

# Global AI Community

## Azure Agentic Prompt Studio Agentic AI | MCP | A2A

PRESENTED BY

Alok Adhao  
UBS - Cloud Specialist



**Microsoft®**  
Most Valuable  
Professional



@casa-cloud



# Overview of Agentic AI Framework – Prompt To Pod

Explore the essential components of building an Agentic AI ChatOps Assistant for AKS.

1

*Introduction to Agentic AI concepts*

2

**Understanding Model Context Protocol**

3

**Exploring Agent to Agent Protocol**

4

**Utilization of Vector Store technology**

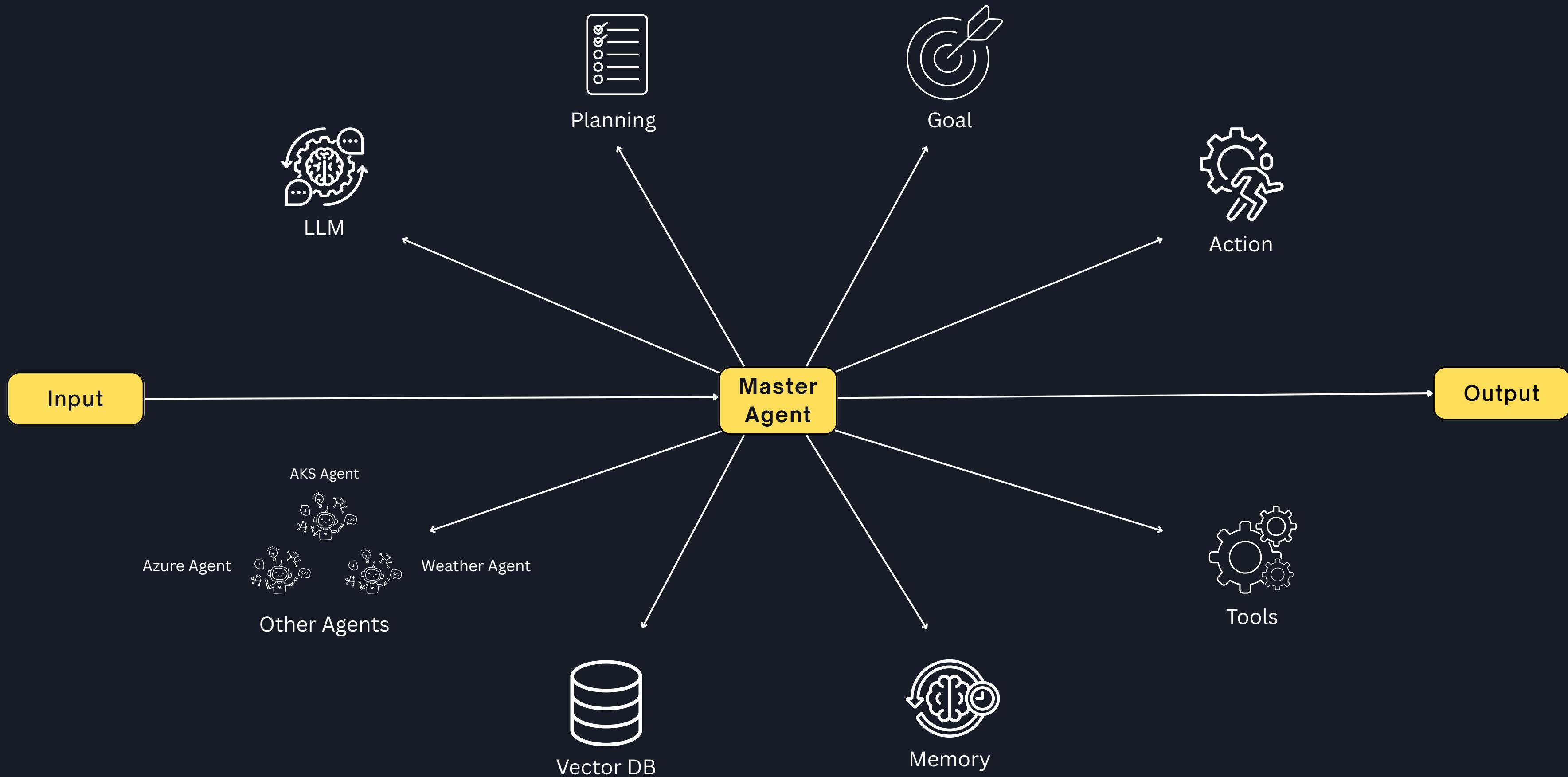
5

**Understanding RAG in AI systems**

6

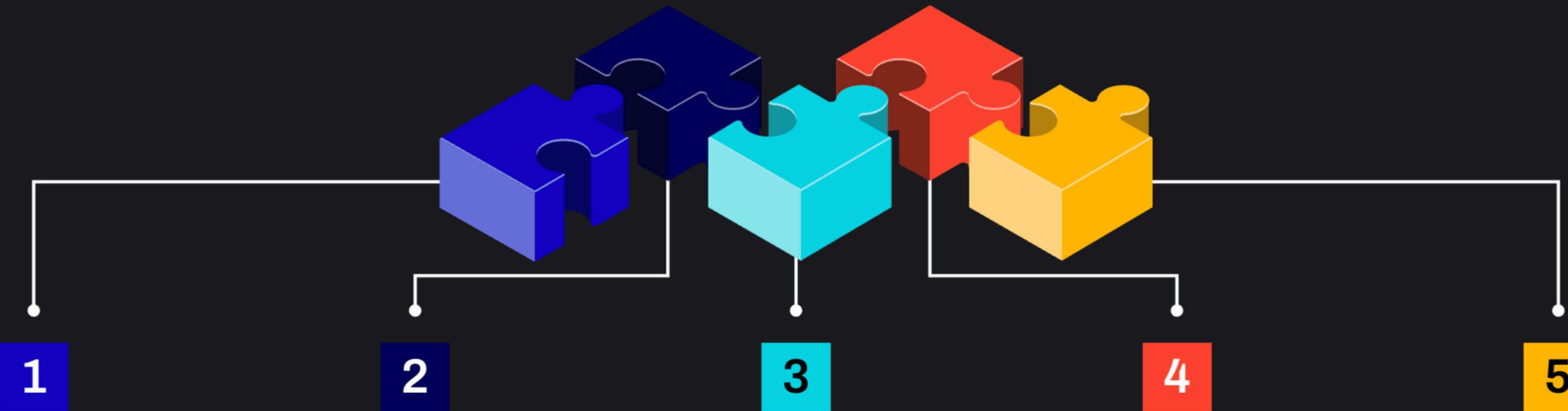
**Implementation of AKS Chatbot Assistant**

# Agentic Framework Flow



# Model Context Protocol Overview

Exploring how MCP enhances AI interactions and integrations



**1**

## MCP Overview

MCP is an open standard to enable connections between AI-powered applications like LLM

