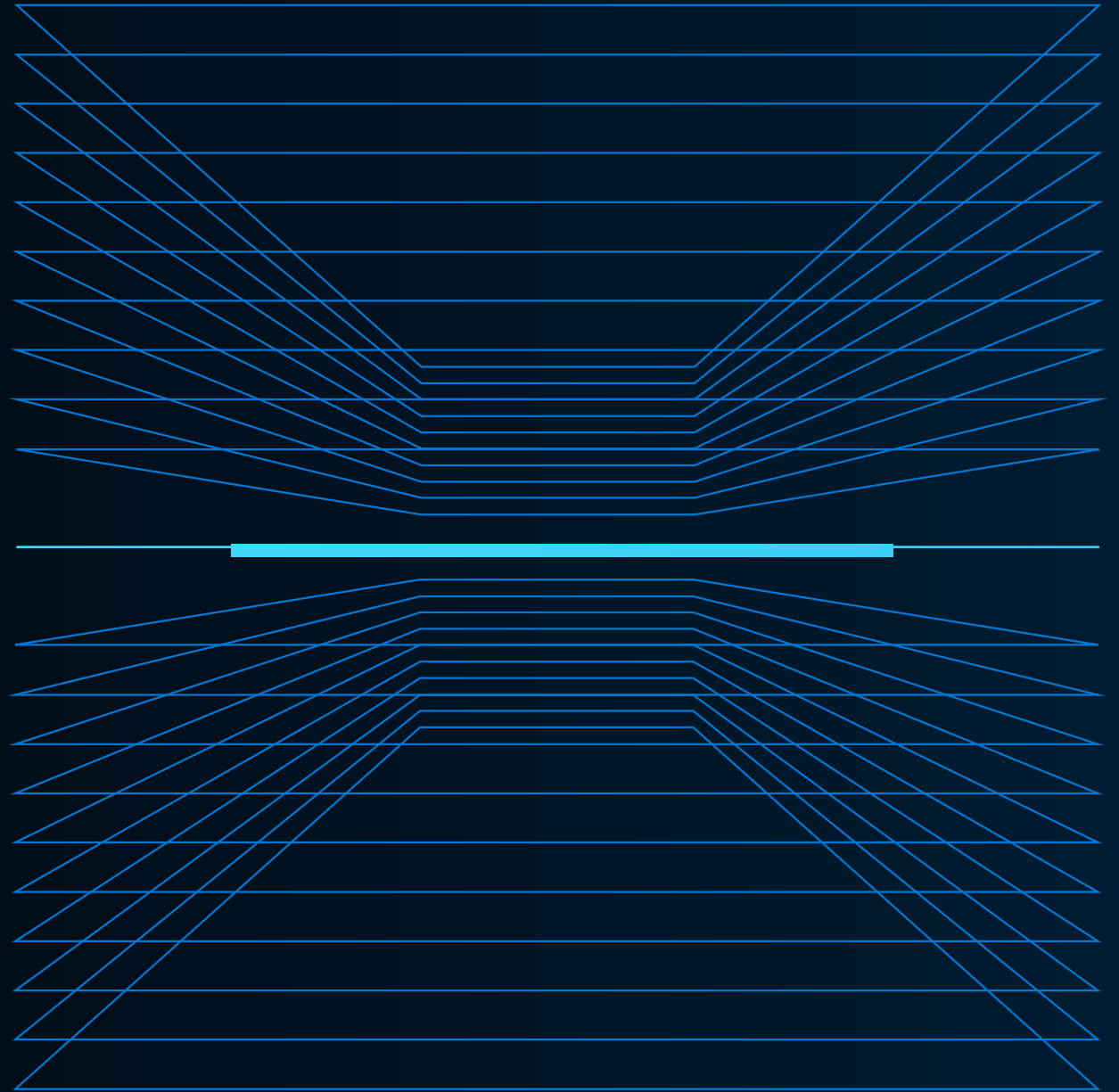




An introduction to Azure OpenAI Service

from the developer
perspective

John Metzger
Microsoft



Brief history of AI

Artificial Intelligence

Machine Learning

Deep Learning

Generative AI



Artificial Intelligence

the field of computer science that seeks to create intelligent machines that can replicate or exceed human intelligence



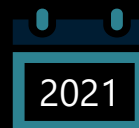
Machine Learning

subset of AI that enables machines to learn from existing data and improve upon that data to make decisions or predictions



Deep Learning

a machine learning technique in which layers of neural networks are used to process data and make decisions



Generative AI

Create new written, visual, and auditory content given prompts or existing data.

