



AI-900

Generative AI

# AI-900 Agenda



1: AI Overview

2: Computer Vision

3: Natural Language Processing

4: Document Intelligence and Knowledge Mining

5: Generative AI 

# LP Agenda

---

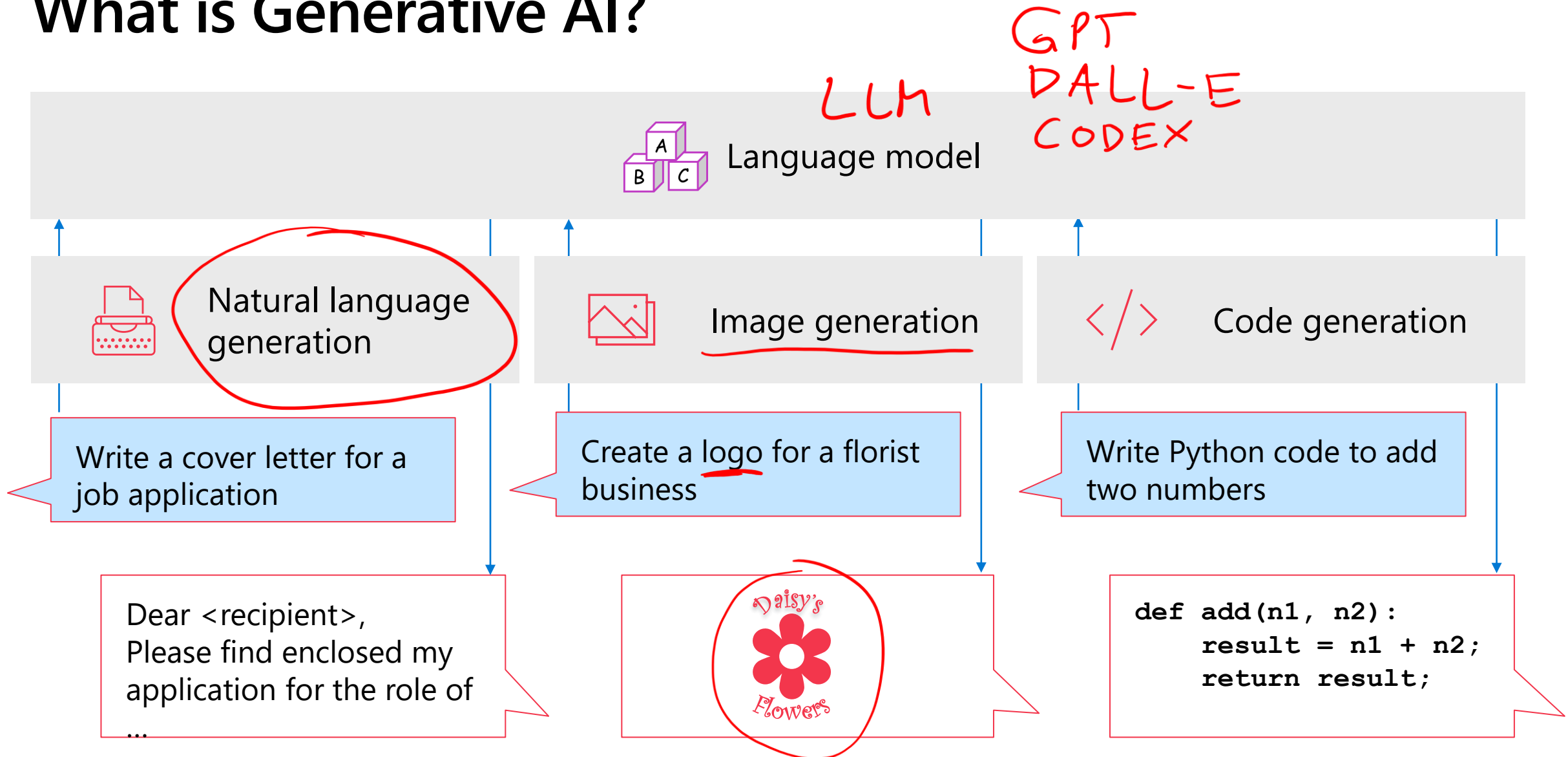
- Fundamentals of generative AI
- Introduction to Azure AI Studio
- Explore responsible generative AI

LLM Embedding Vectors

# Fundamentals of Generative AI

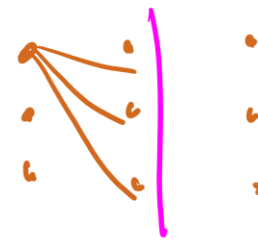


# What is Generative AI?

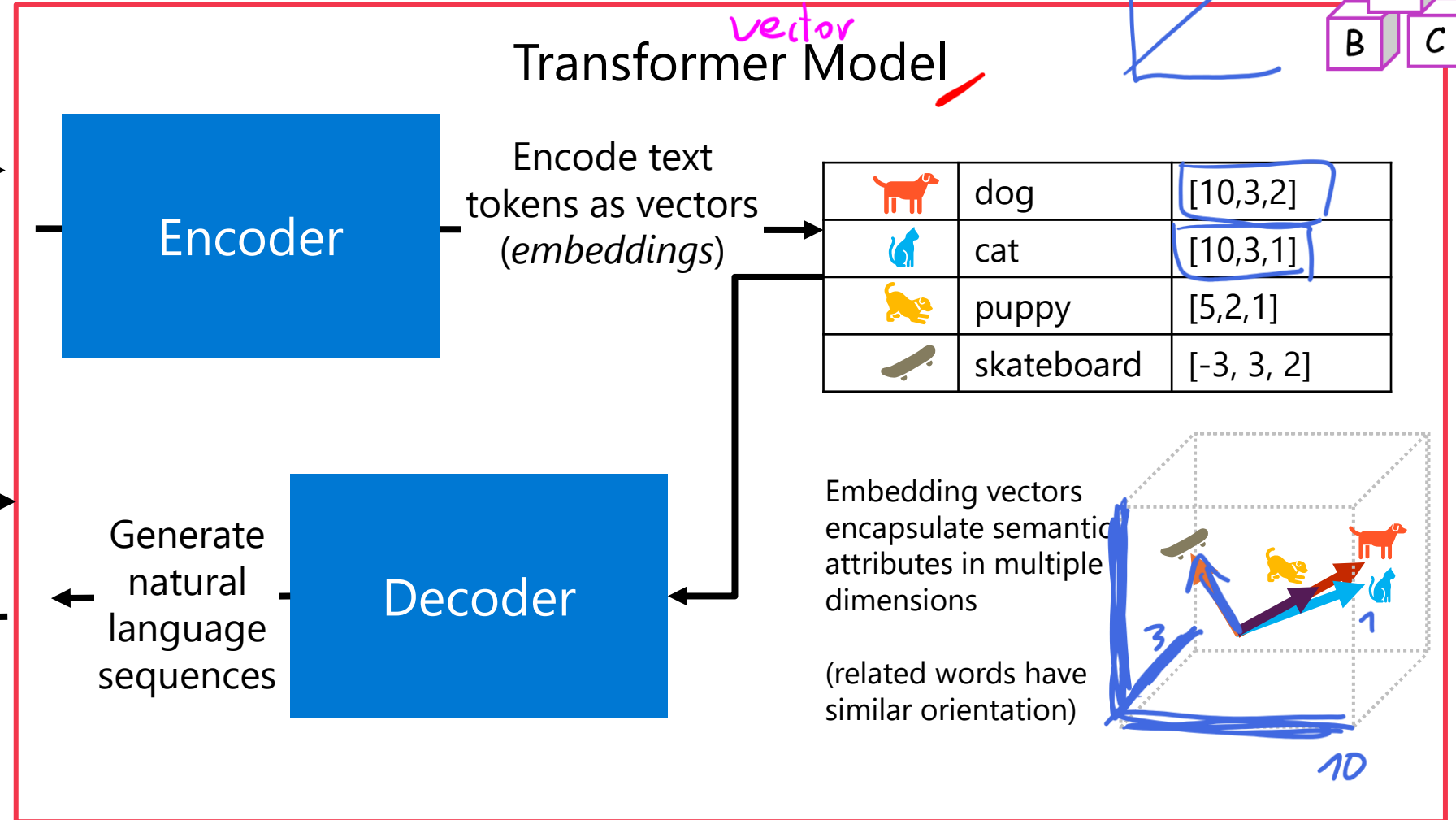
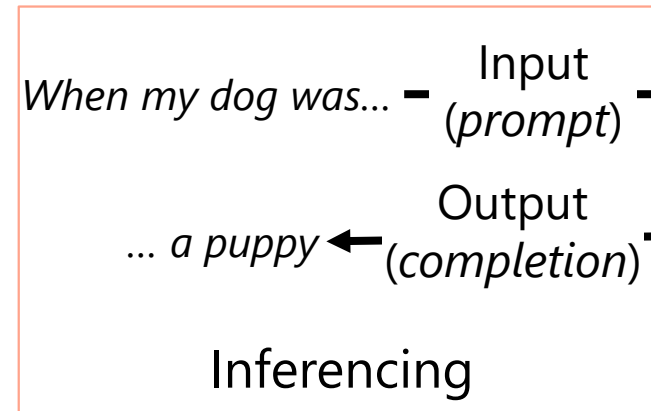
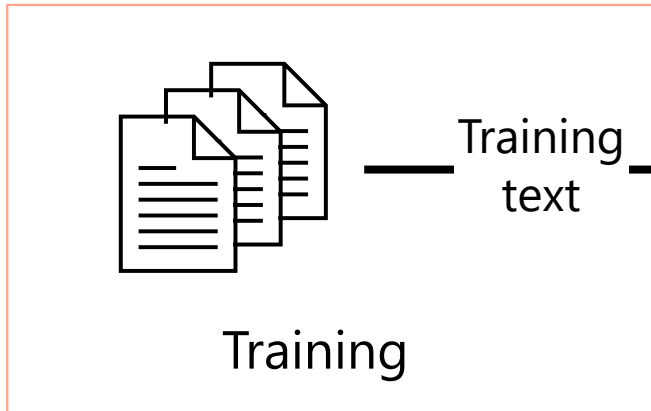
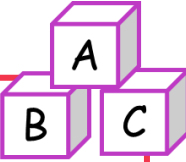
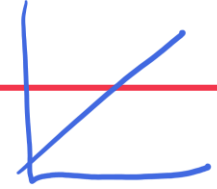


