

Integration von Azure AI Services in AKS

Wolfgang Ladermann & Mustafa Emal
Microsoft Cloud Solution Architects
Digital & Application Innovation



Agenda

Intro to intelligent apps?

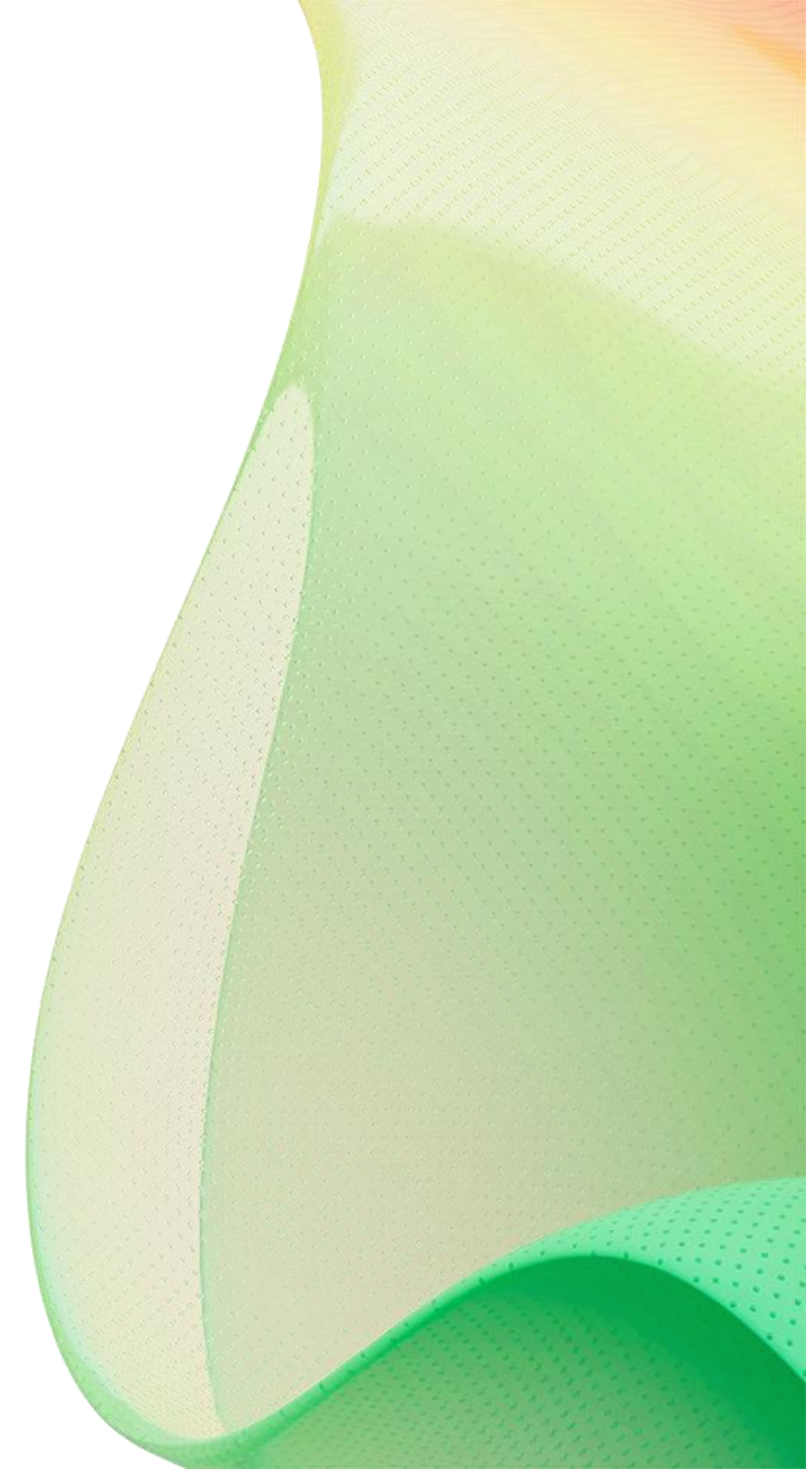
AKS Copilot

Azure OpenAI

Vector Databases

Takeaways

What are intelligent apps?