Azure OpenAI Service

Large pretrained foundation AI models
custom-tunable with your parameters and your data

GPT-3.5

GPT-4 (preview)

ChatGPT (preview)

Codex

DALL·E 2 (preview)

Generative Text Models, with varying capabilities and uses

Specialised Generative Coding Model

Generative Image Model

Azure AI

Ecosystems



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

Azure OpenAI

Top Capabilities and Use Cases



Content generation

Intelligent Virtual Agent:
automatically generate responses to inquiries

Customized Copilot:
Augment decision steps in business processes

Generate engaging web content



Summarization

Subject matter expert document:
summarization (e.g. compliance and training documents)

Social media trends summarization



Code generation

Convert natural language to SQL (or vice versa) for telemetry data

Code documentation



Semantic search

Knowledge Exploration:
Search reviews for a specific product/service

Information discovery and knowledge mining

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

| Microsoft Azure Cloud

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

GPT-3.5

ChatGPT (preview)

GPT-4 (preview)



Large, pretrained Language Models
that use deep learning to produce text outputs



Custom-tunable AI models with your data



Built-in responsible AI to detect and
mitigate harmful use

Comparison of GPT versions

GPT-3.5

- Available in Western Europe
- Use-case specific models to optimize inference time and performance
- Suitable for a large range of use cases

ChatGPT (preview)

- Should be first choice for most use cases
- Most economical GPT model in Azure OpenAI Service
- For all workloads, not just Chat

GPT-4 (preview)

- Improved problem solving and reasoning capabilities
- Iterative refinement:
 - Paste in code errors & GPT-4 will fix for you
 - Iterate on stories
- Increased token limit - works well for long content