# Language models

A (very) high-level overview



vector



# Language models – tokenization

## Step one: tokenization

The first step in training a transformer model is to decompose the training text into *tokens*.

**Example sentence:** *I heard a dog bark loudly at a cat.*

"I"=1

"heard"=2

"a"=3

"dog"=4

"bark"=5

"loudly"=6

"at"=7

"cat"=8

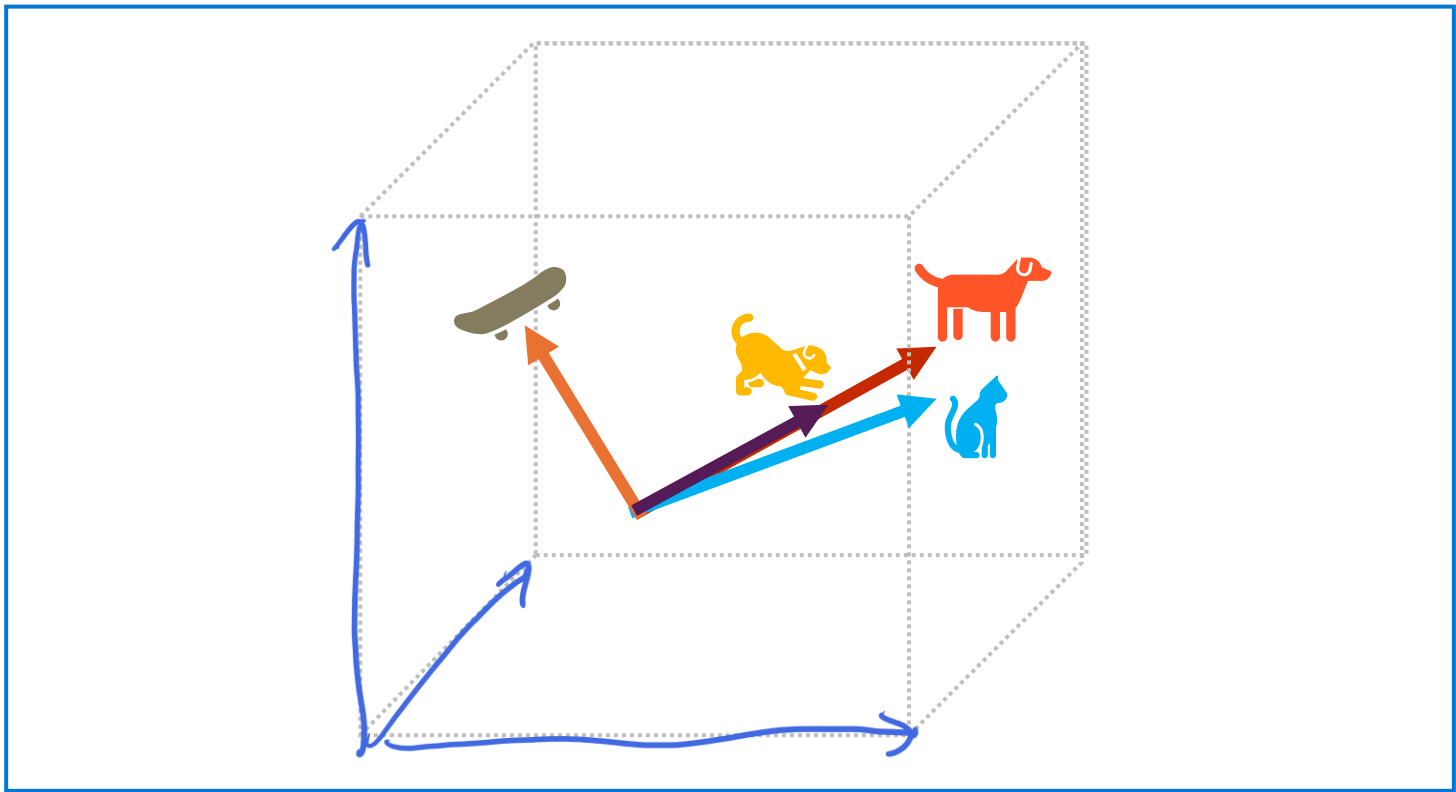
- The sentence is now represented with the tokens: *[1 2 3 4 5 6 7 3 8]*.
- Note "a" is tokenized as 3 only once.
- Similarly, the sentence "I heard a cat" could be represented with the tokens *[1 2 3 8]*.

*Token*





# Language models – embeddings

## Step two: embeddings

The semantic relationship between tokens is encoded in vectors, known as embeddings.



