



Reactor

Build a RAG Agent with Azure AI Search

Cihan Cinar
Software & AI Architect





Reactor

Cihan Cinar

Software & AI Architect

Givaudan

Linkedin: @cihancinar

Github: @cihancinar

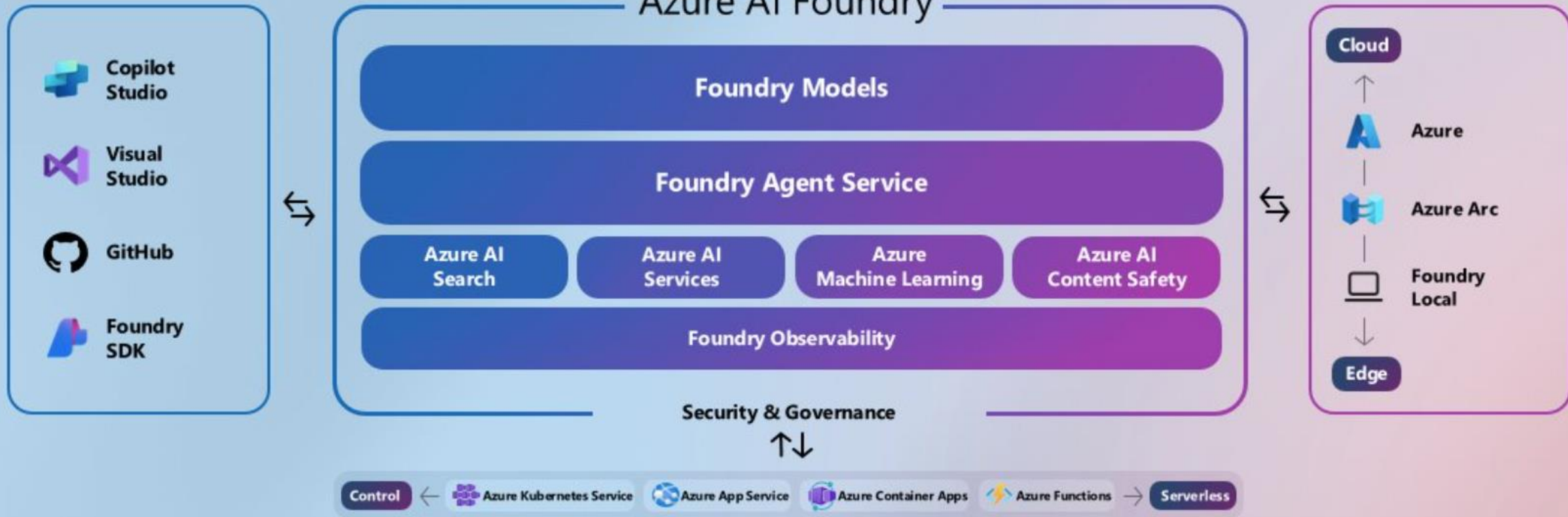
X: @Cihan_Cinar_

<https://blog.cihan-cinar.com/>