Concepts

Service

Model

Model deployments

Tokens and limits

Prompts

Grounding

Embeddings

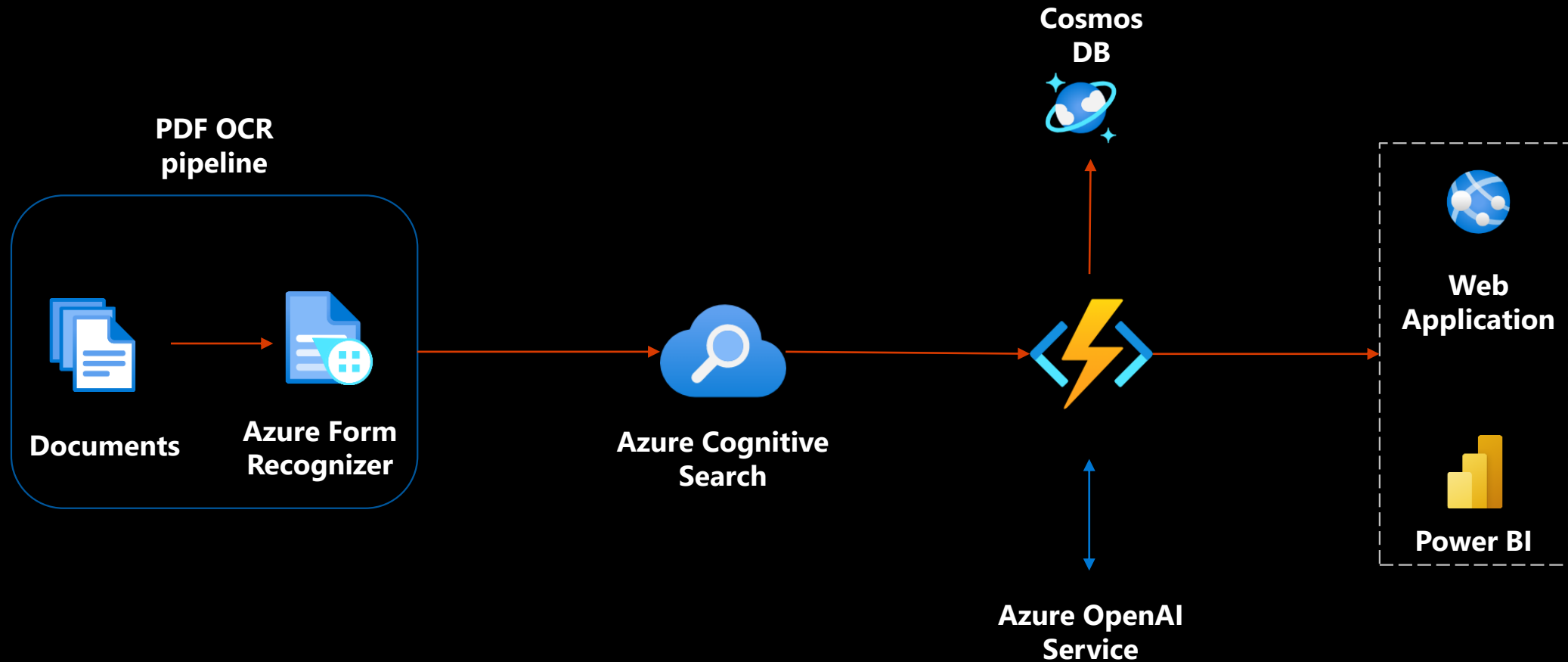
Retrieval Augmented Generation

Fine tuning

Demos

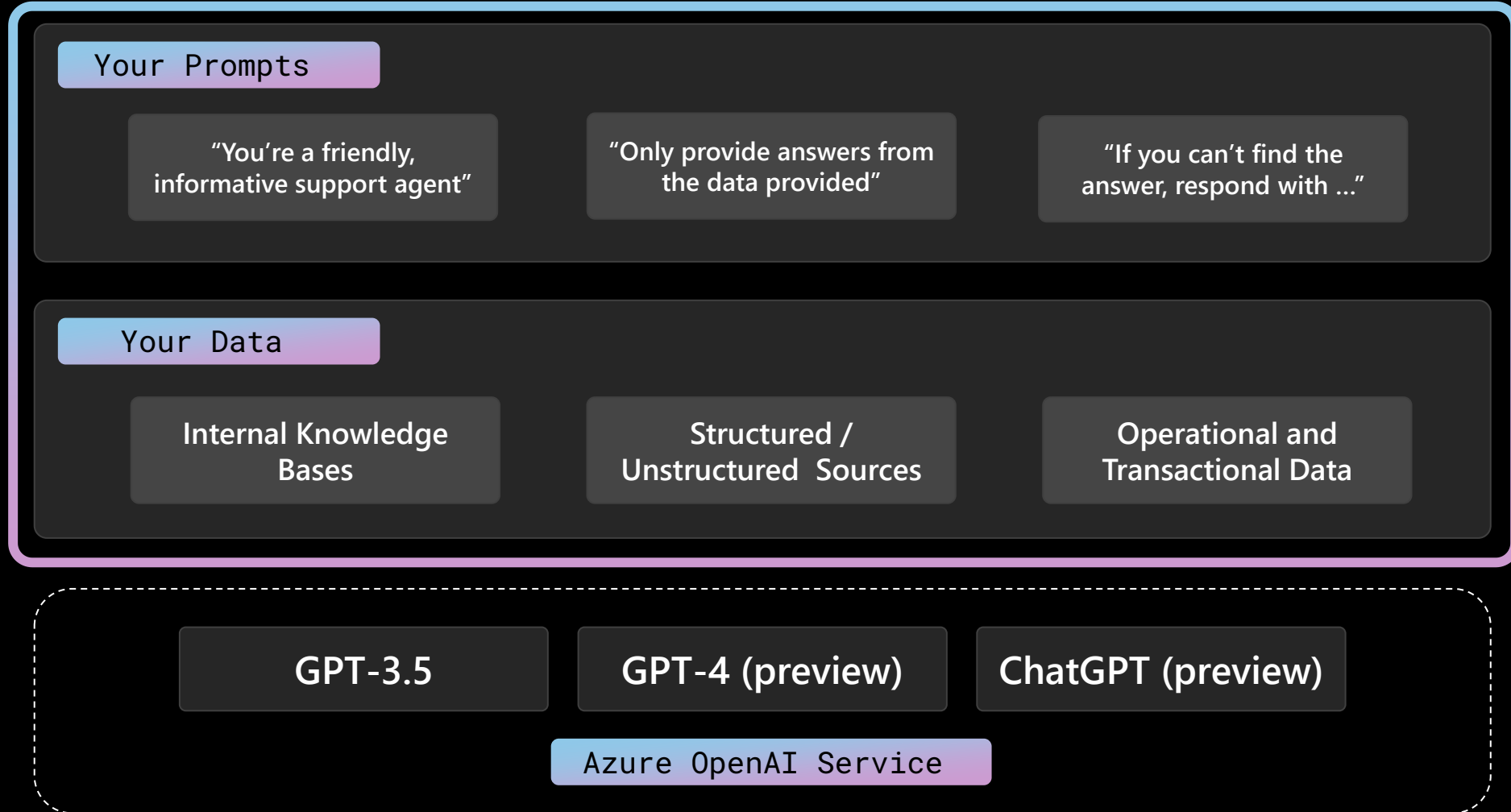
Document Process Automation

Extract rich insights from documents and summarizing them



| Customizing Azure OpenAI

Your Differentiation



| GPT-3 Ideate, Experiment and Fine-Tune

Iterate on ideas with
a general-purpose
text-in/text-out interface

Prompt

Summarize game commentary
into highlights:

Shey Peddy is applying ball pressure at
the top against Sabrina Ionescu. At