Vector

	Token	Word	Embedding
	4	<u>dog</u>	[10,3,2]
	8	cat	[10,3,1]
	9	puppy	[5,2,1]
	10	skateboard	[-3, 3, 2]



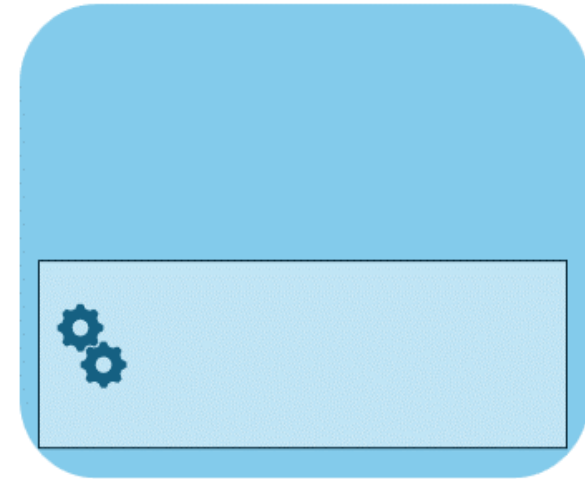
# Language models – attention

## Step three: attention

Capture the strength of relationships between tokens using the attention technique.

### Example:

- Goal: predict token after "**dog**."
- Represent "**I heard a dog**" as vectors.
- Assign "**heard**" and "**dog**" more weight.
- Several possible tokens can come after dog.
- The most probable token is added to the sequence, in this case "**bark**."



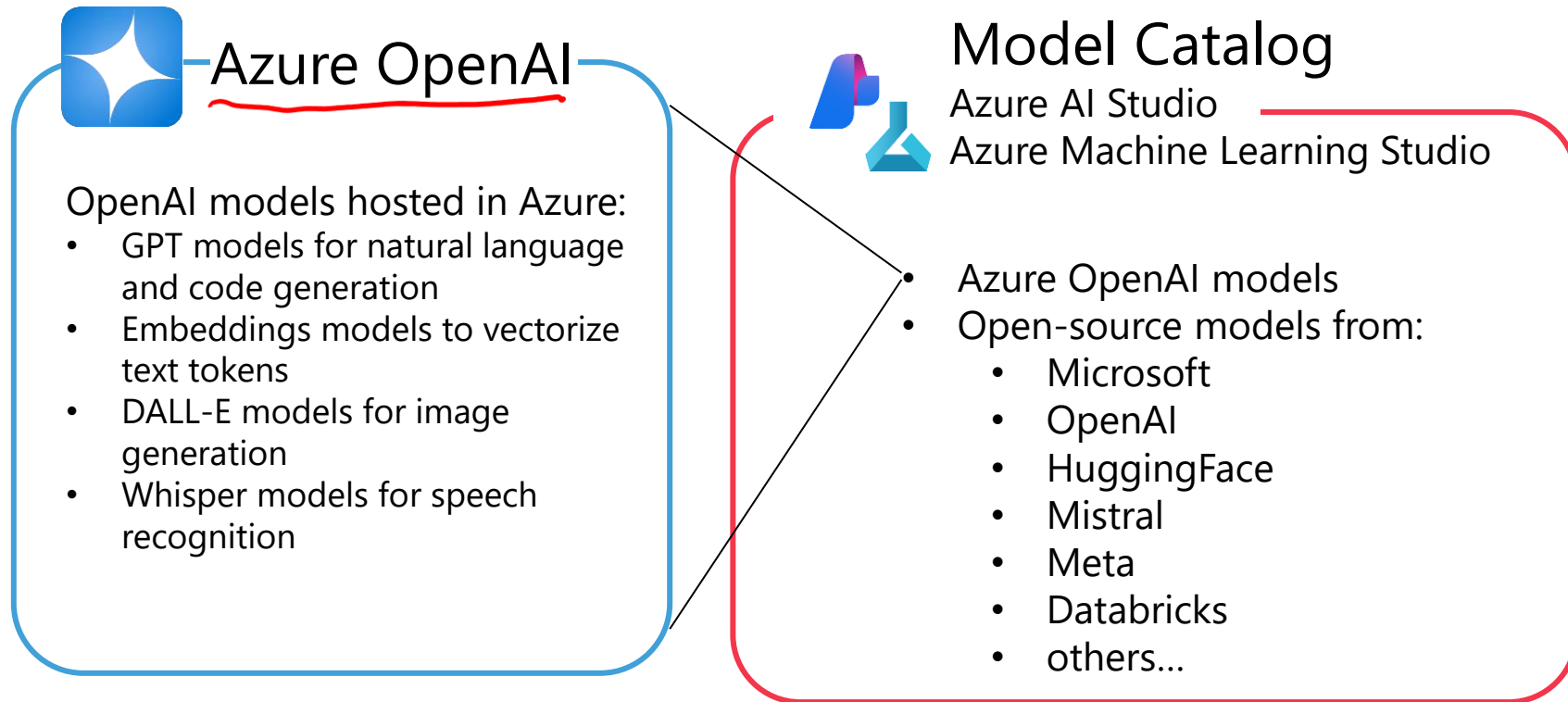
# Language models

## Foundation models

~~X Azure OpenAI Studio~~

↳ Azure AI Studio

- You can train your own language model from scratch, but..
- Most generative AI solutions use a *foundation* model (optionally fine-tuned with your own data)



# Language models

## Large and small language models

### Large Language Models (LLMs)

- Trained with large volumes of general text data
- Many billions of parameters
- Comprehensive language generation capabilities in multiple contexts
- Large size can impact performance and portability
- Time-consuming (and expensive) to fine-tune with your own training data

### Small Language Models (SLMs)

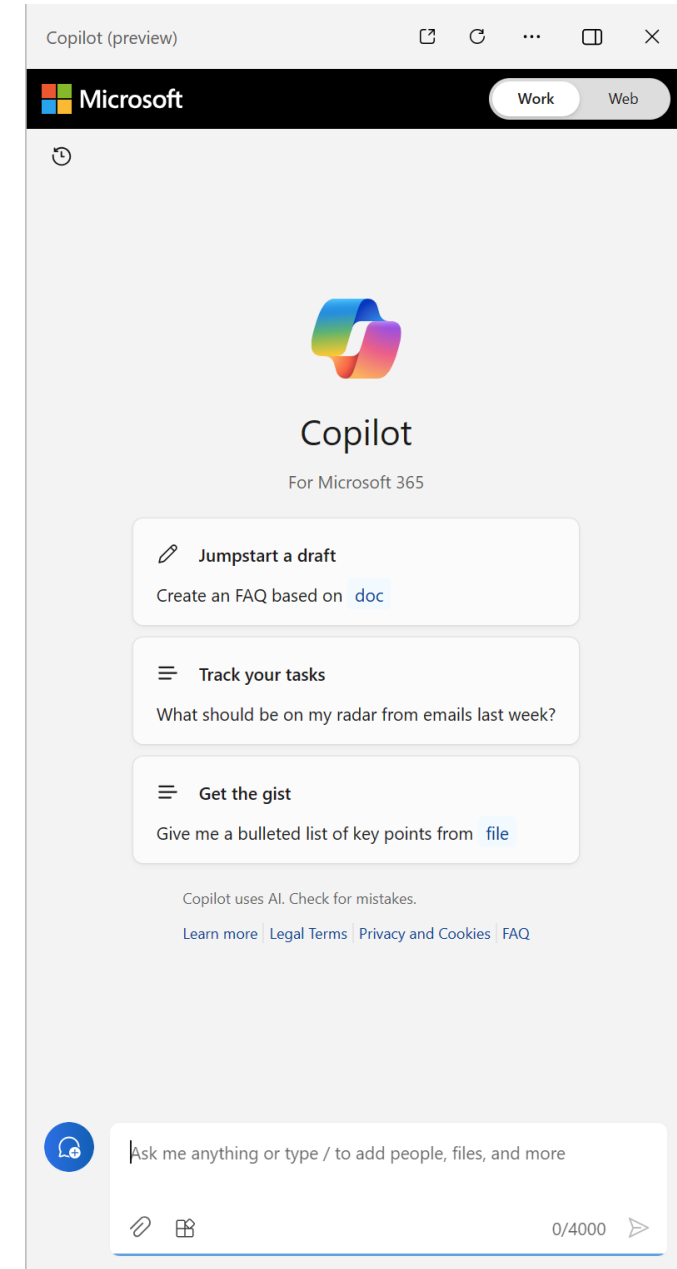
- Trained with focused text data
- Fewer parameters
- Focused language generation capabilities in specialized contexts
- Fast and portable
- Faster (and less expensive) to fine-tune with your own training data
- Examples include:
  - Microsoft Phi-2