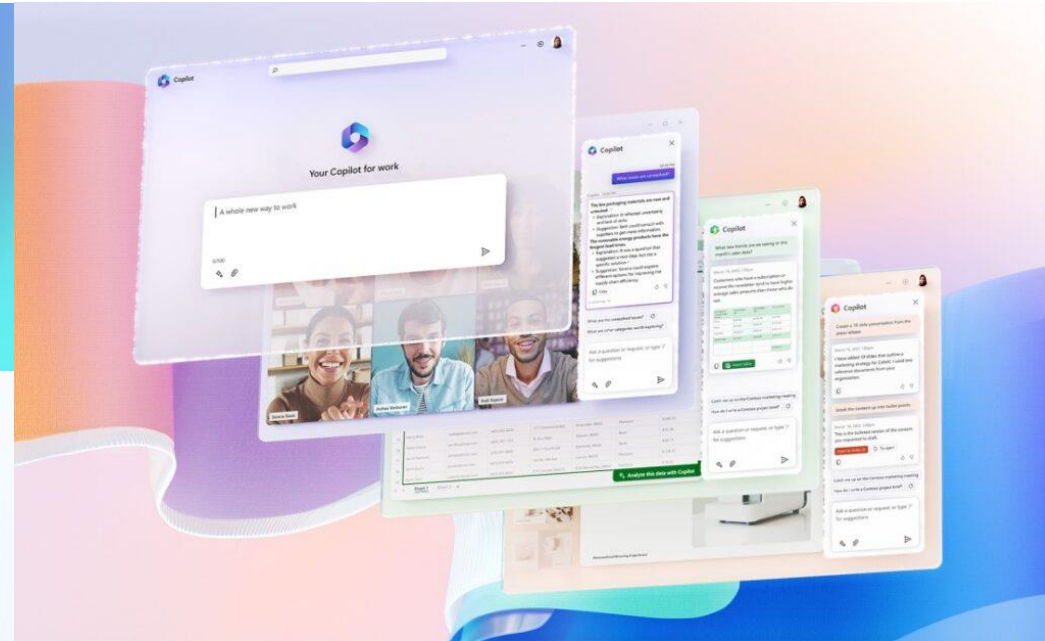
Chatbots and Copilots abound

AI is everywhere, AI in everything



Announcing

Microsoft Security Copilot



Create an image of 3D render of a polar bear

Sure, I'll use Image Creator to draw that for you.

Image Creator

Persistent Chat

Take a ferry ride

Learn more: 1 of 20

1 postcardssentwith.com

Export/Share

How do I pick the best dog breed for me?

Searching for 'how to pick the best dog breed'

Generating answers for you...

Picking the best dog breed for you depends on several factors, such as your lifestyle, your energy level, your house size, and your preferences.

Stop responding...

Ask me anything

Bing Chat

Discover

Bing Open Preview

Introducing the new Bing. Be the one of the first to enjoy it.

Start chatting

Open Preview

The New Bing: AI-powered search with chat

Chat History

- Gluten-free and vegan recipes
- Little-known baseball facts
- 25 blog ideas for gardening
- Ball trip planning
- Best upcoming RPGs

Actions

- Searching for: New restaurants in NYC
- Using: OpenTable
- Generating answers for you...

Type message

More Visual Answers

Pollination is carried out by nectar feeding bats, birds and insects.

Learn more

Compose

Write about an email to congratulate my colleague on her promotion

5/3/2000

Tone

Professional Casual Enthusiastic

Informational Funny

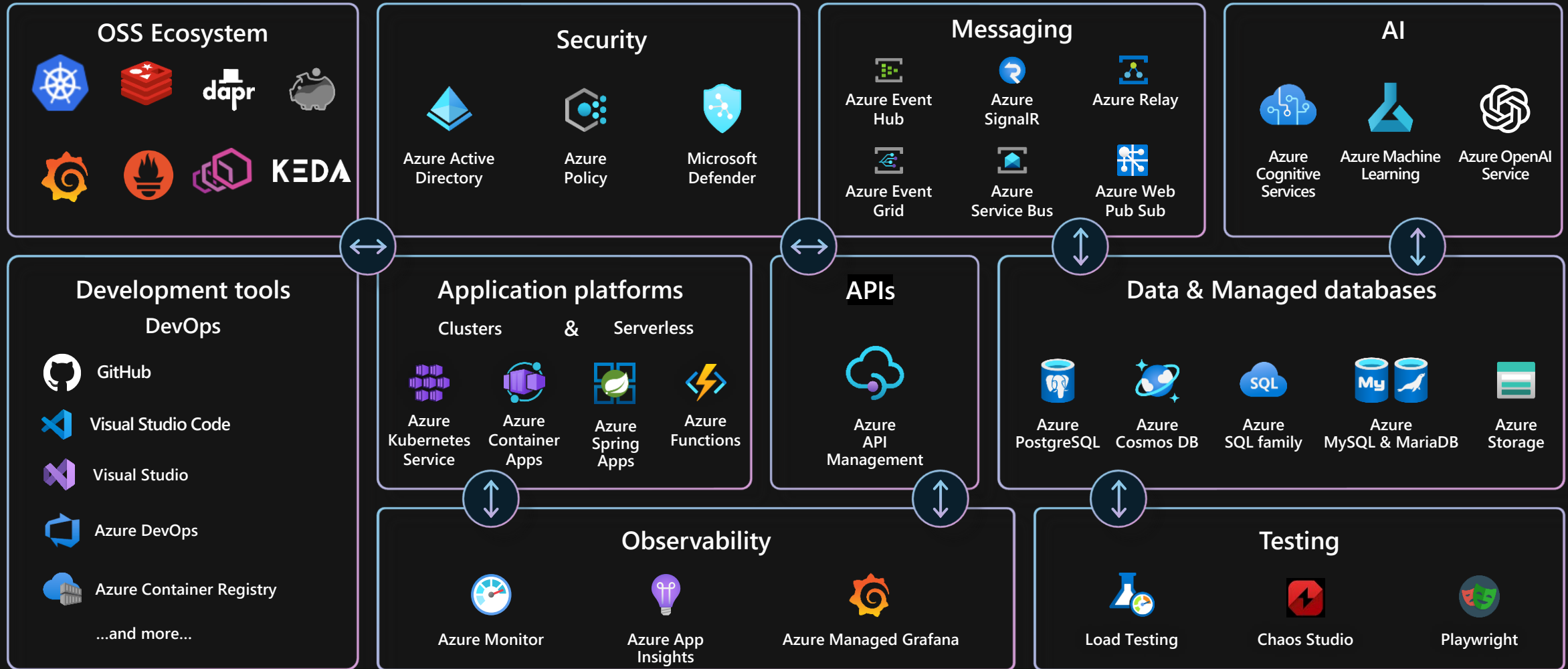
More Languages for Image Creator

Visual Graph

Name	Salary (Dollar)
Stephen Curry, PG	48,750,000
Russell Westbrook, PG	47,000,000
LeBron James, SF	46,400,000