7:48 remaining in the quarter; Peddy

What are the main highlights of the
game so far?

Sample response

The game has been close with Phoenix
leading New York 7-5. Shey Peddy has
been key for Phoenix.

Refine with examples
(‘few shot learning’) with
a simple UX

Prompt

Turn game commentary into highlights:

Commentary: What a pickup she has

Main highlights: New York has domina

###

Commentary:

1. Turner is so important defensively to
2. Griner pulled way out, Hartley with
3. At 1:54 remaining in the quarter, Pho

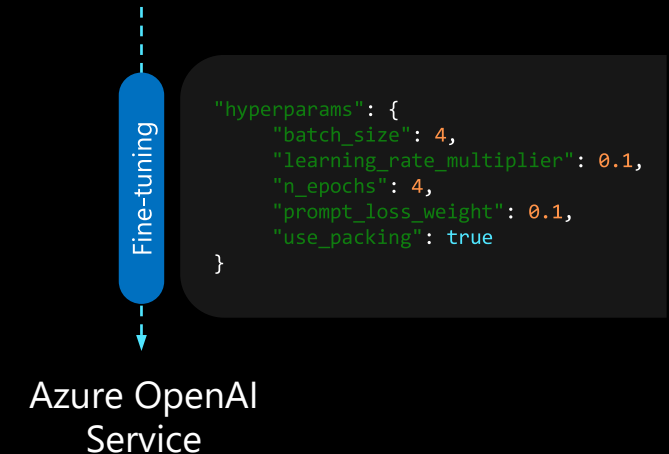
Sample response

Main highlights:

1. New York has had a strong run in the
2. Phoenix leading by 1 point, 24-23
3. New York Liberty's comeback has be

Optimize accuracy and
latency to validate proof
of concept fast

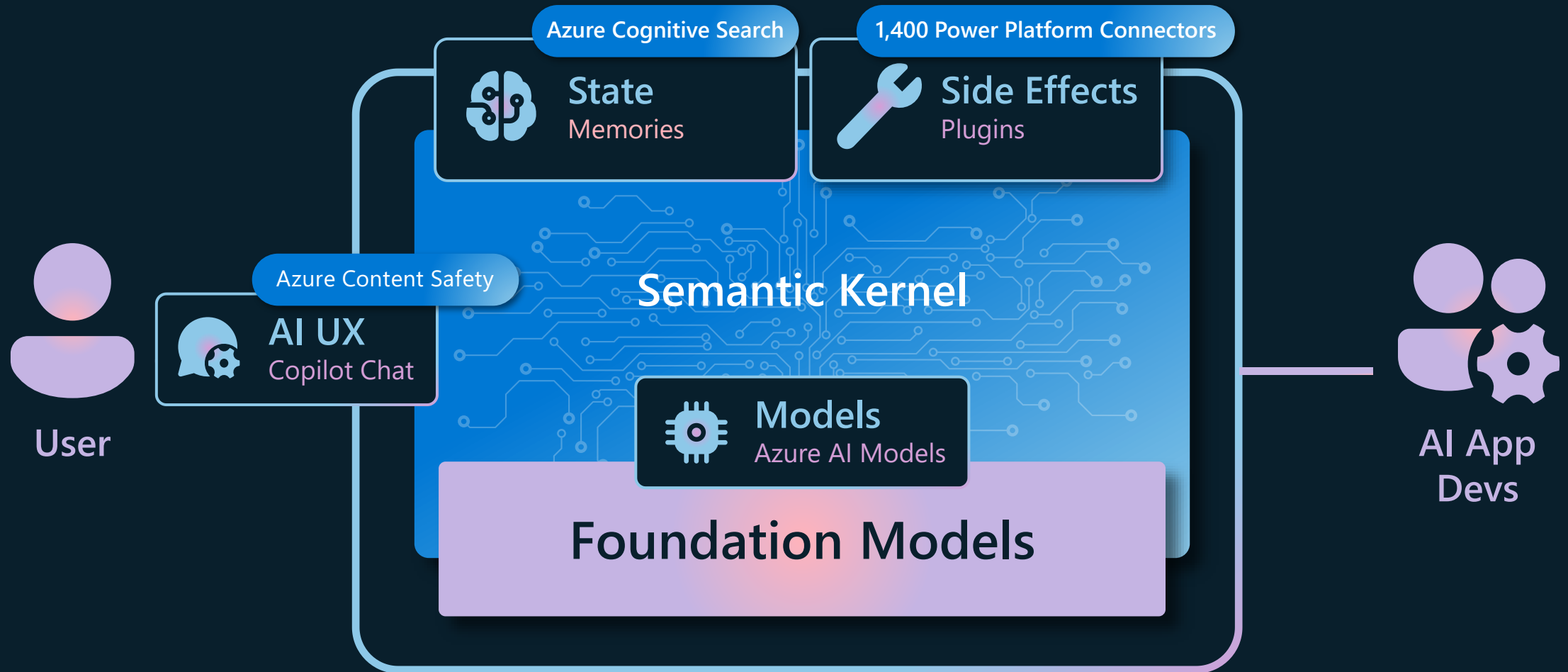