# Copilots

## What are copilots?

Generative AI powered assistants integrated into applications, often as chat interfaces

- Copilots provide contextual assistance with everyday tasks; typically using a language model to interpret and generate natural language
- Business users can use copilots to boost their productivity and creativity with AI-generated content and automation of tasks
- Developers can:
  - Extend copilots by creating plug-ins that integrate them into business processes and data
  - Create custom copilots to build generative AI capabilities into apps and services



# Copilots

## Levels of copilot adoption

1

### 1. Unlock productivity across your business with Microsoft Copilot

- Use copilot to optimize the way you work in the tools you use every day
- Empower employees to be more creative
- Spend less time on mundane tasks and more on high impact activities

2

### 2. Extend Microsoft Copilot for your business with *plug-ins*

- Integrate business-specific tasks and processes into Microsoft 365
- Get actionable insights from your own data and systems in productivity apps
- Build 3<sup>rd</sup>-party plug-ins so your customers can use your services in Microsoft 365

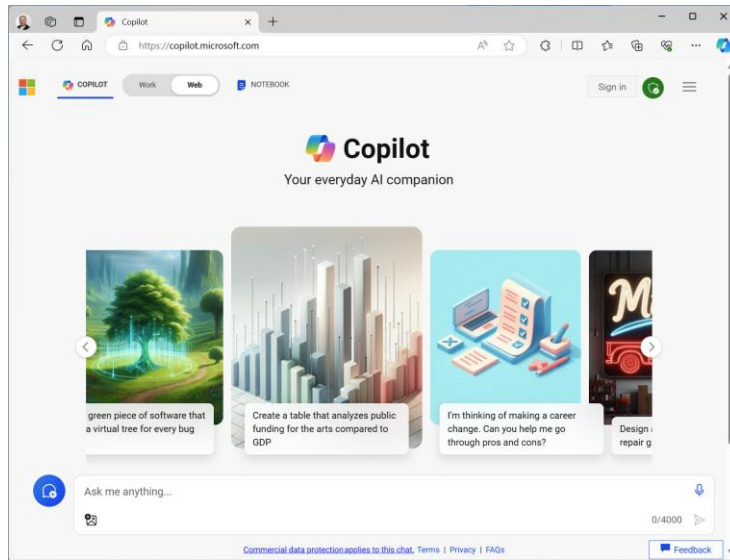
3

### 3. Build custom copilots of your own

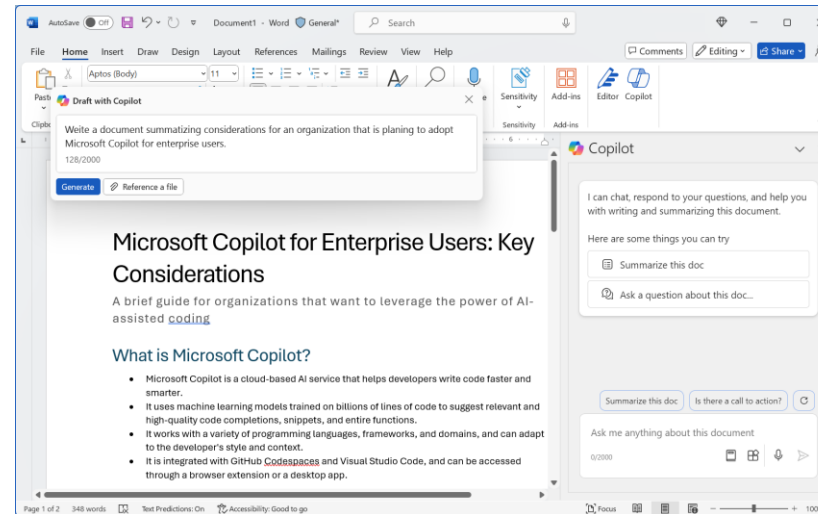
- Integrate generative AI into custom workloads and apps
- Create compelling customer experiences and commercial products
- Have full control over design and development

# Microsoft Copilot Examples

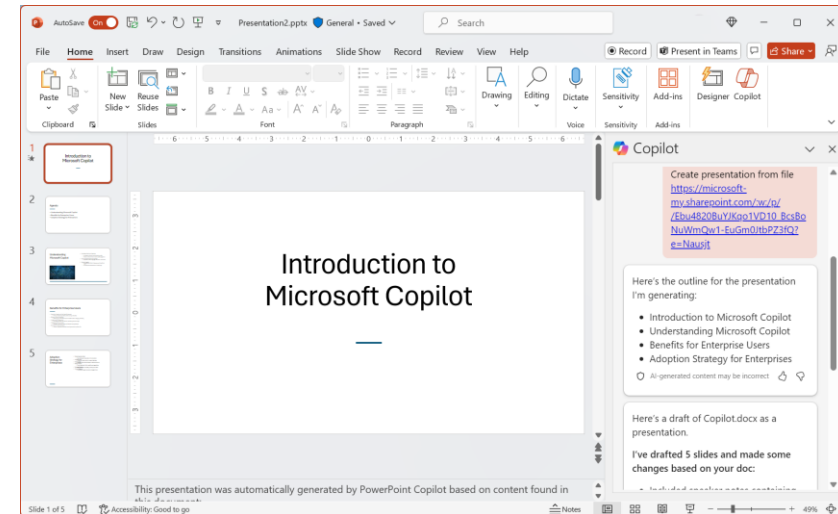
## Generative AI to support information communication



**AI on the web** – accessible to all:  
Use Copilot to answer questions,  
create content, and search the web at  
*<https://copilot.microsoft.com>*