**2**

## Standardised Framework

MCP serves as a **standard open-source** framework

**3**

## Universal Adapter

It acts as a **universal adapter** for AI systems

**4**

## Natural Conversations

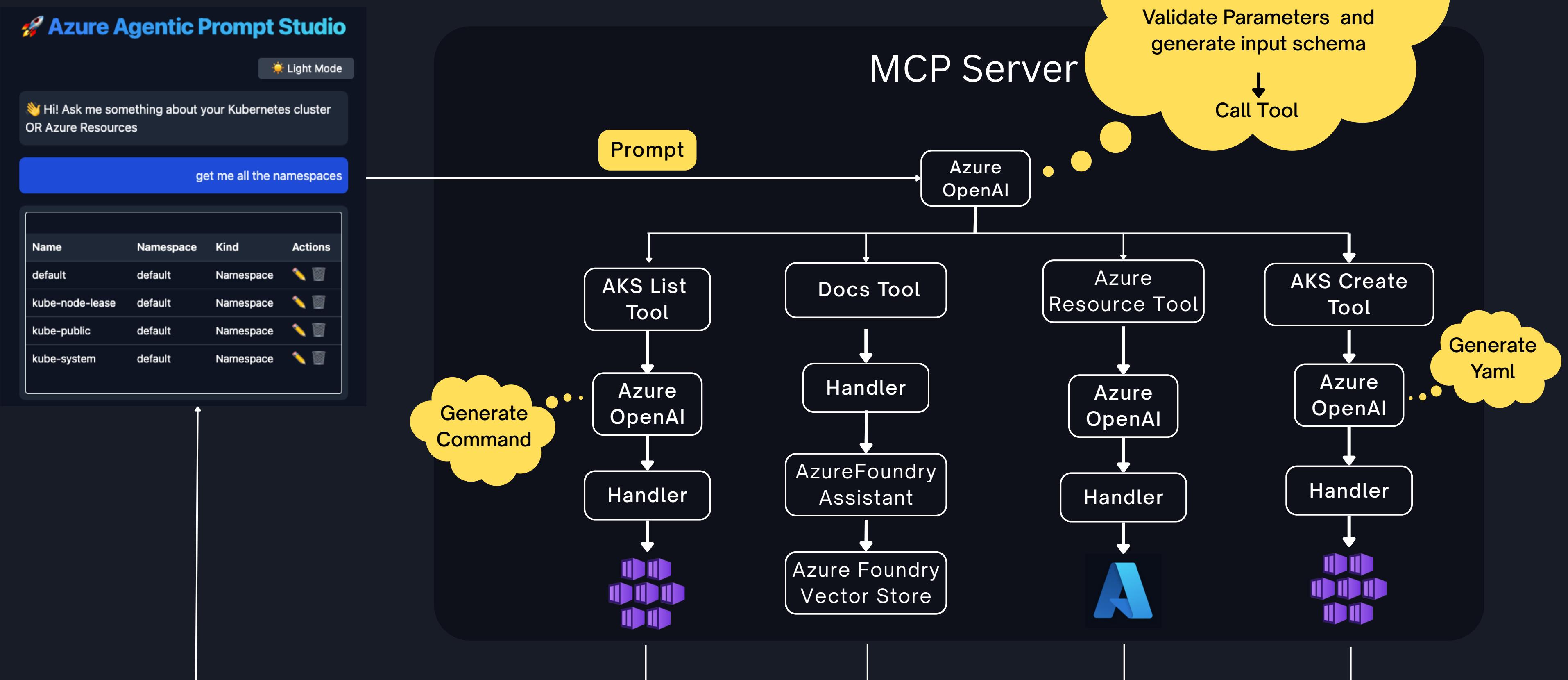
By maintaining context, MCP facilitates **more natural and coherent** conversations

**5**

## Integration with LLMs

Designed particularly for **Large Language Models (LLMs)**

# Model Context Protocol



# Agent to Agent Communication Framework

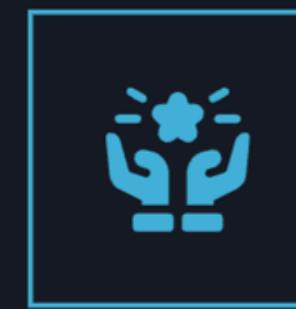
Enhancing collaboration between AI agents

## A2A Protocol Overview



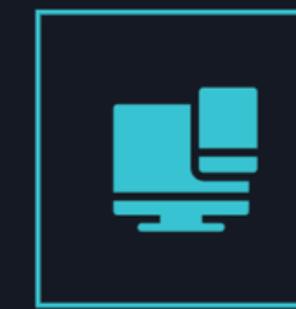
The **Agent to Agent Protocol** enables **AI agents** to communicate effectively, enhancing collaboration and task execution across platforms.

## Enhanced Task Coordination



The protocol allows AI agents to **coordinate actions**, leading to improved efficiency in task management