Building cloud-native apps with AI on Azure



Cloud operations anywhere



AKS Copilot: AI-powered assistant for Kubernetes

Streamline your operations

Reduce your skilling efforts

Empower your developers

Available free for all clusters



Home > build-ai
build-ai | Workloads
Kubernetes service

Search

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems
- Microsoft Defender for Cloud
- Kubernetes resources**
- Namespaces
- Workloads**
- Services and ingresses
- Storage
- Configuration
- Settings**
- Node pools
- Cluster configuration
- Networking
- Extensions + applications
- Backup (preview)
- Open Service Mesh
- GitOps
- Automated deployments

+ Create Delete Refresh Show labels Give feedback Open AKS Copilot

Deployments Pods Replica sets Stateful sets Daemon sets Jobs Cron jobs

Filter by deployment name: Enter the full deployment name
Filter by namespace: All namespaces
Add label filter

<input type="checkbox"/>	Name	Namespace	Ready	Up-to-date	Available	Age ↓
<input type="checkbox"/>	coredns	kube-system	✓ 2/2	2	2	42 minutes
<input type="checkbox"/>	coredns-autoscaler	kube-system	✓ 1/1	1	1	42 minutes
<input type="checkbox"/>	konnnectivity-agent	kube-system	✓ 2/2	2	2	42 minutes
<input type="checkbox"/>	metrics-server	kube-system	✓ 2/2	2	2	42 minutes
<input type="checkbox"/>	tigera-operator	tigera-operator	✓ 1/1	1	1	42 minutes
<input type="checkbox"/>	calico-typha	calico-system	✓ 3/3	3	3	40 minutes
<input type="checkbox"/>	calico-kube-controllers	calico-system	✓ 1/1	1	1	40 minutes
<input type="checkbox"/>	ama-logs-rs	kube-system	✓ 1/1	1	1	34 minutes

Search

- + Create
- Connect
- Start
- Stop
- Delete
- Refresh
- Give feedback
- Open in mobile
- Open AKS Copilot

Essentials View Cost JSON View

Resource group : aks-amanda-rg	Kubernetes version : 1.25.6
Status : Succeeded (Running)	API server address : build-ai-dns-eyyl6269.hcp.westus2.azmk8s.io
Location : West US 2	Network type (plugin) : Kubenet
Subscription : ComputePM LibrarySub - 040	Node pools : 2 node pools
Subscription ID : 821664e1-b43e-4312-a018-4d2d284c2b9c	
Tags (edit) : Click here to add tags	

- Get started
- Properties**
- Monitoring
- Capabilities (3)
- Recommendations
- Tutorials

Kubernetes services	
Encryption type	Encryption at-rest with a platform-managed key
Virtual node pools	Not enabled
Node pools	
Node pools	2 node pools
Kubernetes versions	1.25.6
Node sizes	Standard_D8ds_v5, Standard_DS2_v2
Configuration	
Kubernetes version	1.25.6
Auto Upgrade Type	Disabled
Authentication and Authorization	Local accounts with Kubernetes RBAC
Local accounts	Enabled

Networking	
API server address	build-ai-dns-eyyl6269.hcp.westus2.azmk8s.io
Network type (plugin)	Kubenet
Pod CIDR	10.244.0.0/16
Service CIDR	10.0.0.0/16
DNS service IP	10.0.0.10
Docker bridge CIDR	172.17.0.1/16
Network Policy	Calico
Load balancer	Standard
HTTP application routing	Not enabled
Private cluster	Not enabled
Authorized IP ranges	Not enabled
Application Gateway ingress controller	Not enabled

Extensions + applications	Integrations
	Container insights Enabled

Azure OpenAI | Top 4 Capabilities & Use Cases



Content Generation

Call Center Analytics: Automatically generate responses to customer inquiries

Generate personalised UI for your website



Summarization

Call Center Analytics: Summary of customer support conversation logs

Subject Matter Expert Document Summarization (e.g. financial reporting, analyst articles)

Social Media Trends Summarization



Code generation

Convert Natural Language to SQL (or vice versa) for telemetry data

Convert Natural Language to Query Proprietary Data Models

Code Documentation



Semantic Search

Search reviews for a specific product / service

Information Discovery and Knowledge Mining

Examples of Multiple Model Use Cases

End to End Call Center Analytics: Classification, Sentiment, Entity Extraction, Summarization and Email Generation

Customer 360: Hyper-personalisation using timely Summarization of customer queries & trends, Search, and Content Generation

Business Process Automation: Search through structured & unstructured documentation, Generate Code to query data models, Content Generation

Azure OpenAI Service content creation by API



Prompt

Write a tagline for an ice cream shop.

Response

We serve up smiles with every scoop!

Prompt

```
Table customers, columns =  
[CustomerId, FirstName,  
LastName, Company, Address,  
City, State, Country,  
PostalCode]
```

Create an SQL query for all customers in Texas named Jane
query =

Response

```
SELECT *  
FROM customers  
WHERE State = 'TX' AND  
FirstName = 'Jane'
```

Prompt

A ball of fire with vibrant colors to show the speed of innovation at our media and entertainment company

Response



Azure OpenAI | Sample Use Cases


Playground

Deployments

text-davinci-002

Examples

Load an example

 Code View

Customer conversation

Hi there I just had a car accident and wanted to report it. OK, I hope you're alright and in safe place to have this conversation.
Yes, I'm fine thank you. Can you please describe to me what happened? I was driving on the M23 and I hit another car. Are you OK? Yeah, I'm just a little shaken up. That's understandable. Can you give me your full name? Sure, it's Sarah Standl. Do you know what caused the accident? I think I might have hit a pothole. OK, where did the accident take place? On the M23 near junction 10. Was anyone else injured? I don't think so. But I'm not sure. OK, well we'll need to do an investigation. Can you give me the other drivers information? Sure, his name is John Radley. And your insurance policy number. OK. Give me a minute. OK, it's 546452.
OK, is there any damage to your car? Yes, Headlights are broken and the airbags went off. Are you going to be able to drive it? I don't think so. I'm going to have to have it towed. Well, we'll need to get it inspected. I'll go ahead and order you a tow van. I'll also start the claim and we'll get everything sorted out. Thank you.