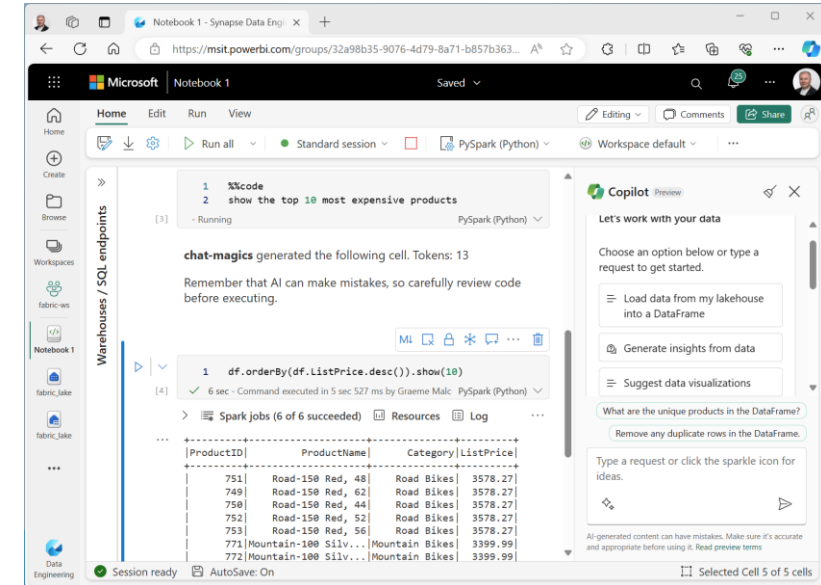
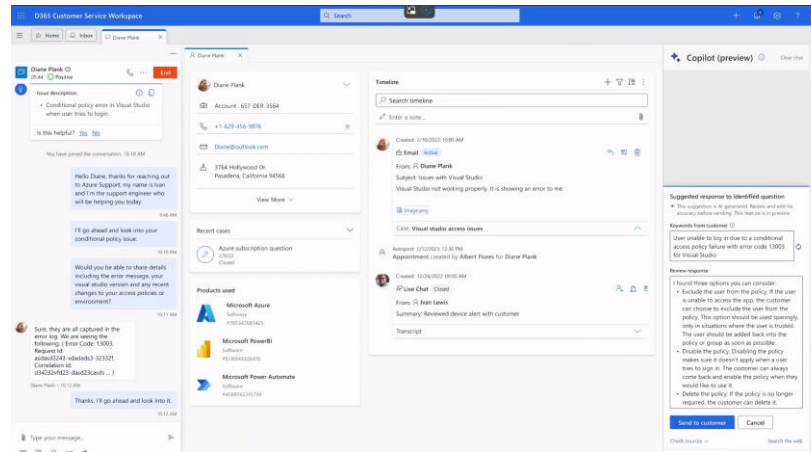
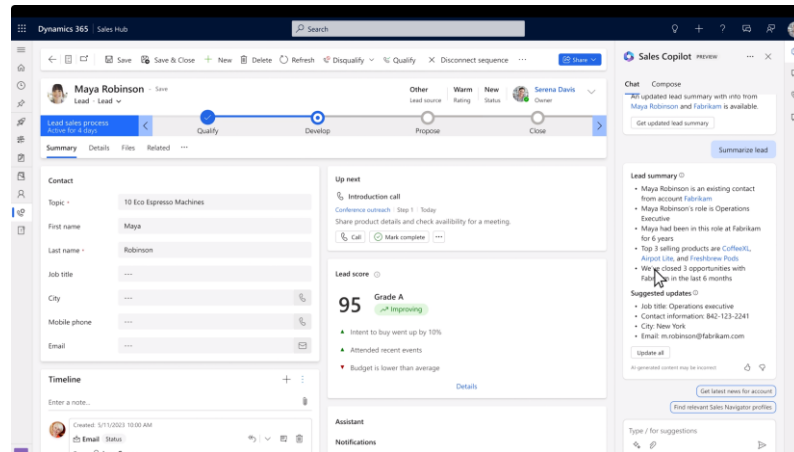


**AI in Microsoft 365** – accessible with subscription:  
Write, edit, and analyze documents in Word, create and enhance presentations in PowerPoint, get support with communication on Outlook and Teams, and more.



# Microsoft Copilot Examples

## Generative AI to support business processes



### AI in Microsoft Dynamics 365:

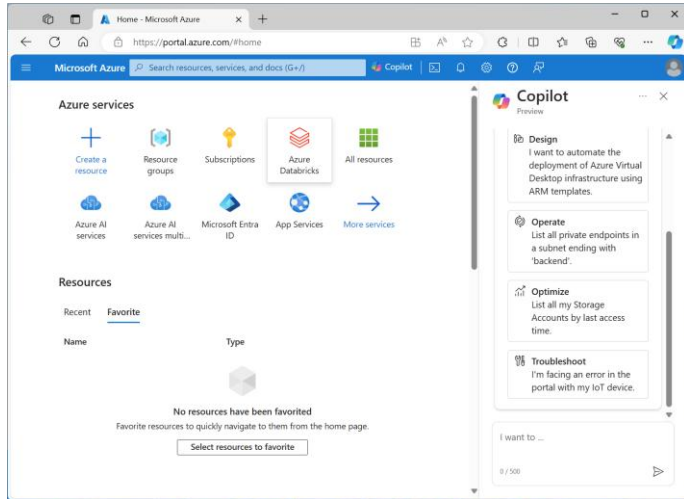
For instance, use Copilot in Microsoft Dynamics 365 Sales to find relevant customer information, qualify leads, and prepare proposals. In another instance, use Copilot in Dynamics 365 Customer Service to quickly research and resolve customer issues.

### AI in Microsoft Fabric:

Use Copilot in Microsoft Fabric to write the code needed to analyze, manipulate, and visualize your data.

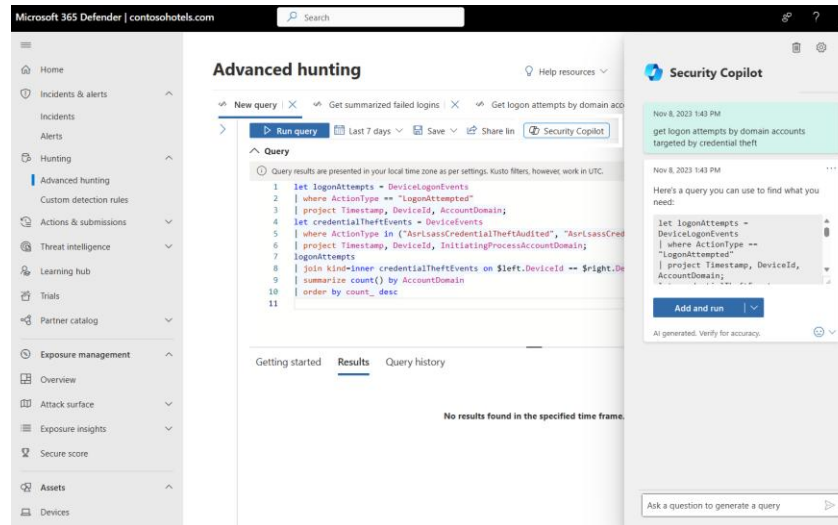
# Microsoft Copilot Examples

Generative AI to support IT infrastructure, security, and software development



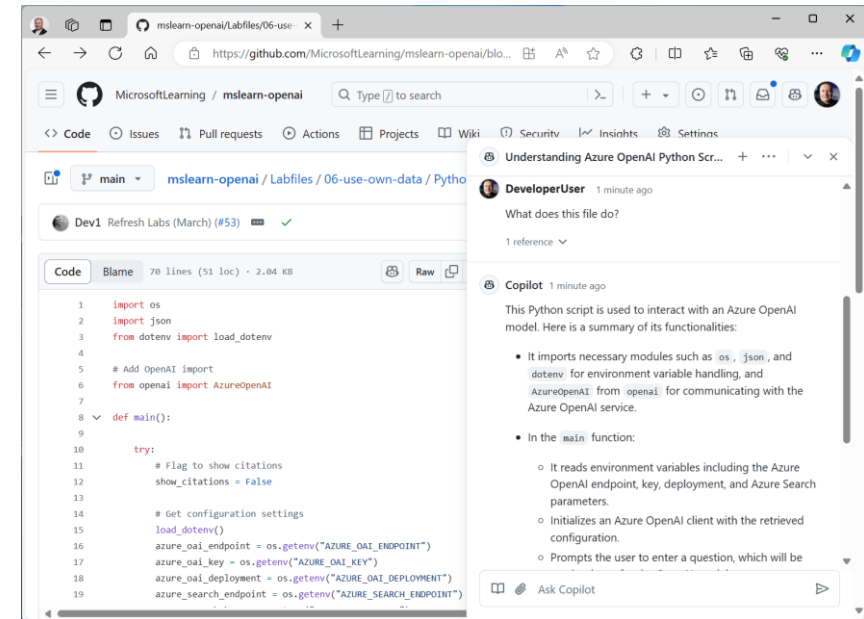
## AI in Microsoft Azure:

Use Copilot in the Azure portal to design, operate, optimize, and troubleshoot cloud infrastructure.



