

Développez votre activité en tirant le meilleur parti d'Azure OpenAI.

Bechtle-IT Forum | 13.06.2023 | SwissTech Convention Center Lausanne

Alexandra Grozea, Partner Technology Strategist, Microsoft
Louis Grillon, Consultant Data & Analytics, Bechtle Suisse

Agenda

Azure OpenAI Basics

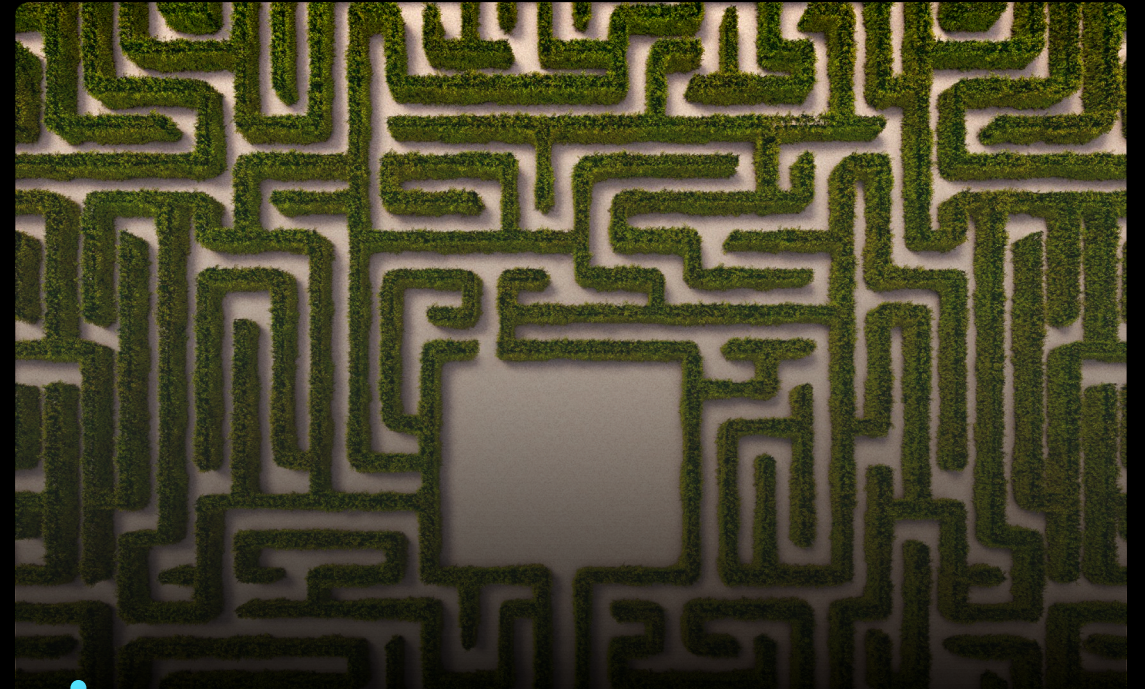
Azure OpenAI Capabilities

Use Cases + Demo

Large Language Models transform how we think about AI



Simple model:
Purpose-built for one use case



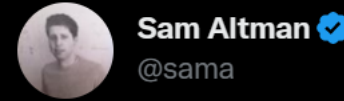
Foundation model:
Broader intelligence with many applications

Microsoft and OpenAI



Satya Nadella ✓
@satyanadella

I think of this next generation of AI where we have built very unique infrastructure. In fact, the best AI infrastructure is on Azure in which we trained models like GPT from OpenAI



Sam Altman ✓
@sama

microsoft, and particularly azure, don't get nearly enough credit for the stuff openai launches. they do an amazing amount of work to make it happen; we are deeply grateful for the partnership. 🙏 they have built by far the best AI infra out there.

GPT-3.5 and GPT-4
Text

ChatGPT
Conversation

Codex
Code

DALL·E 2
Images

Our AI Offerings



Azure Cognitive Services



Pretrained Models



Generic Use Cases

Is it a common use case?

Azure OpenAI Service



Generative AI



Generate Content

Do we need to generate content or a complex reasoning logic?

Azure Machine Learning



Custom Models



Workspace for Data Scientists and ML engineers

Is the necessary skill set present?