Prompt and completion examples



SEMANTIC KERNEL

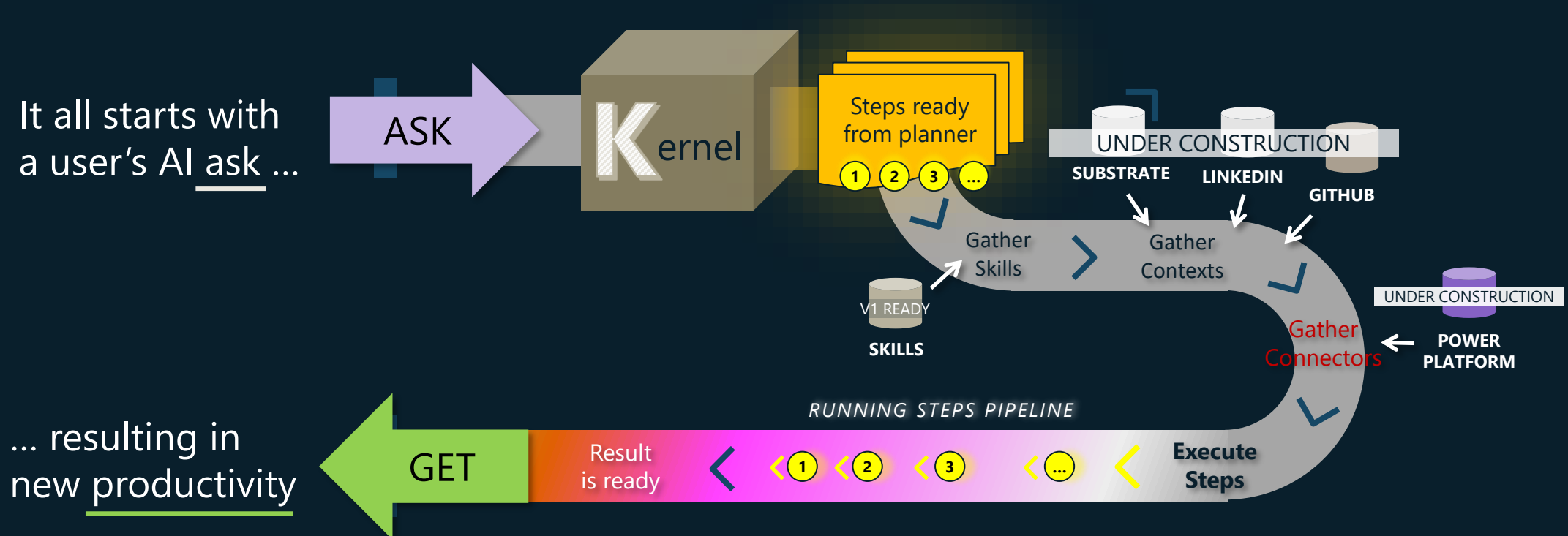
Plugins, Planners, Personas

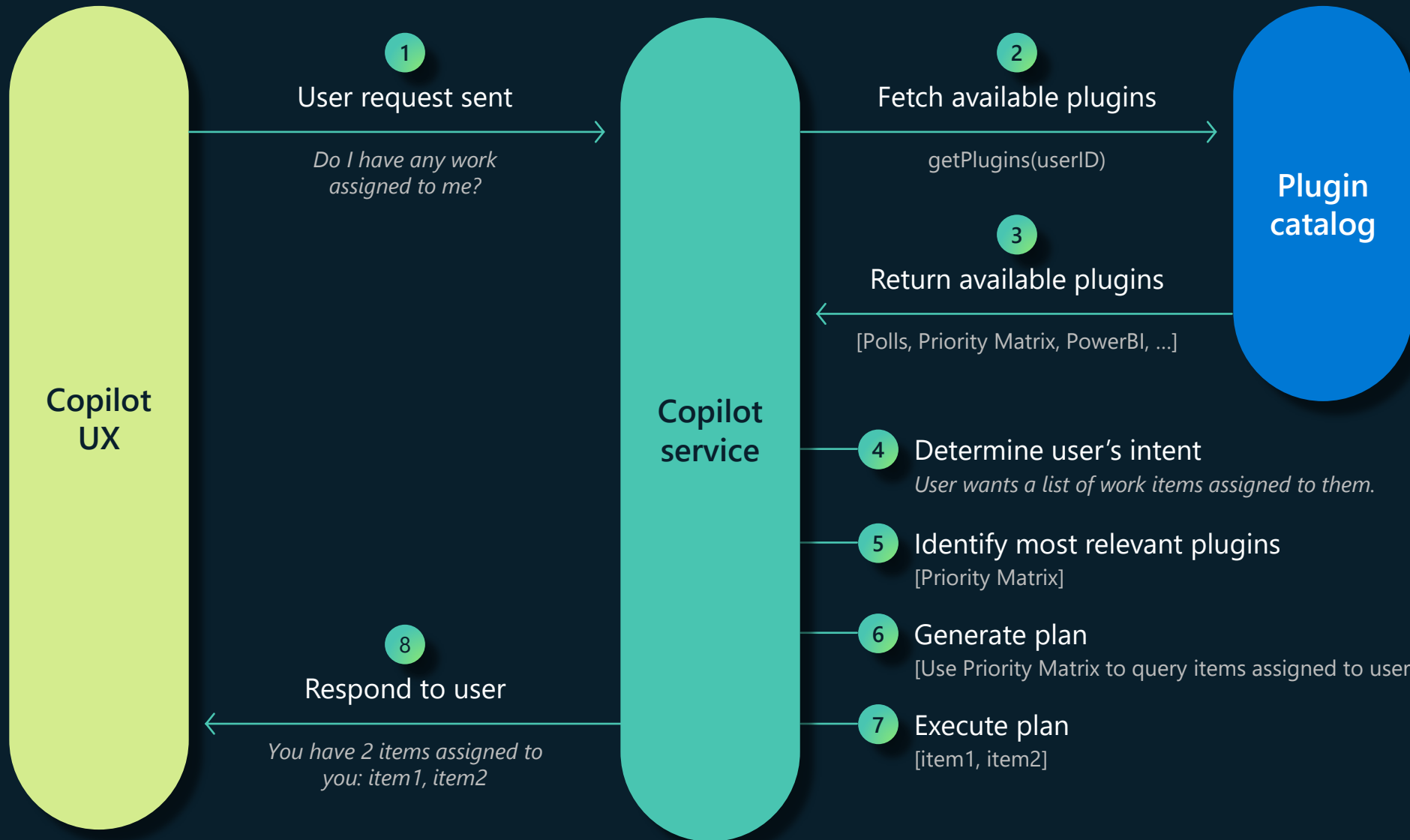
Semantic Kernel made it easier for AI App Devs to get going.