## AI in Microsoft Copilot for Security:

Use Microsoft Copilot for Security as a standalone security portal or embedded in security tools to assess and respond to security threats.

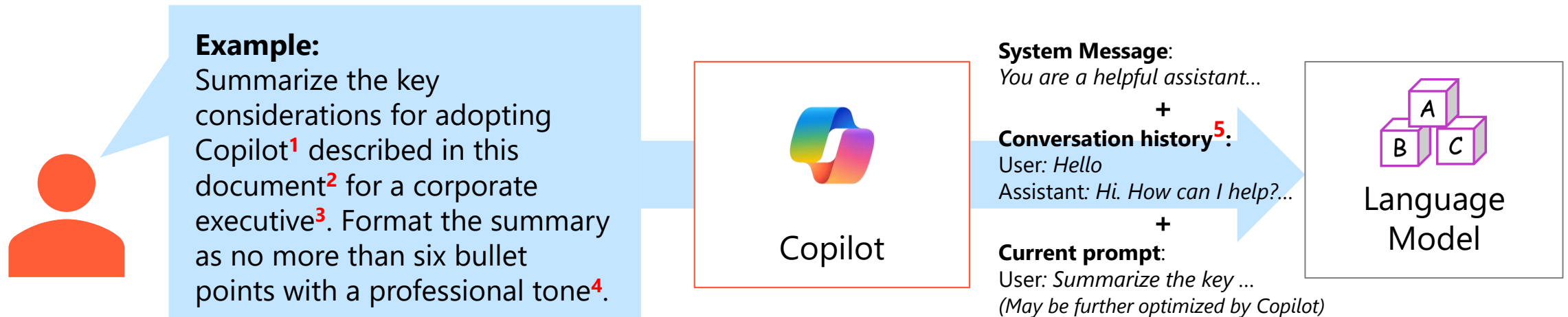


## AI in GitHub Copilot:

Use GitHub Copilot to maximize developer productivity by analyzing and explaining code, adding code documentation, generating new code based on natural language prompts.



# Considerations for copilot prompts

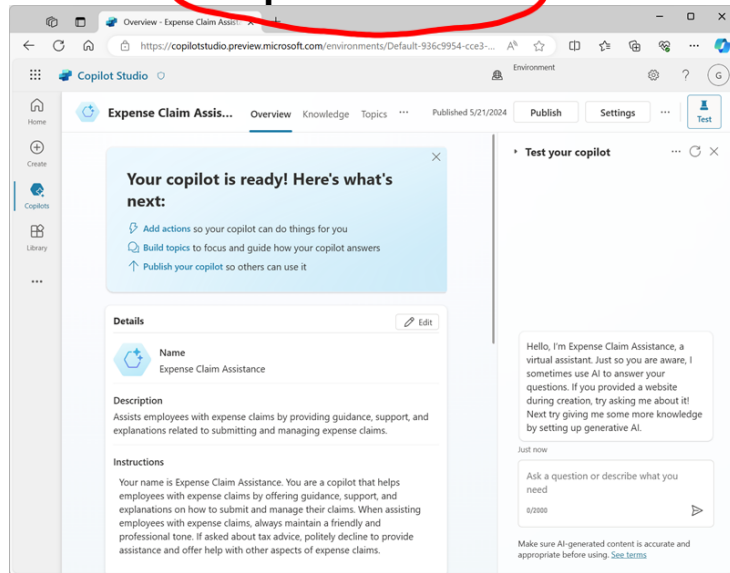


1. Start with a specific goal for what you want the copilot to do
2. Provide a source to *ground* the response in a specific scope of information
3. Add context to maximize response appropriateness and relevance
4. Set clear expectations for the response
5. Iterate based on previous prompts and responses to refine the result

# Extending and developing copilots

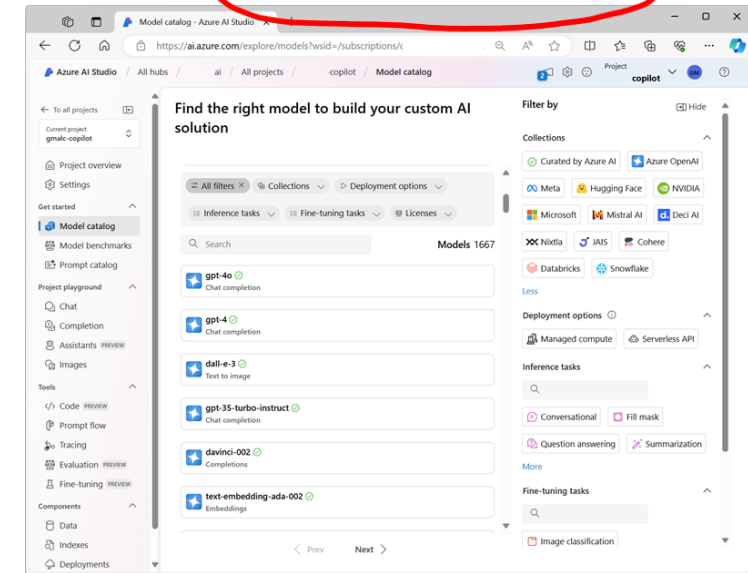
## Choosing a copilot development tool

### Copilot Studio



- Low-code development of copilots and plug-ins using Azure OpenAI models for generative AI
- Fully managed, hosted SaaS service
- Dialog and conversational orchestration
- Built-in analytics with security and governance controls
- Deploy to common chat channels like web, apps, social channels, and Teams

### Azure AI Studio



- Pro-code development with full catalog of models and fine-tuning capabilities
- PaaS services with full control over cloud infrastructure
- Prompt and model orchestration
- Evaluations engine to test performance, reliability, scalability, and responsible AI safety
- Deploy as an endpoint in Azure for use in custom apps and services

# Exercise: Explore generative AI with Microsoft Copilot

**Estimated time:**  
40 minutes



Use the hosted environment and credentials provided for this exercise

The instructions are also available online:

- Microsoft 365 version:  
<https://go.microsoft.com/fwlink/?linkid=2270745>
- Microsoft Edge version:  
<https://go.microsoft.com/fwlink/?linkid=2249955>

# Knowledge check



## 1 What are Large Language Models?

- ☐ Models that detect additional meaning in paragraphs of text.
- ☐ Lists of words and code that computers use to generate text.
- ☐ Models that use deep learning to process and understand natural language on a massive scale.

## 2 Which Microsoft Copilot should a customer support agent use to research and resolve a support issue?

- ☐ Microsoft Copilot for Microsoft Edge
- ☐ Microsoft Copilot for Dynamics 365 Customer Service
- ☐ Microsoft Copilot for Security

## 3 Which tool should a professional developer use to build a custom copilot and deploy it as a service endpoint in Azure?

- ☐ Microsoft Copilot for Azure
- ☐ Microsoft Copilot Studio
- ☐ Microsoft Azure AI Studio

# Knowledge check



## 1 What are Large Language Models?

- ☐ Models that detect additional meaning in paragraphs of text.
- ☐ Lists of words and code that computers use to generate text.
- ☒ Models that use deep learning to process and understand natural language on a massive scale.

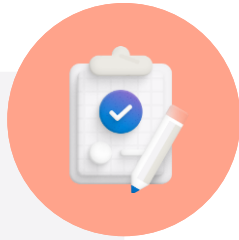
## 2 Which Microsoft Copilot should a customer support agent use to research and resolve a support issue?

- ☐ Microsoft Copilot for Microsoft Edge
- ☐ Microsoft Copilot for Dynamics 365 Customer Service
- ☒ Microsoft Copilot for Security

## 3 Which tool should a professional developer use to build a custom copilot and deploy it as a service endpoint in Azure?

