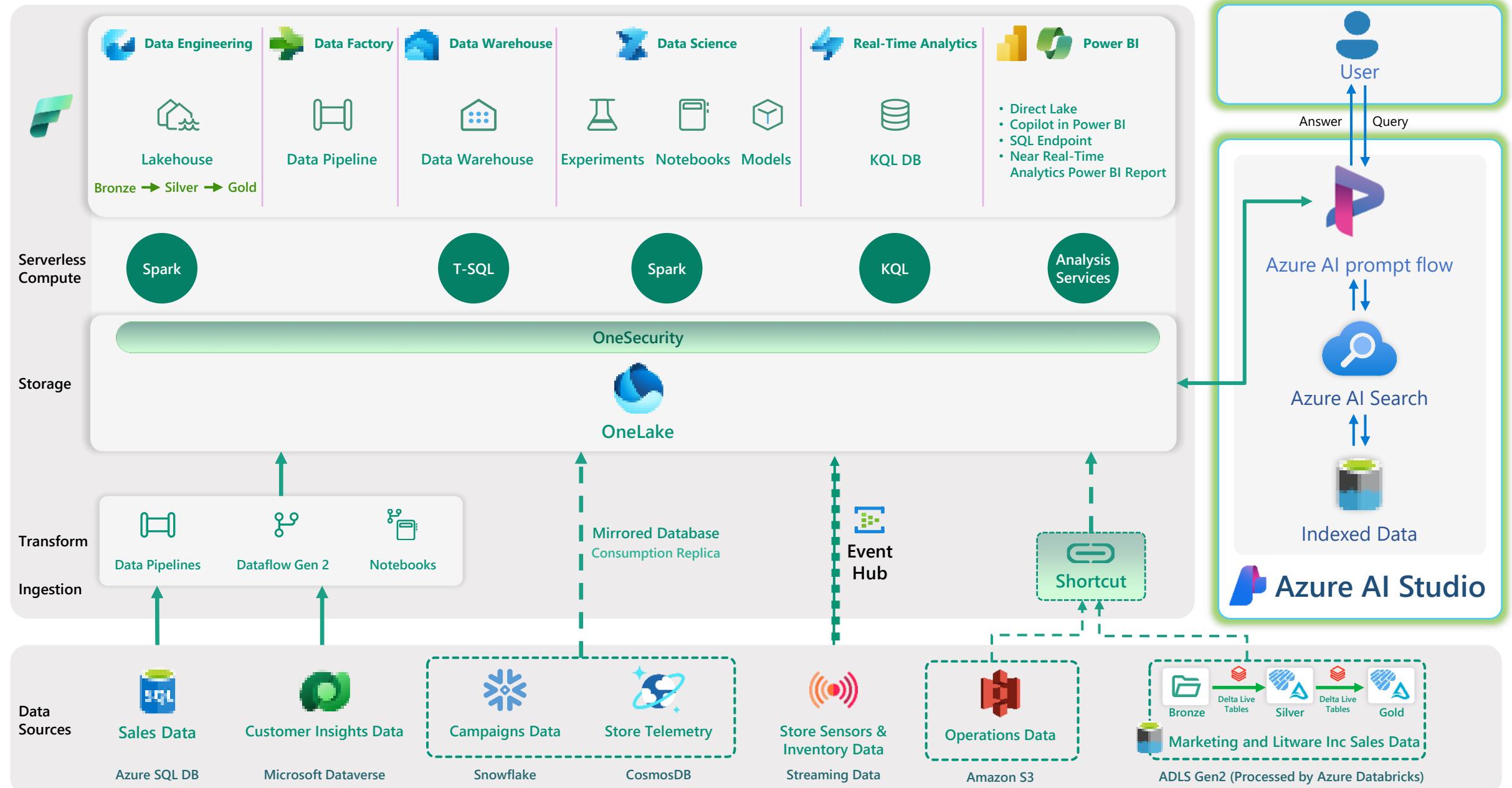


# Microsoft Fabric and Azure AI Studio integration

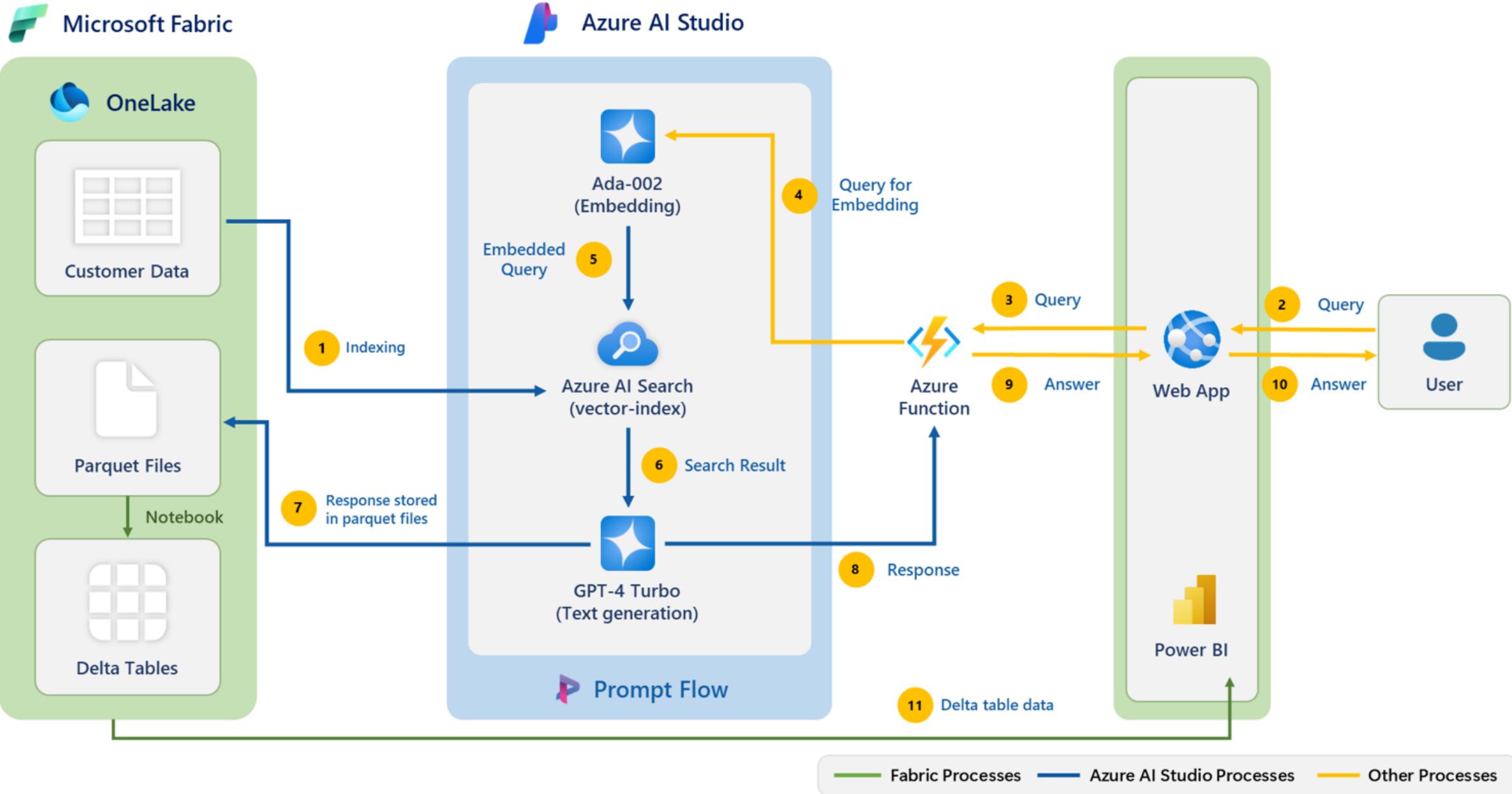
- 1 Data and AI Pros access, use and process data across the AI lifecycle with OneLake as the connective tissue.
- 2 Data Pros can pre-process and share AI-ready data with AI Pros. Once in Azure AI Studio, AI Pros can create their own generative AI applications and copilot experiences.
- 3 AI Pros can make their models and generative AI apps accessible in Microsoft Fabric for Data Pros to enrich their analytics workflows in their lakehouse and serve through Power BI.