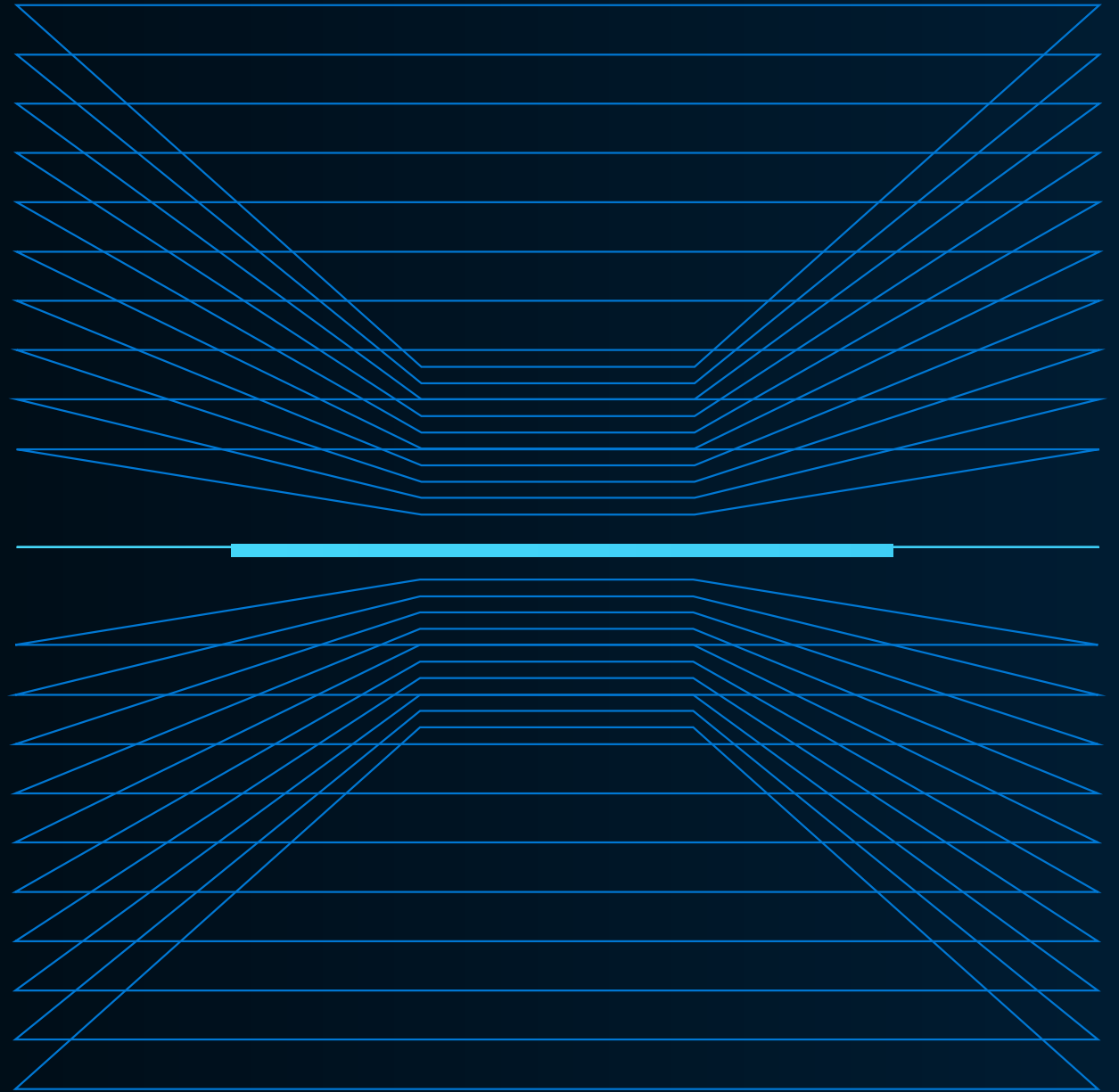
Semantic Kernel Elements and Execution

It all starts with
a user's AI ask ...





Questions?





Thank you