- ☐ Microsoft Copilot for Azure
- ☐ Microsoft Copilot Studio
- ☒ Microsoft Azure AI Studio

# Summary



## Key learning points:

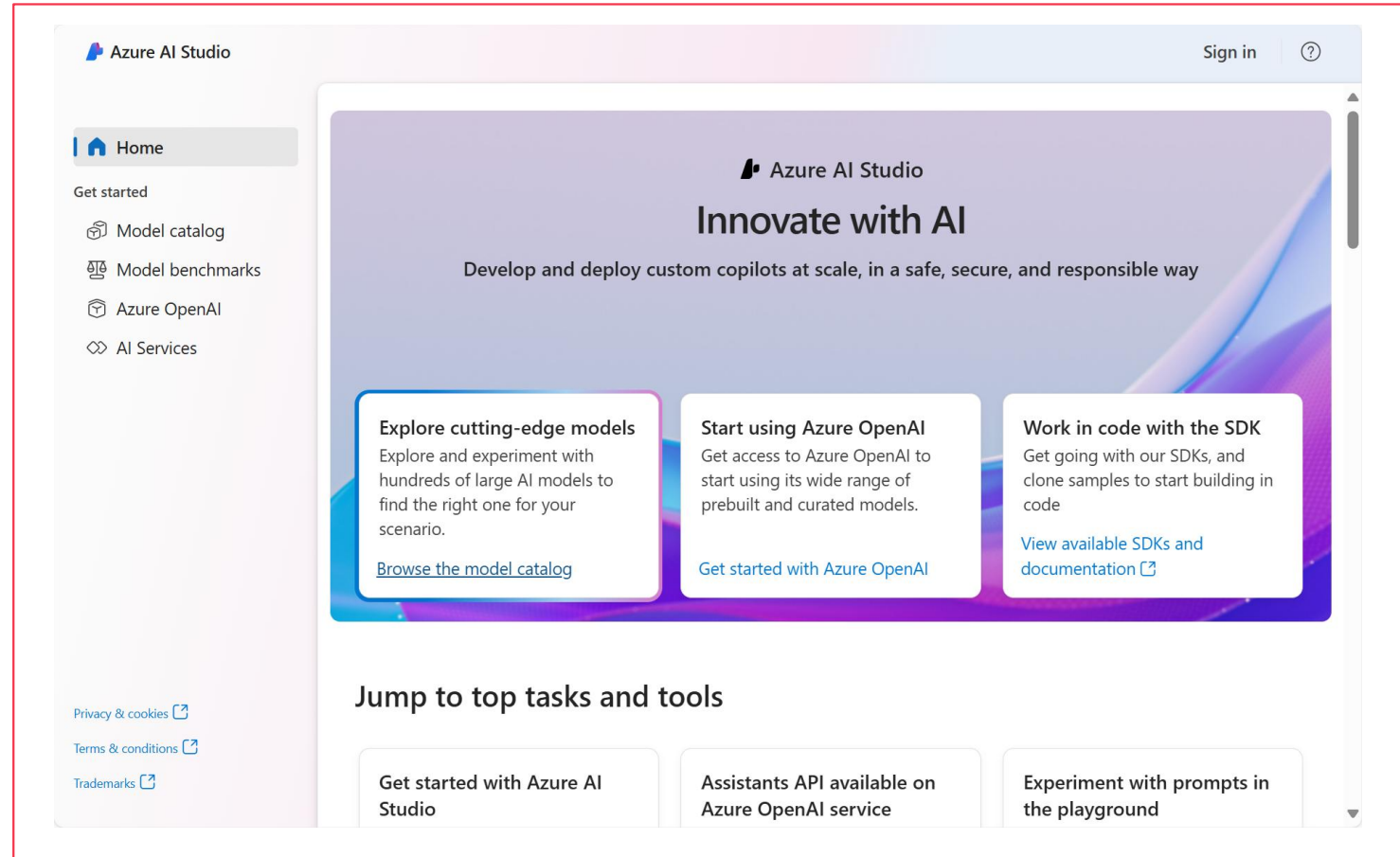
- Understand generative AI's place in the development of artificial intelligence.
- Understand large language models and their role in intelligent applications.
- Describe examples of copilots and good prompts.

# Introduction to the Azure AI Studio



# What is Azure AI Studio?

- A Platform as a Service (PaaS) for developers
- Provides a collaborative pro-code environment with enterprise-grade security for AI innovation
- Use to explore, build, test, and deploy AI solutions that are reliable, scalable, and adhere to responsible AI safety





# What solutions does Azure AI Studio bring together?



**Azure OpenAI service:**  
Microsoft's cloud solution for deploying, customizing, and hosting language models.



**Azure Machine Learning:** Microsoft's cloud solution for training and deploying models and managing machine learning operations (MLOps).



**Azure AI services** encompass Microsoft's cloud solutions for creating cutting-edge, market-ready, and responsible applications with prebuilt and customizable APIs and models.

# How does Azure AI Studio work?

**AI hub resources** allow you to collaborate with others as a team. **AI projects** are contained within a hub and allow for AI customization and *orchestration* (process automation).

## AI hub resource



Create and manage connections



Create and manage compute



Security setup and governance

## AI projects



Deploy and test



Augment and build



Evaluate and manage

# Use Azure AI Studio to:

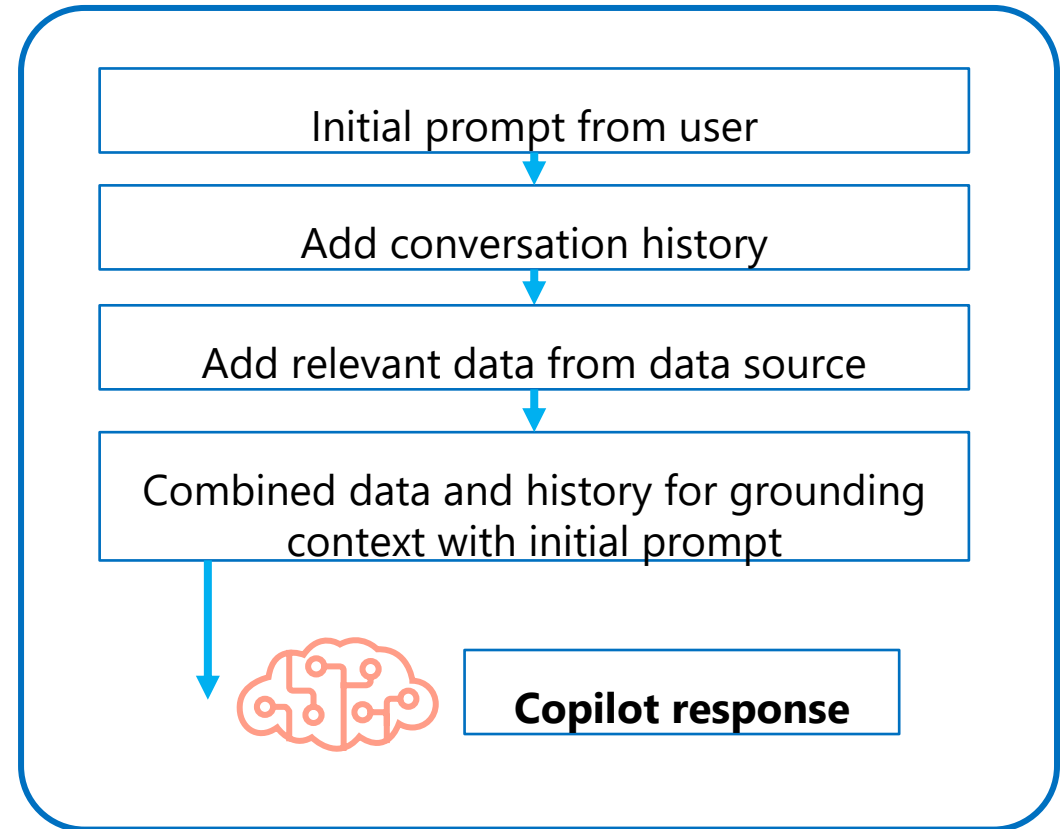
- **Create and manage AI projects:** provides a centralized hub for all your AI projects
- **Develop generative AI applications:** create applications that can generate content or build your own prompt flow
- **Explore available AI models:** Experiment with various AI models in Azure AI Studio's model catalog
- **Leverage Retrieval Augmented Generation (RAG):** enhance the quality and relevance of generated content.
- **Monitor and evaluate AI models:** ensure AI models meet performance metrics
- **Integrate with Azure services:** work seamlessly with other Azure services
- **Build responsibly:** ensure that your applications adhere to ethical standards and best practices

# What are prompt flows?

**Prompt:** input sent to a model, consisting of the user input, system message, and any examples.

**Flow:** an executable instruction set that can implement AI logic such as data retrieval, calling language models, executing code.