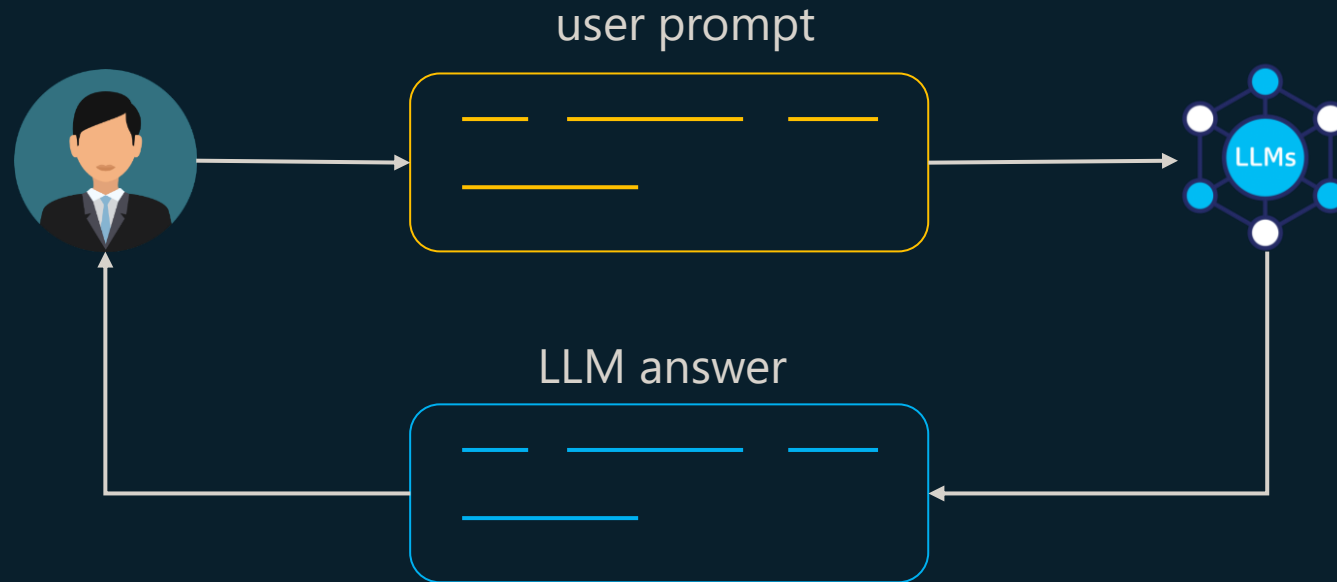




Azure AI Foundry

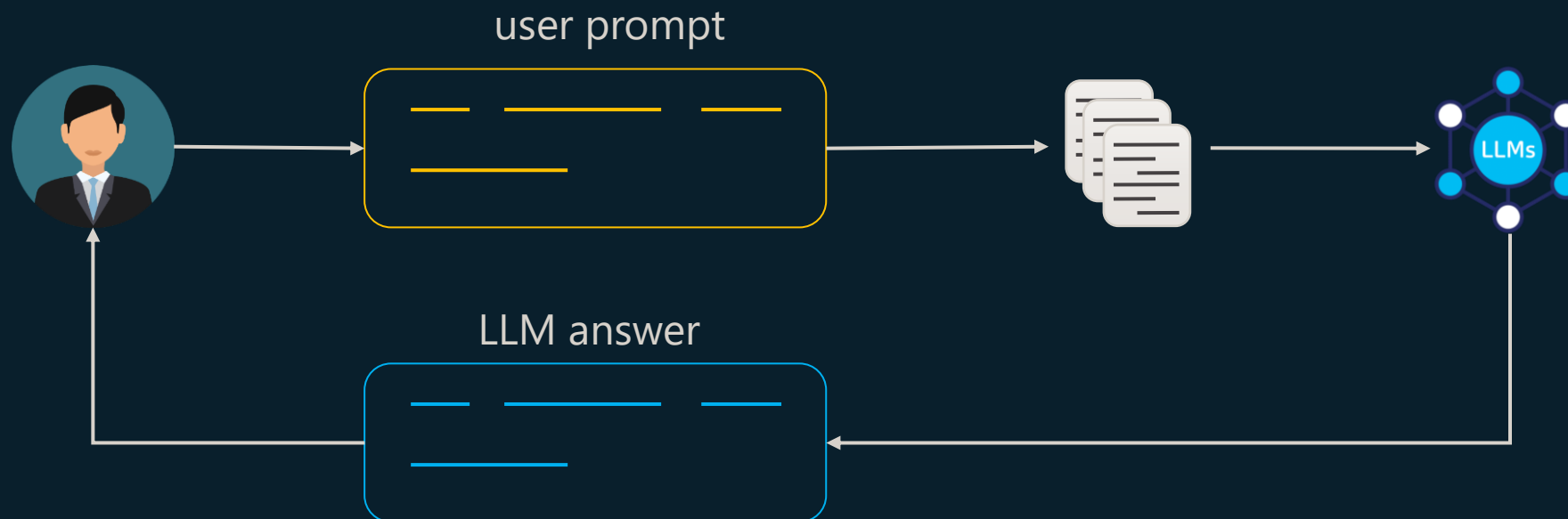


Chat with LLM



Limited to the LLMs knowledge and user input

LLM Chat with RAG



Relevant answers, but limited to data sources

 Microsoft

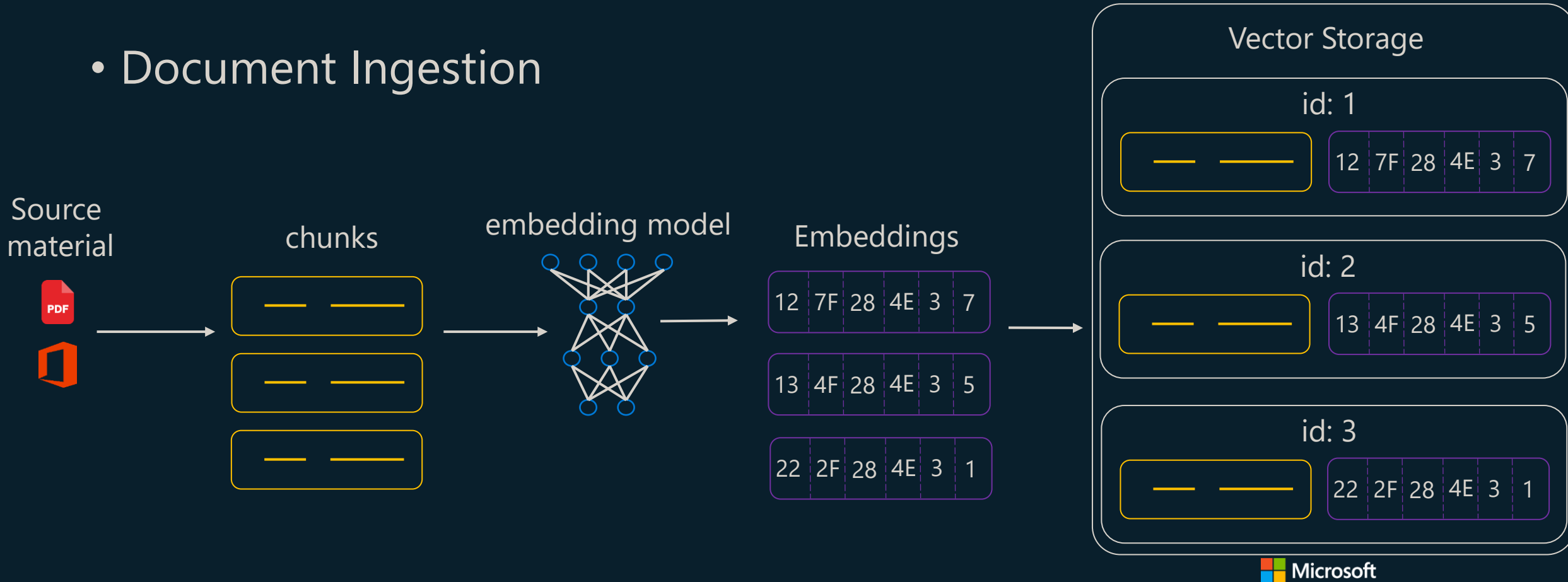
Reactor