Azure OpenAI Instruction

Extract the following from the conversation:

1. Main reason of the conversation
2. Sentiment of the customer
4. How did the agent handle the conversation?
5. what was the Final outcome of the conversation
6. create a short summary of the conversation

Output

1. The main reason of the conversation was to report a car accident.
2. The sentiment of the customer was shaken up but OK.
4. The agent handled the conversation well and was empathetic.
5. The final outcome of the conversation was that the agent started a claim and ordered a tow truck.
6. The customer was in a car accident and the agent started a claim and ordered a tow truck. The customer was shaken up but OK.

**Example:
Extract
information
from text and
summarize it**

Azure OpenAI | Sample Use Cases

Playground

Deployments

text-davinci-002

Examples

Summarize Text

 Code View

You must extract the following information from the phone conversation below:

1. Call reason (key: reason)
2. Cause of the incident (key: cause)
3. Names of all drivers as an array (key: driver_names)
4. Insurance number (key: insurance_number)
5. Accident location (key: location)
6. Car damages as an array (key: damages)
7. A short, yet detailed summary (key: summary)

Make sure fields 1 to 6 are answered very short, e.g. for location just say the location name. Please answer in JSON machine-readable format, using the keys from above. Format the output as JSON object called "results". Pretty print the JSON and make sure that is properly closed at the end.

**Azure OpenAI
Instruction in English**

Phone conversation:

Hi I just had a car accident and wanted to report it. OK, I hope you're alright, what happened? I was driving on the I-18 and I hit up. That's understandable. Can you give me your full name? Sure, it's Sarah Standl. Do you know what caused the accident? I think I might have hit a pothole. OK, where did the accident take place? On the I-18 freeway. Was anyone else injured? I don't think so. But I'm not sure. OK, well we'll need to do an investigation. Can you give me the other drivers information? Sure, his name is John Radley. And your insurance number. OK. Give me a minute. OK, it's 546452. OK, what type of damages has the car? Headlights are broken and the airbags went off. Are you going to be able to drive it? I don't know. I'm going to have to have it towed. Well, we'll need to get it inspected. I'll go ahead and start the claim and we'll get everything sorted out. Thank you.

**Customer conversation in
English (or other language)**

```
results = {  
  "reason": "car accident",  
  "cause": "hit a pothole",  
  "driver_names": ["Sarah Standl", "John Radley"],  
  "insurance_number": 546452,  
  "location": "I-18 freeway",  
  "damages": ["headlights broken", "airbags went off"],  
  "summary": "Sarah Standl was driving on the I-18 when she hit another car. She thinks she hit a pothole. John Radley was the other driver. There were no injuries but both cars have damage."  
}
```

Output as JSON

Call Center Automation

Example: OpenAI Prompt

Azure OpenAI | Sample Use Cases (cont.)

Playground

Deployments

text-davinci-002



Examples

Load an example



 Code View

Wartungsfenster für den 14. Januar von 14 bis 15 Uhr geplant, erwarten Sie Ausfallzeiten.

Input - German

Generate a formal customer communication in English

- offer help at info@operations.com
- be apologetic where possible
- use 0:00 as time format

Please output in a letter format.

Azure OpenAI
Instruction - English

Dear Customer,

We are writing to inform you that we have scheduled maintenance for January 14th from 14:00 to 15:00. We apologize for any inconvenience this may cause and offer help at info@operations.com should you have any questions.