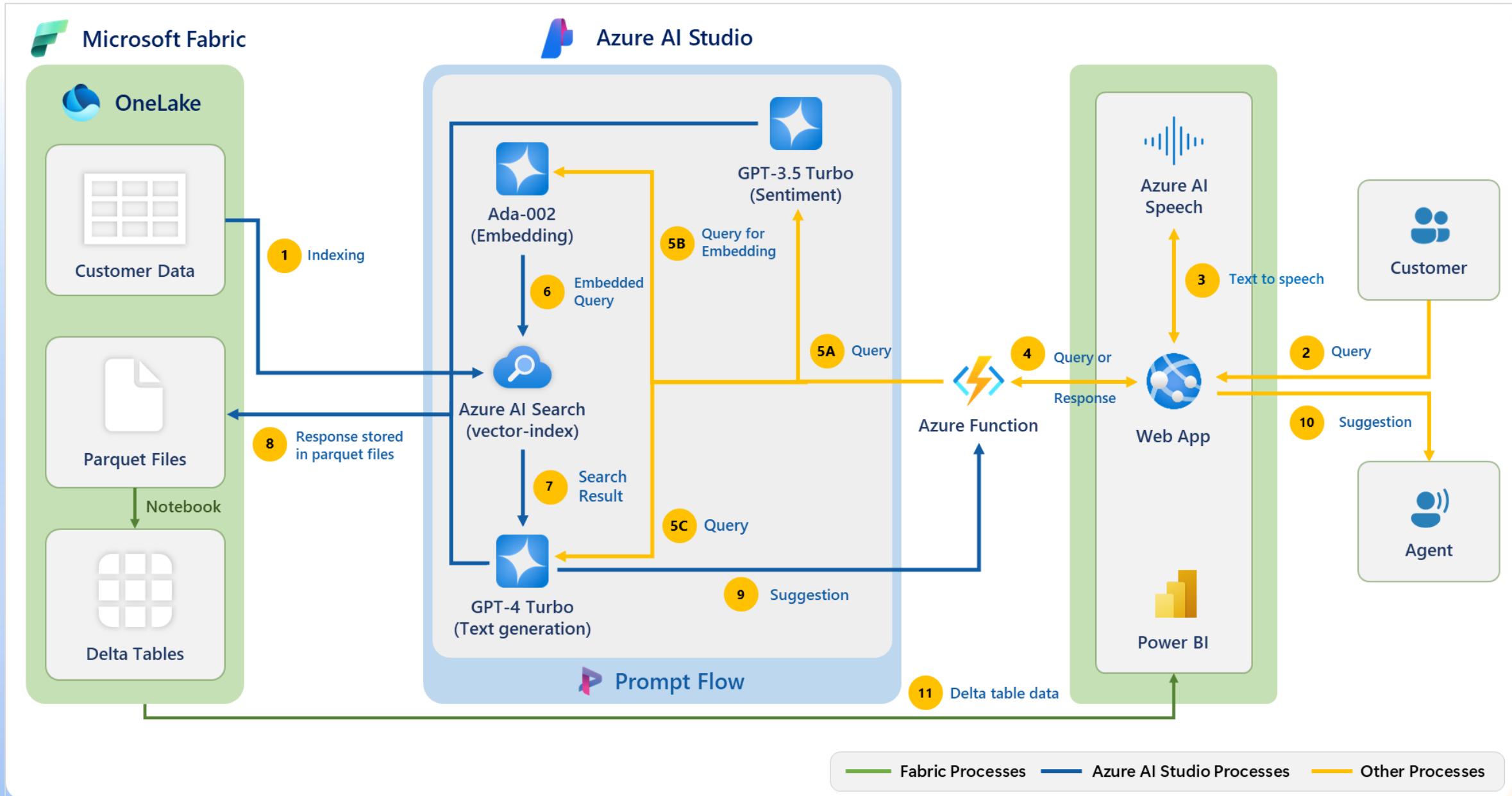
# Microsoft Fabric and Azure AI Studio Architecture