## Cross-Platform Integration

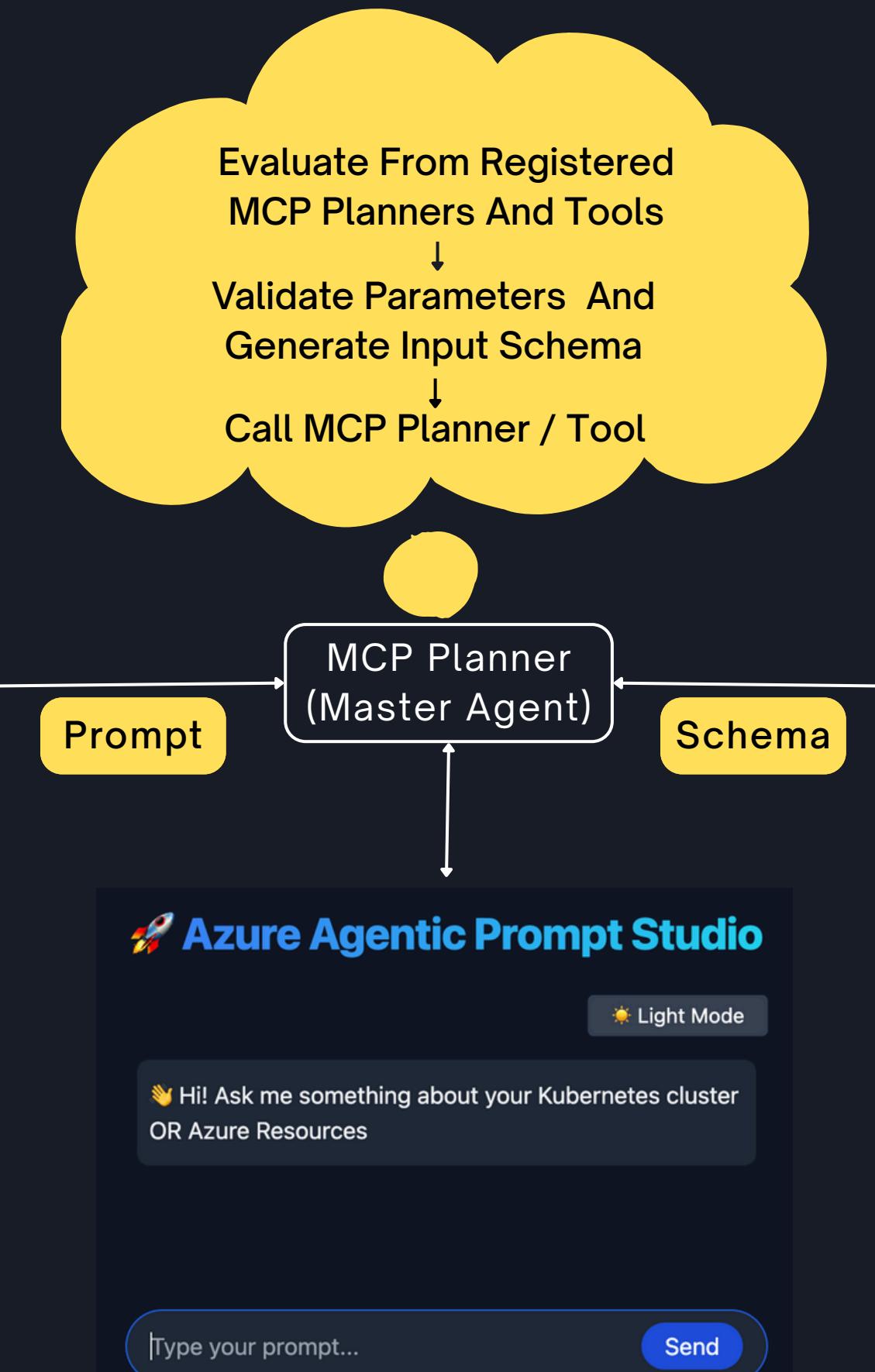
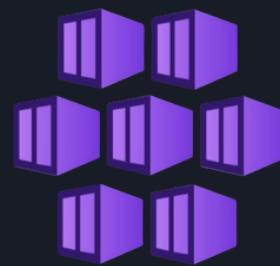


A2A facilitates communication across **various enterprise platforms**, allowing a seamless flow of information between different applications.

# Agent To Agent Protocol

MCP Planner (Agent Card)  
- AKS

```
public > .well-known > {} agent.json > ...
1  {
2    "name": "k8screateagent",
3    "description": "create a Kubernetes resource",
4    "tools": [
5      {
6        "name": "k8screateTool",
7        "description": "create a Kubernetes resource",
8        "planner": true,
9        "input_schema": {
10          "type": "object",
11          "properties": {
12            "prompt": { "type": "string" }
13          },
14          "required": ["prompt"]
15        }
16      ],
17      {
18        "type": "agent",
19        "endpoints": {
20          "mcp": "http://localhost:9999/v1/mcp"
21        }
22    }
```



MCPTool(Agent Card)  
- Weather

```
public > .well-known > {} agent.json > ...
1  {
2    "name": "weather-agent",
3    "description": "Provides weather data from external API",
4    "tools": [
5      {
6        "name": "WeatherTool",
7        "description": "Returns weather data for a city",
8        "input_schema": {
9          "type": "object",
10          "properties": {
11            "city": { "type": "string" },
12            "field": { "type": "string" }
13          },
14          "required": ["city"]
15        }
16      ],
17      {
18        "type": "agent",
19        "endpoints": {
20          "mcp": "http://localhost:8888/v1/mcp"
21        }
22    }
```