Microsoft AI Portfolio



Applications



Partner Solutions

Application Platform

AI Builder



Power BI



Power Apps



Power Automate



Power Virtual Agents

Scenario-Based Services

Applied AI Services



Bot Service



Cognitive Search



Form Recognizer



Video Indexer



Metrics Advisor



Immersive Reader

Customizable AI Models

Cognitive Services



Vision



Speech



Language



Decision

Azure OpenAI Service

ML Platform



Azure Machine Learning



Business Users



Developers & Data Scientists



The Differences - Why Azure OpenAI?

Azure runs on trust

Your data is your data.

Data is stored encrypted in *your* Azure subscription

Your data from any fine-tuning is not used to train the foundation AI models.

Azure OpenAI Service provisioned in *your* Azure subscription

Model fine tuning stays in *your* Azure subscription and never moves into the foundation AI models

Your data is protected by the most comprehensive enterprise compliance and security controls.

Encrypted with Customer Managed Keys

Private Virtual Networks, Role Based Access Control

Soc2, ISO, HIPPA, CSA STAR Compliant

Microsoft's responsible AI principles



Where are You in Your OpenAI Journey?

Please give us your feedback



Azure OpenAI Capabilities

Killer Apps and User Experiences

Your Product Vision & Your Use Cases

Your Platform and Your Ecosystem

Your Proprietary Corpuses of Data

Embeddings

Fine-Tuning

Your Proprietary Instances of Models to customize

API calls

Prompt Engineering

GPT-3.5

GPT-4

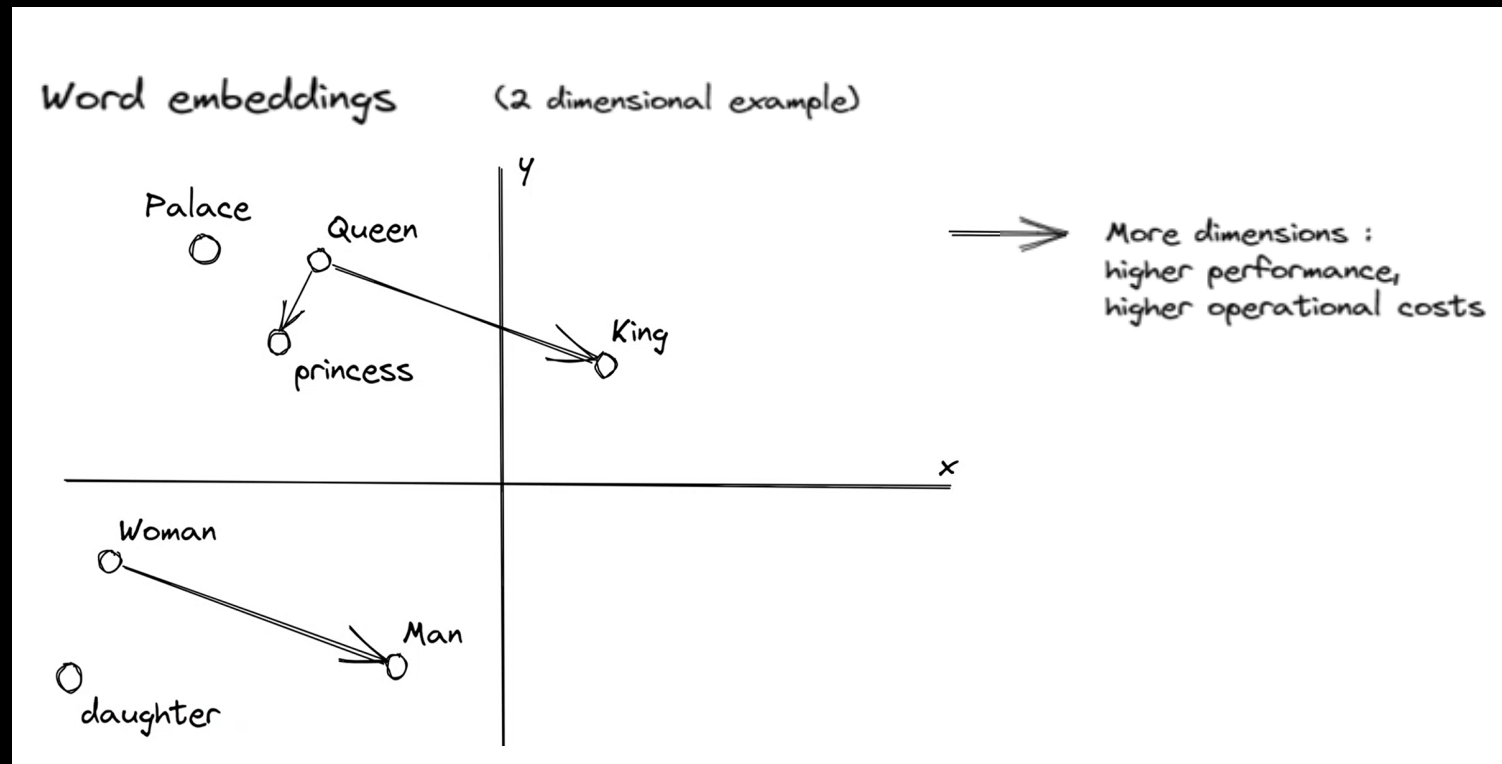


DALL-E

Codex

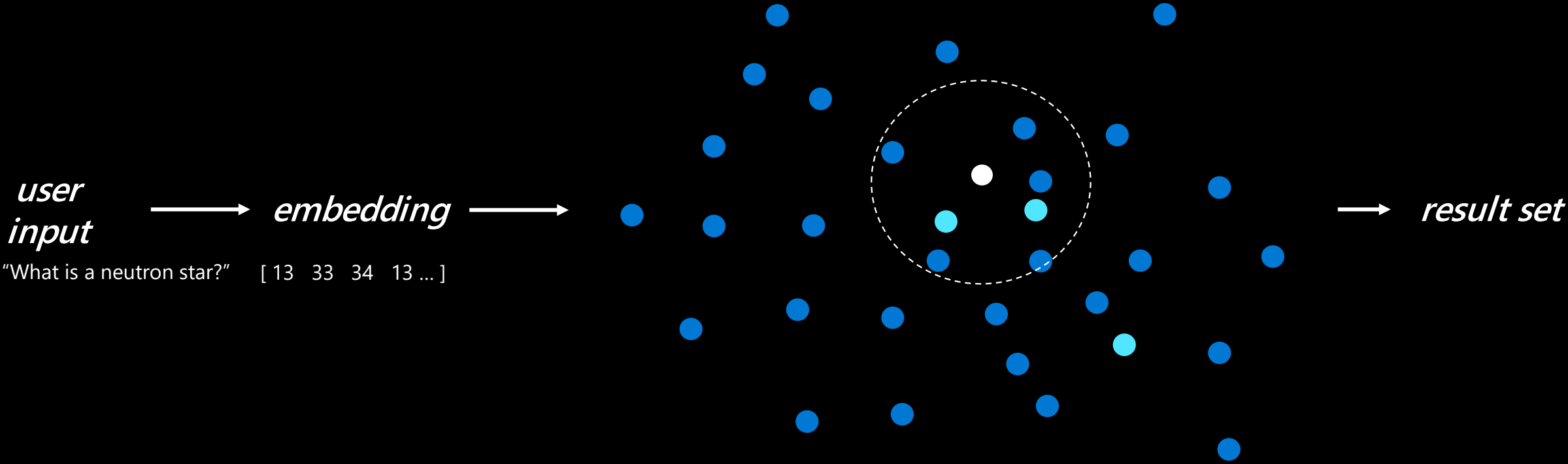
Embeddings

We need embeddings to convert words into numbers:



Similarity Search with embeddings

Once you encode your content as embeddings, you can then get an embedding from the user input and use that to find the most semantically similar content. Vector databases can be used to store those embeddings and assist when doing a research.



Prompt Engineering



"The art of telling chatGPT what to do"

GPT-3 Playground

Deployments: text-davinci-003 | Examples: Load an example | [View code](#)

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.

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 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 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.