**Prompt flow:** a feature that can be used to generate, customize, or run a flow.



# What capabilities does the model catalog support?

Azure AI Studio provides an extensive model catalog that allows you to discover, evaluate, fine-tune, and deploy models provided by Microsoft and Non-Microsoft companies.

Use these models to build generative AI applications that have:

Natural language generation capabilities

Code generation capabilities

Image generation capabilities, and *more*.

# Knowledge check



**1** Which of the following best describes Azure AI Studio?

- ☐ An online marketplace where you can buy and sell AI models.
- ☐ A collaborative development environment for AI projects on Azure.
- ☐ A graphics editing application that uses AI to generate images.

**2** How can you enable a colleague to collaborate with you on an AI project?

- ☐ Add them as a member in the appropriate role to your Azure AI hub.
- ☐ Deploy a model as a web app and configure authentication for the web app service.
- ☐ Tell the colleague to create their own Azure AI hub in the same Azure subscription as yours.

# Knowledge check



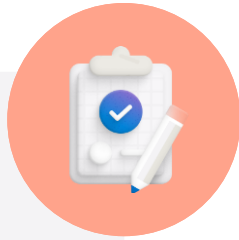
**1** Which of the following best describes Azure AI Studio?

- ☐ An online marketplace where you can buy and sell AI models.
- ☒ A collaborative development environment for AI projects on Azure.
- ☐ A graphics editing application that uses AI to generate images.

**2** How can you enable a colleague to collaborate with you on an AI project?

- ☒ Add them as a member in the appropriate role to your Azure AI hub.
- ☐ Deploy a model as a web app and configure authentication for the web app service.
- ☐ Tell the colleague to create their own Azure AI hub in the same Azure subscription as yours.

# Summary



## Key learning points

- Azure AI Studio provides a single tool for AI development with multiple Azure AI services.
- An Azure AI resource defines a collaborative workspace for AI development with Azure AI Studio.
- An Azure AI project provides a shared collection of assets and code for a particular AI solution.

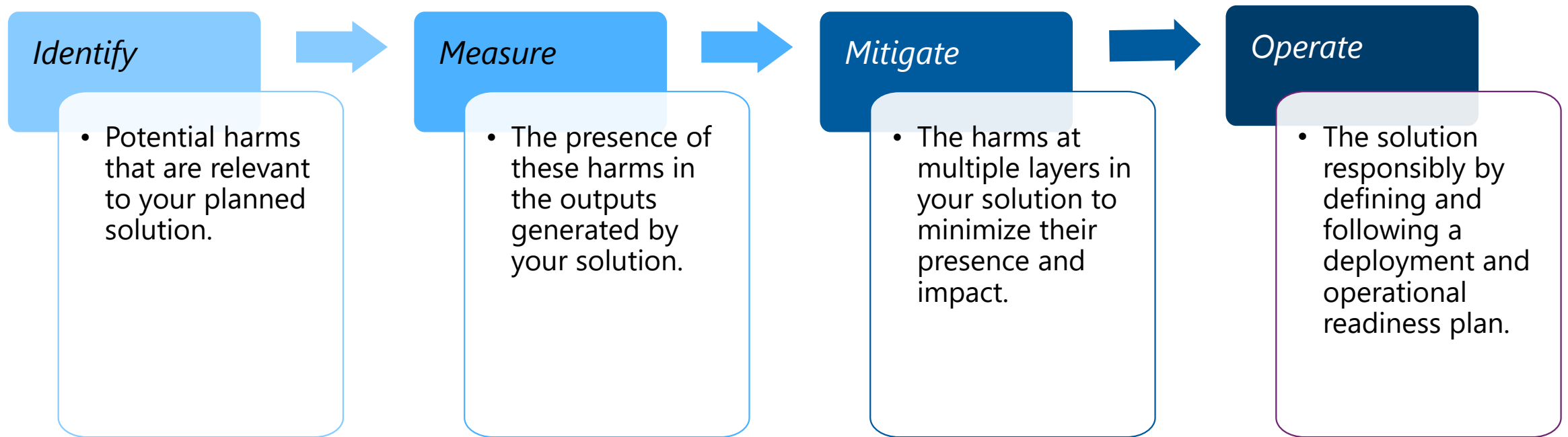


Implement responsible  
generative AI in AI Studio



# Plan a responsible generative AI solution

Four stage process to develop and implement a plan for responsible AI are:



# Azure AI Content Safety

Feature	Functionality
Prompt shields	Scans for the risk of user input attacks on language models
Groundedness detection	Detects if text responses are grounded in a user's source content
Protected material detection	Scans for known copyrighted content
Custom categories	Define custom categories for any new or emerging patterns

# Demo: Explore Azure AI Content Safety



In this demo, you'll see **Azure AI**'s content filters in action.

1. Follow along on the exercise page at: <https://aka.ms/explore-content-filters>

# Knowledge check



## 1 What does groundedness refer to in the context of generative AI?

- ☐ To make a legal case that indemnifies you from responsibility for harms caused by the solution
- ☐ To document the purpose, expected use, and potential harms for the solution
- ☐ To evaluate the cost of cloud services required to implement your solution

## 2 What capability of Azure AI Studio helps mitigate harmful content generation at the Safety System level?

- ☐ DALL-E model support
- ☐ Fine-tuning
- ☐ Content filters

## 3 Why should you consider a multiple-stage process for your generative AI solution?

- ☐ To enable you to gather feedback and identify issues before releasing the solution more broadly
- ☐ To eliminate the need to identify, measure, and mitigate potential harms
- ☐ To enable you to charge more for the solution

# Knowledge check



## 1 What does groundedness refer to in the context of generative AI?

- ☐ To make a legal case that indemnifies you from responsibility for harms caused by the solution
- ☒ To document the purpose, expected use, and potential harms for the solution
- ☐ To evaluate the cost of cloud services required to implement your solution

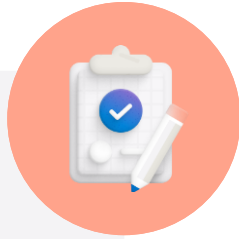
## 2 What capability of Azure AI Studio helps mitigate harmful content generation at the Safety System level?

- ☐ DALL-E model support
- ☐ Fine-tuning
- ☒ Content filters

## 3 Why should you consider a multiple-stage process for your generative AI solution?

- ☒ To enable you to gather feedback and identify issues before releasing the solution more broadly
- ☐ To eliminate the need to identify, measure, and mitigate potential harms
- ☐ To enable you to charge more for the solution

# Summary



## Key learning points

- Plan a responsible AI generative AI solution
- Operate a responsible generative AI solution