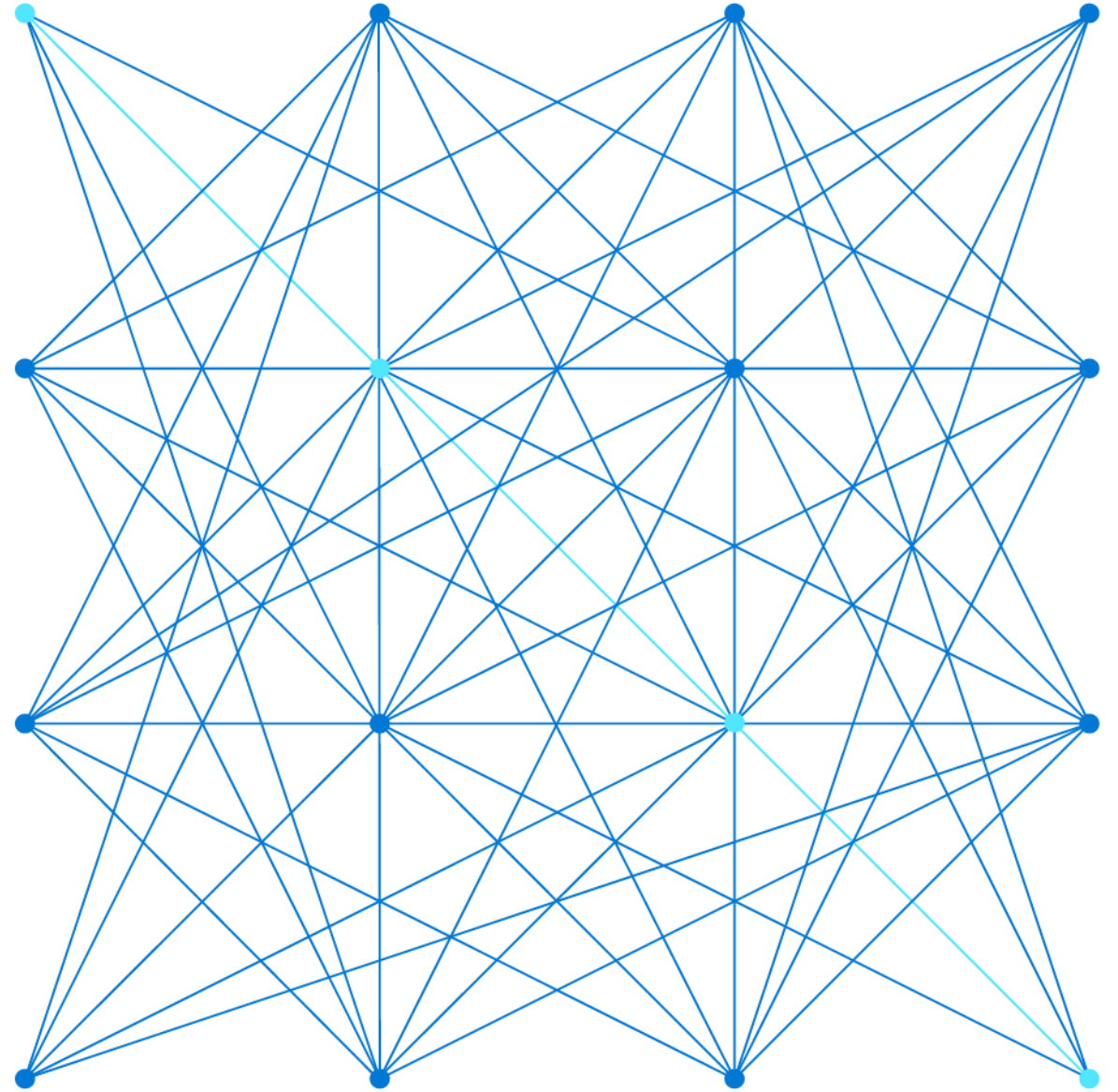
Sincerely,

Your Operations Team

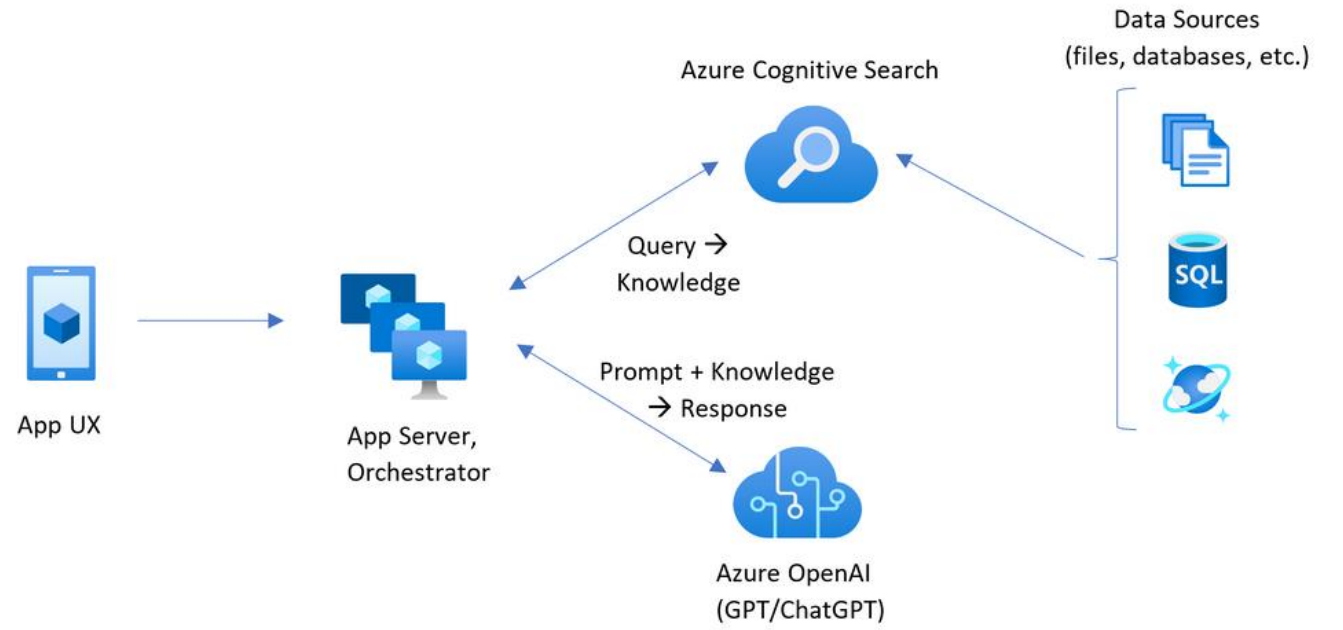
Output - English

Example:
Generate
auto text in
any
language

Demo: Azure OpenAI Studio