JSON:

```
{
  "reason": "Car accident report",
  "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 and John Radley were involved in a car accident on the I-18 freeway. Sarah believes she hit a pothole, causing the accident. No one was injured, but Sarah's car has suffered damages to the headlights and airbags. Sarah will need to have her car towed and inspected."
}
```

Prompt Engineering Guide



1. Give clearer instructions

2. Split complex tasks into simpler subtasks

3. Structure the instruction to keep the model on task

4. Prompt the model to explain before answering

5. Ask for justifications of many possible answers, and then synthesize

6. Generate many outputs, and then use the model to pick the best one

7. Fine-tune custom models to maximize performance

Best results with lots of experimenting and trial-and-error

Tips for Prompt Engineering



Include

Start by **including instructions** to tell the model what to do. It makes sense to include to not make up stuff.

Be precise in stating what is asked.



Restrict

Restrict the output (e.g., choose from a confined list.)
Tell the model what not to do.



Explain

Give reasons why things are important.
Define a role for the model
"act as a twitter user"



Context

Dynamically inject content to give context
Balance model freedom vs control
Split prompts into multiple requests to make responses more predictable.

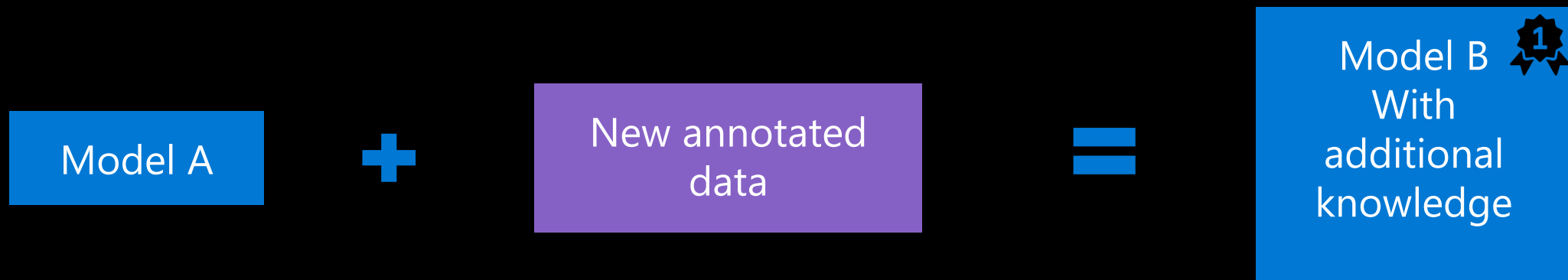


Position

Position most important instructions at first.
Sometime repeating instructions helps.
Add Chain of Thought style of instruction, "Solve the problem step by step."

Fine-Tuning

- Fine Tuning is the process of actually retraining a model with additional data, to get increased accuracy.
- Fine Tuning is not possible with newer generation of models.
- It has been shown that augmented prompt engineering usually outperforms fine-tuned models.



Azure OpenAI Service Use Cases



Language

- Contact Centers
 - Classification – route mails to appropriate team
 - Sentiment – prioritize angry customers
 - Mail and call transcript summarization
 - Customer response email generation
- Rapid response marketing campaigns: classification, sentiment, summarization, content generation, image generation
- Chatbot Knowledge Hub internal to a company
- Automatic entity extraction to create reports on news, documents, articles and analyze liability, risk, trends.

Codex

- Natural Language to Code
- Natural Language to SQL
- Code to Natural Language
- Code documentation
- Refactoring