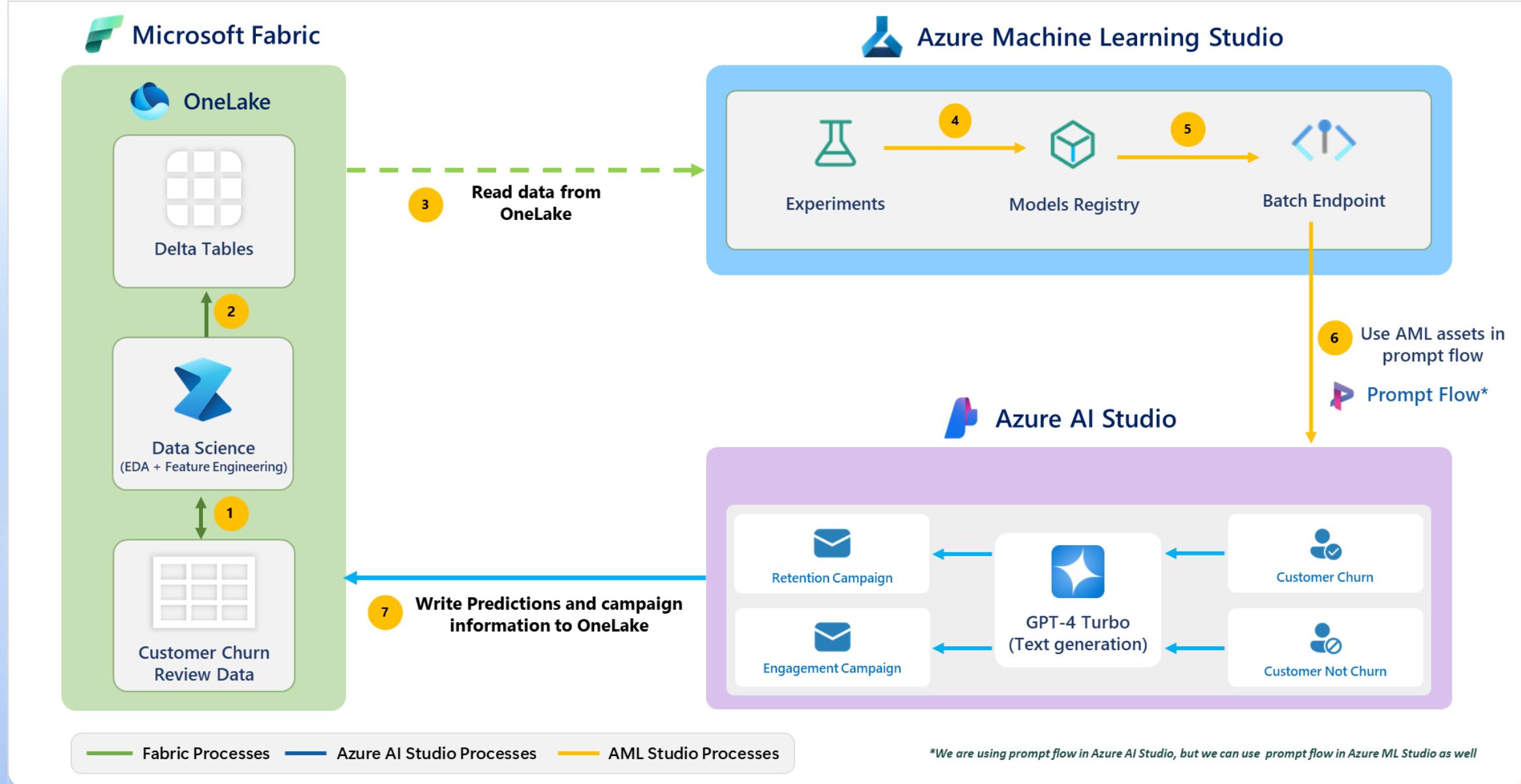
# Enhance Data Insights



# Analyze Customer Interactions



# Customize Machine Learning Models



# Use cases

# Top use cases for

## Azure AI Studio



**Build your