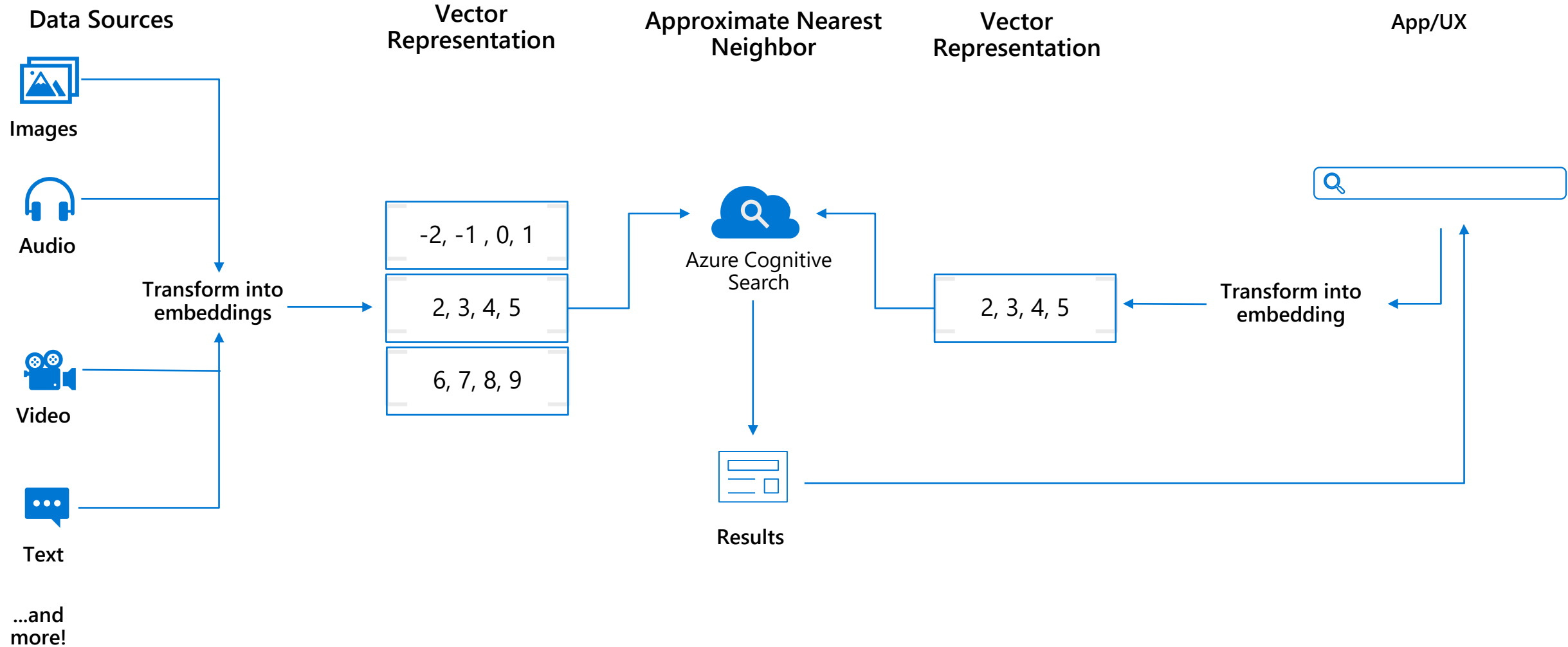


Vector Search/Databases



| What is Vector search?

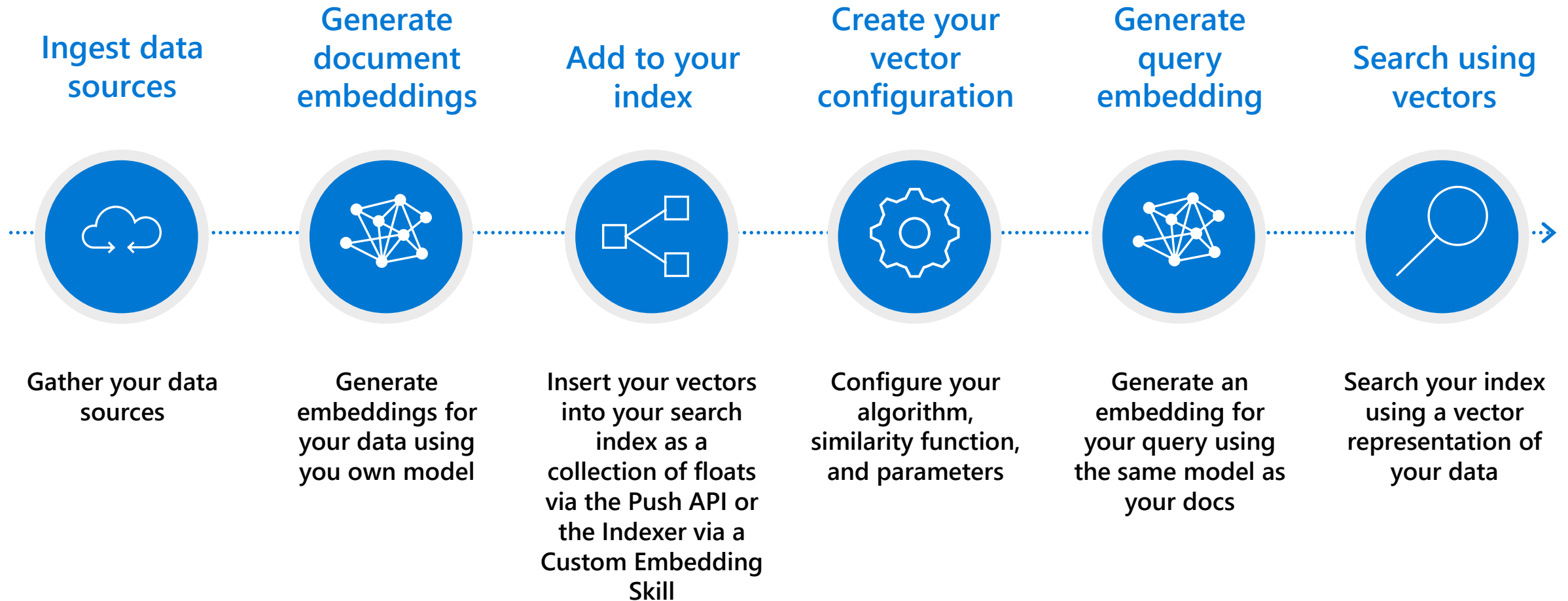
Convert data into vector representations where distances represent similarity