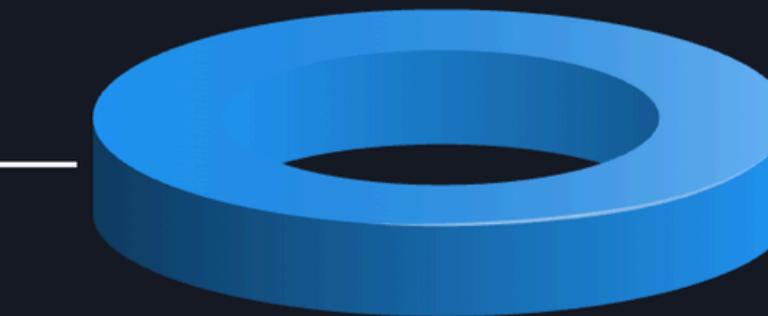
# Azure Vector Stores

Leveraging vector databases for AI performance improvement

- **Normal DB:**  
Query: `SELECT * FROM quotes WHERE text LIKE '%moon%'`  
→ Finds only exact word matches.
- **Vector DB:**  
Query: *"Find quotes similar in meaning to: 'Reach for the stars.'"*  
→ Finds results like "Aim high," "Dream big," or "The sky is not the limit."

## Fast Similarity Searches

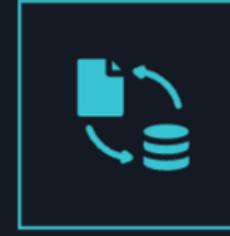
Vector stores enable quick retrieval of data points that are similar, enhancing performance.



- Normal DB: Only searches by filename or tag ( "astronaut.jpg" ).
- Vector DB: Can find visually *similar* images even if filenames differ — because it compares embeddings, not text.

## Efficient Data Retrieval

Vector stores allow for efficient management of large datasets, improving access to relevant information.



## High-Dimensional Data

These databases are optimized for managing high-dimensional vectors, crucial for AI applications.

## Optimized for AI Systems

Vector databases are specifically designed to enhance the performance of AI systems by speeding up data access.

*"Traditional databases store facts; vector databases store understanding.  
Together, they form the backbone of modern Agentic AI and RAG applications."*

# Understanding Retrieval-Augmented Generation

Enhancing AI responses with contextual data



## RAG Framework

Retrieval-Augmented Generation (RAG) integrates **information retrieval** with generative capabilities, enhancing AI's performance.

## Integration with LLMs

RAG enhances **large language models** (LLMs) by connecting them to external knowledge sources for better accuracy.