own copilot**  
Your data. Your apps. Your people

**Enterprise chat**  
Provide multi-modal knowledge mining

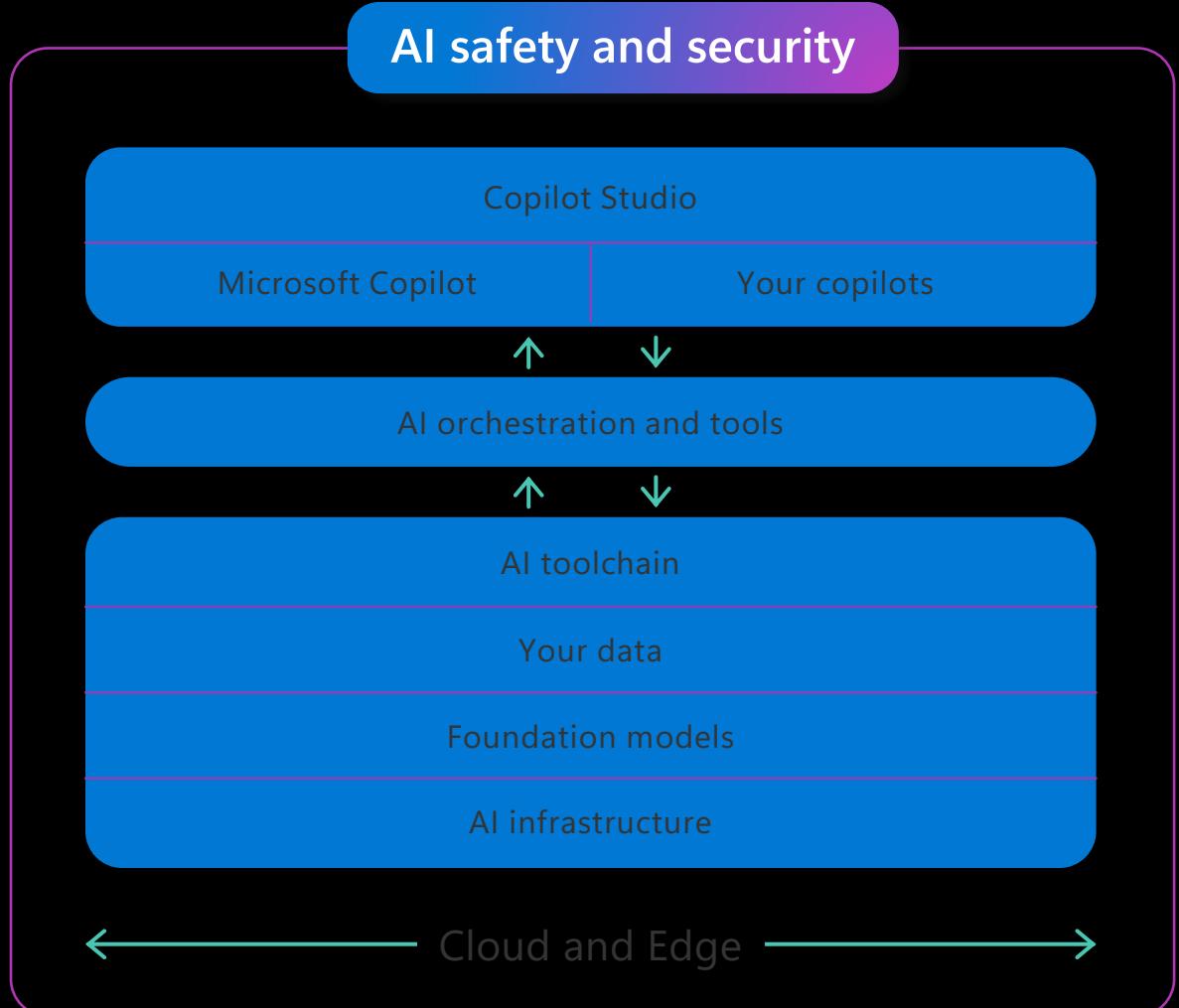
**Speech analytics**  
Improve interactions and service