| Embeddings



| How do I get started with Vector search?

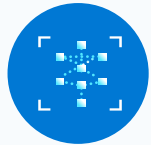


| What is the value of Azure Cognitive Services?

A portfolio of customizable cognitive APIs for vision, speech, language, and decision-making



Vision



Computer vision

Custom vision

Face API

Speech



Speech to text

Text to speech

Speech translation

Speaker recognition

Language



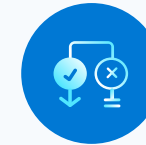
Language understanding

Custom question
answering

Text analytics

Translator

Decision

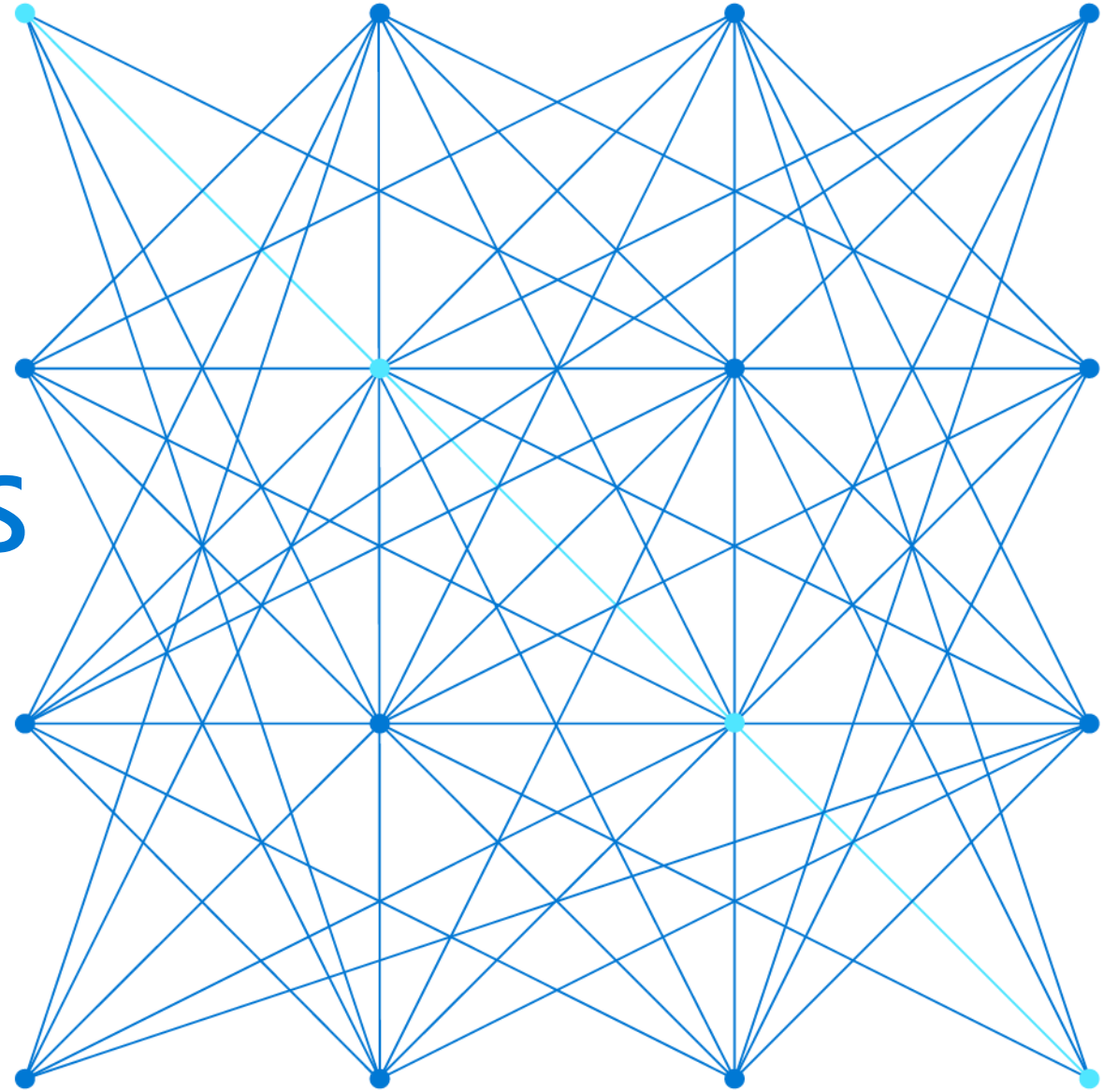


Anomaly detection

Content moderator

Personalizer

Demo: Intelligent App with Azure AI Services & AKS



Intelligent App: Business Scenario

- Pet supply store admin portal
- As new products are added, we would like to add meaningful descriptions by “asking” OpenAI