DALL·E 2

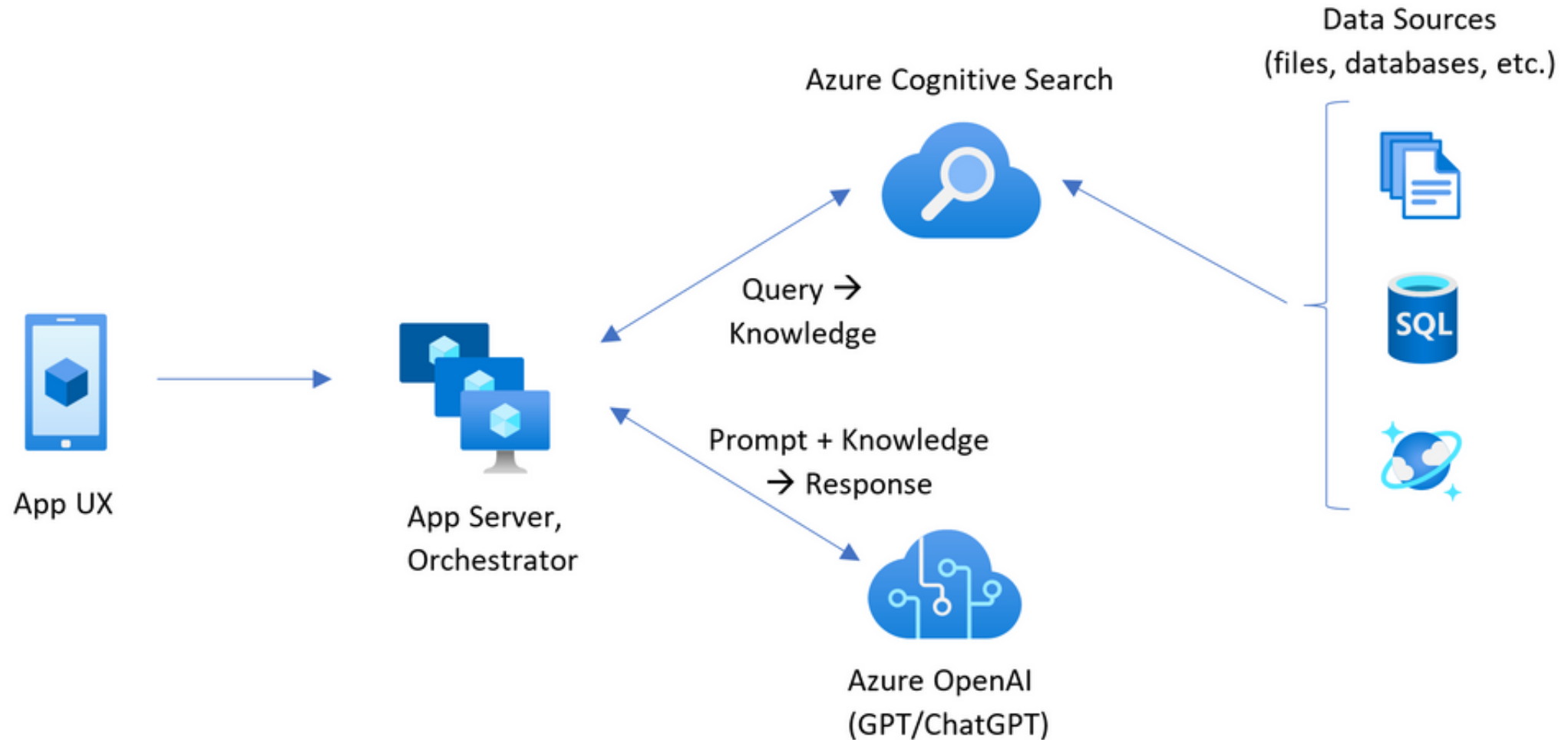
- Creative ideation
- Podcast and music playlist images
- Content syndication
- Marketing campaign personalization
- Hyper-personalization

Company knowledge hub: Business Context

- A fake company wants to enhance employee performance and reduce costs by implementing a chatbot knowledge hub.
- The company has gathered PDF files containing processes and internal information.
- The goal is to create an application similar to chatGPT that can effectively answer employee questions, only based on custom data, and avoid losing time going through countless files and folders !



Custom ChatGPT architecture






Chat with your data

Ask anything or try an example

What is included in my Northwind Health Plus plan that is not in standard?

What happens in a performance review?

What does a Product Manager do?

Type a new question (e.g. does my plan cover annual eye exams?) 

How to get started with Azure OpenAI?

Understand industry challenges/AI use cases

Identify business problems and look out for text processing hot spots

Ask professionals for strategic or technical support

Experiment and do own micro-projects and PoCs

Leverage Azure, Power BI and Power Platform as entry point



Merci!

Des questions? Contactez-nous: it-forum.ch@bechtle.com

Plus d'informations :
[bechtler.com](https://www.bechtler.com)