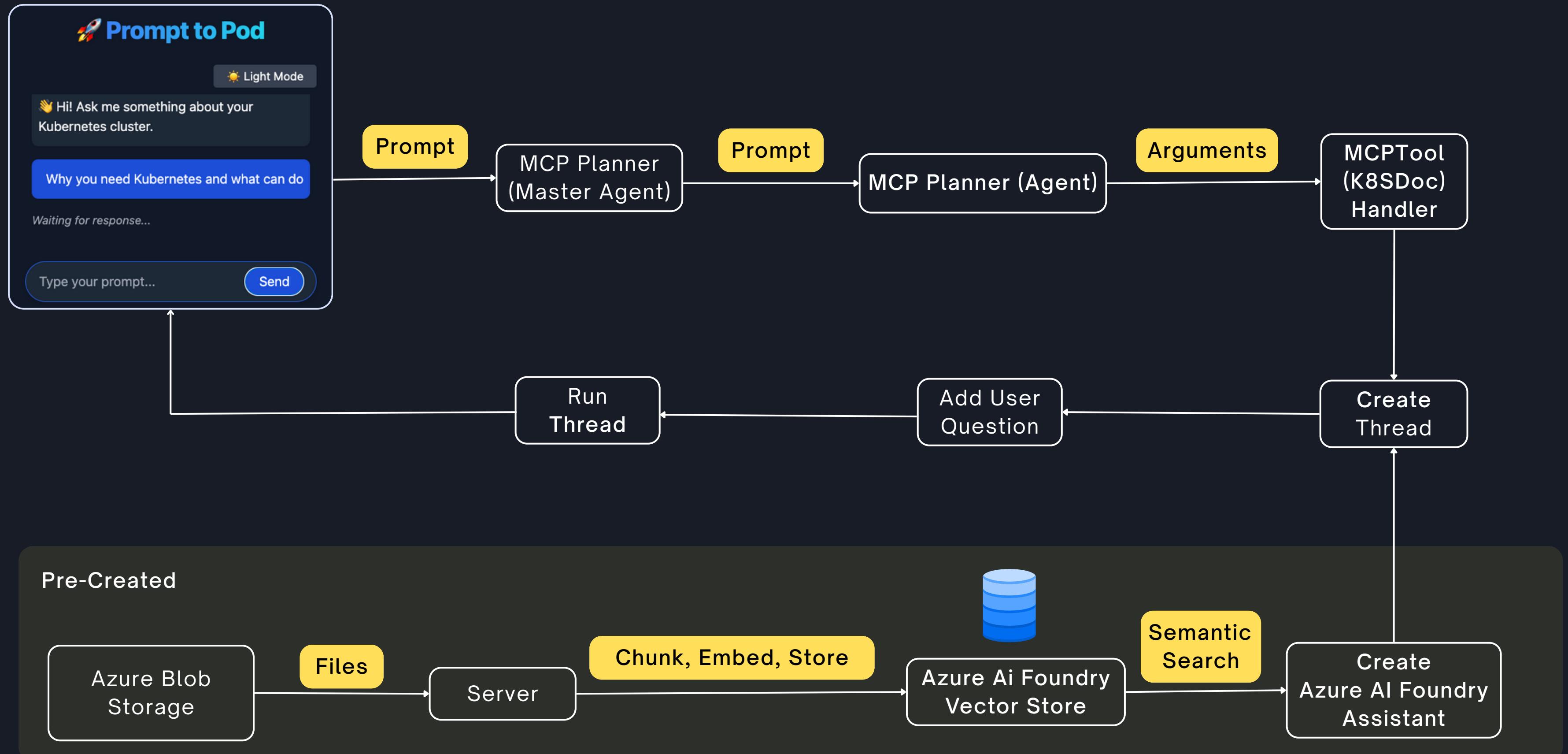
## Improved Accuracy

By leveraging **external data**, RAG ensures responses are relevant and up-to-date, improving overall output quality.

## Contextual Relevance

RAG provides **contextually relevant** information that allows AI to generate more tailored responses to user queries.

# RAG - Implementation



# AKS Chatbot Assistant

Enabling Natural Language Commands

## 1 Natural language interface

Allows users to interact with Kubernetes clusters using intuitive language, improving accessibility and ease of use.

## 2 Kubernetes integration

Seamlessly connects with AKS (Azure Kubernetes Service), facilitating effective management and operations of clusters.

## 3 Real-time command execution

Processes user commands in real-time, providing immediate feedback and responses, enhancing operational efficiency.

## 4 User-friendly design

Focuses on creating a straightforward user experience, enabling users of all technical levels to utilize the assistant effectively.

## 5 Context-aware responses

Utilizes AI to understand and respond to user queries in context, improving the relevance and accuracy of interactions.

## 6 MCP

Implemented MCP for structured tool invocation by interpreting natural language prompts and routing them to appropriate AI tools.

## 7 A2A

Multi Agent seamless connection is done using Agent 2 Agent protocol for inter Planner and Planner to Tool communication.

## 8 RAG

Real time gathering of useful information from vector store from N number of documents using natural language.

# Demo:- Azure Agentic Prompt Studio





CasaCloud

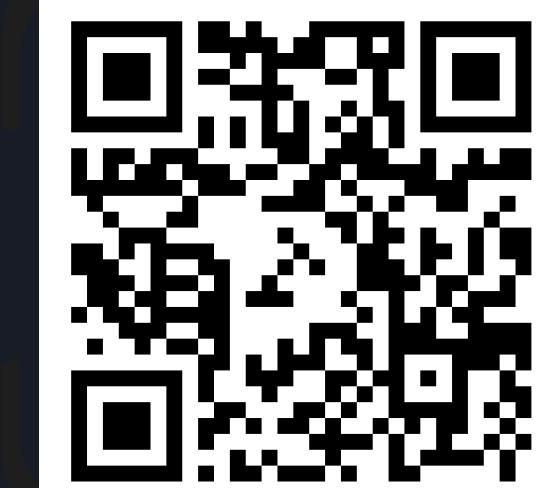
CASA CLOUD.COM



[youtube.com/@casa-cloud](https://youtube.com/@casa-cloud)



in



[linkedin.com/in/alokadhao](https://linkedin.com/in/alokadhao)