Azure Pet Supplies - Admin Portal

Orders

Products

Save Product

Name Indestructible Chew Toy

Price 12.99

Keywords puppy dog chew toy teething training

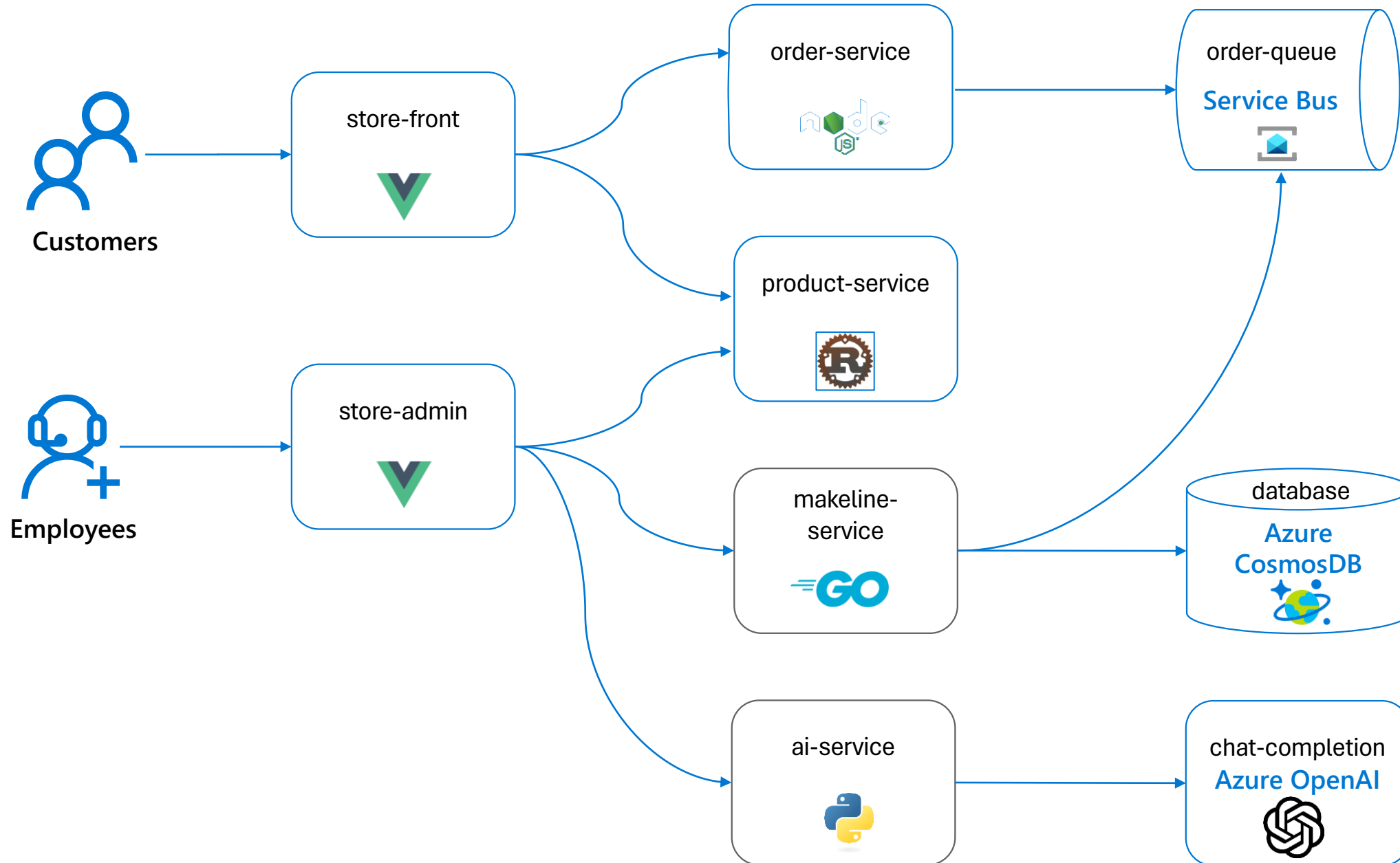
Description

Train your puppy's chewing habits with the Indestructible Chew Toy. Made from tough materials, it withstands even the most aggressive teething while promoting healthy chewing habits. Perfect for training and playtime.

Ask OpenAI

Image /placeholder.png

| An Intelligent App example



| Our Deployments:

- Provision AKS cluster
- Provision Azure OpenAI account
- Deploy gpt-35-turbo (or gpt-4 now) model
- Deploy AKS store demo app (without AI)
- Deploy AI component
- <https://aka.ms/openai-aks>

Takeaways



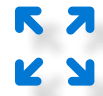
AI is truly revolutionizing the way we think about apps



Cloud-native approach allows for faster innovation



AKS platform can help you operate more reliably at any scale



Infuse AI into enterprise applications or build new

| Additional resources/next steps

- [Kubernetes Learning Path](#)
- [Microsoft Reactor Summer Series: Intro to AKS and deploying apps](#)
- [Semantic Kernel: What it is and why it matters](#)
- [Deploy and run an Azure OpenAI/ChatGPT app on AKS \(Terraform\)](#)
- [Deploy and run an Azure OpenAI/ChatGPT app on AKS \(Bicep\)](#)
- <https://aka.ms/openai-aks>
- <https://aka.ms/secure-aoai-aks-lab>
- <https://aka.ms/AKSLZA/aiscenarioworkshop>
- [Vector Database | Microsoft Learn](#)

**HERZLICHEN DANK FÜR EURE ZEIT
UND EUER ENGAGEMENT!!!**