**Content generation**  
Deliver new products and services

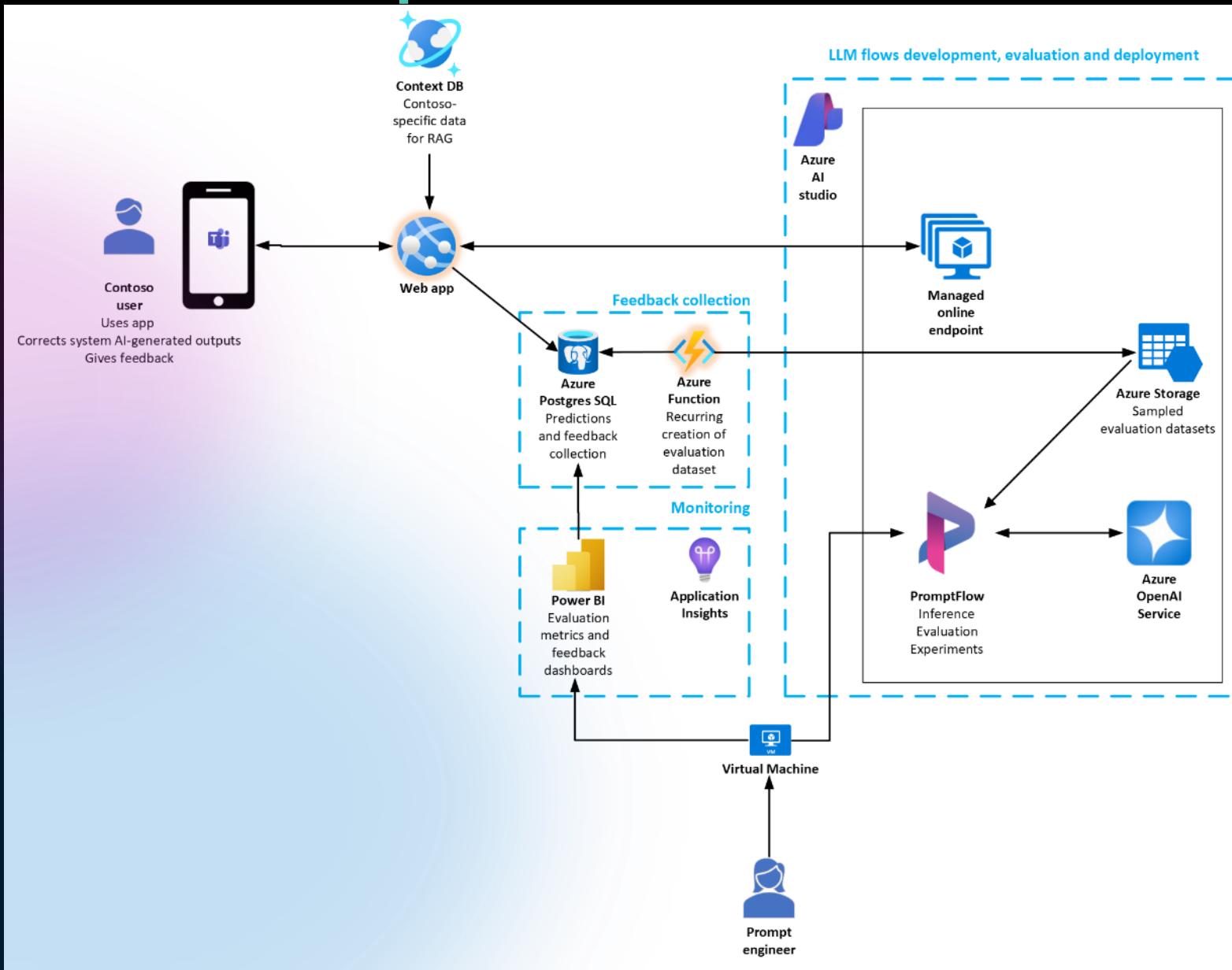
**Hyper-personalization**  
Support better sales and marketing

# Copilot development stack

this will become an expectation for how all software works

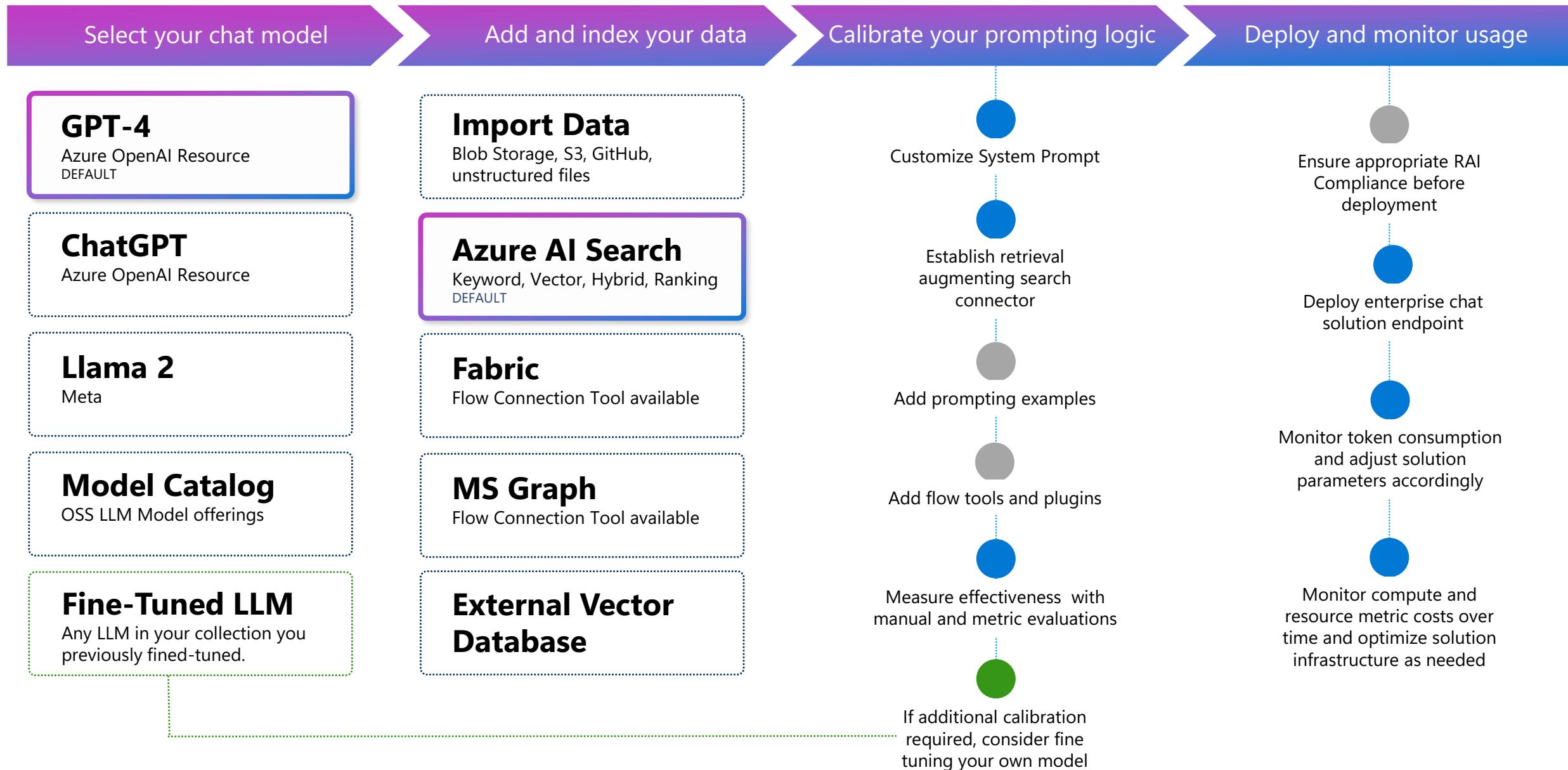


# Build your own copilot with Microsoft Teams



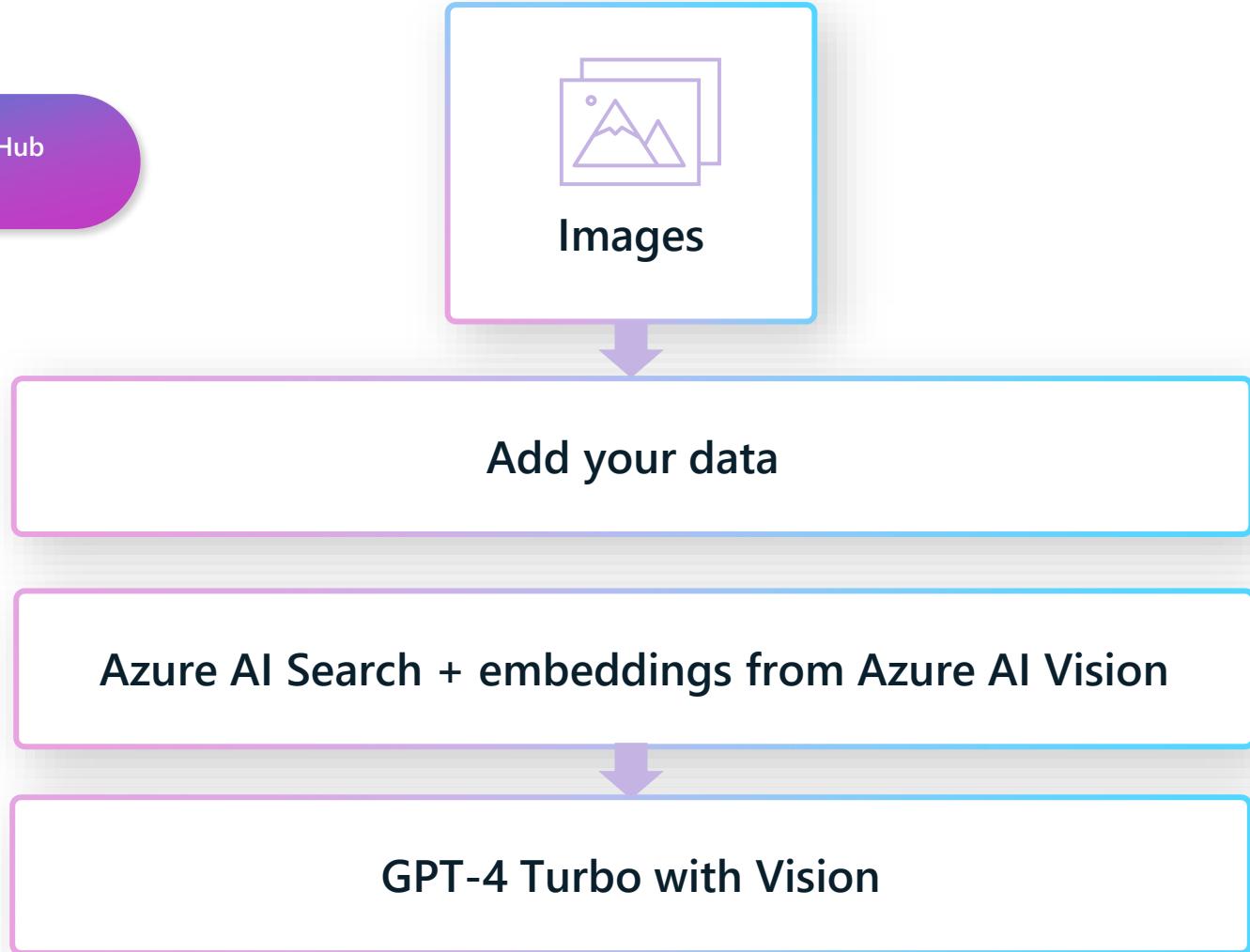
# Enterprise chat High-level Architecture

Enterprise chat



# Let's look at an example: Contoso Outdoor Company

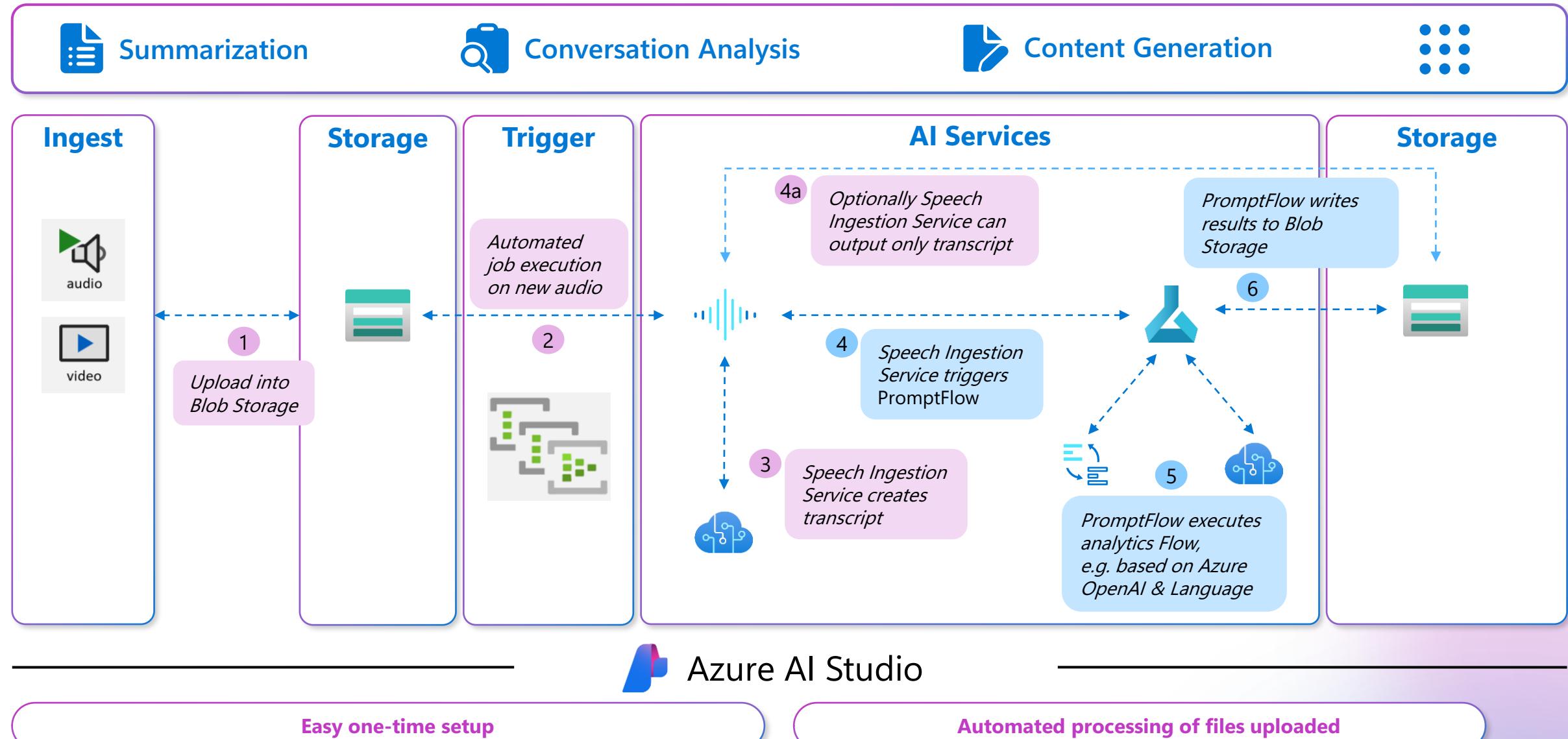
Get the sample code on GitHub  
[Contoso Chat](#)



# Speech analytics High-level Architecture

Speech analytics

Analytics results



# Getting started

## Next Steps

- Speak to your Microsoft account team to
  - Schedule deep dive sessions
  - Get Help building a POC
- Kick the tires on the Platform and Azure AI studio
- If you have an existing use case in development, speak to Microsoft account team on how you can scale it enterprise grade

Organizations are asking not only how, but *how fast*, they can apply this next generation of AI to address the biggest opportunities and challenges they face — **safely and responsibly.**”

Satya Nadella, Microsoft Chairman  
and CEO

# Accelerate AI deployment with Azure AI Studio

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



Define scope & requirements



Prepare data



Build, train, & evaluate



Review & approve app for production



Deploy & monitor



Govern and monitor continuously

## Microsoft is built on trust

1

Your data **is**  
**your data**

2

Your data **is**  
**not used** to  
train or enrich  
foundation  
AI models

3

Your data and  
AI models **are**  
**protected** at  
every step

4

Our Customer  
Copyright  
Commitment

# Ready to get started?



Build with  
[Azure AI Studio](#)



Learn how to build  
AI solutions  
[Azure AI Studio](#)  
[Training Modules](#)



Explore technical  
documentation to  
scale your project  
[Azure AI Studio](#)  
[documentation](#)



Get the latest  
[Azure AI news](#)  
[and resources](#)

**Thank you**