How RAG Works (Retrieval-Augmented Generation)

- Document ingestion
- Transform, the user query into a vector
- Search in vector DB for relevant documents
- Add found documents to the context

RAG (Retrieval-Augmented Generation)

- Document Ingestion

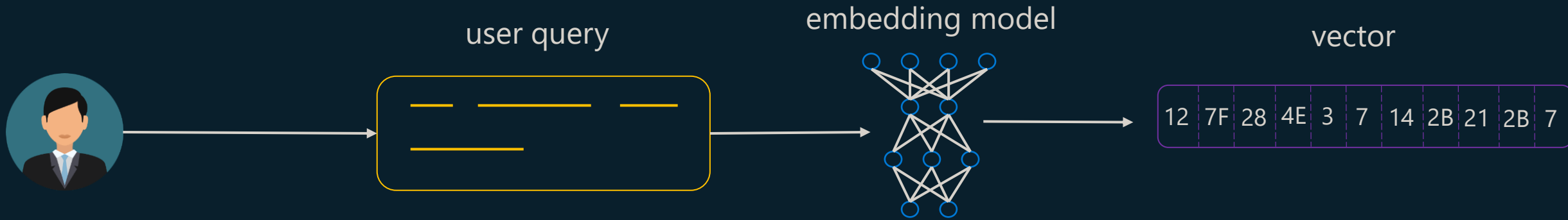


Microsoft

Reactor

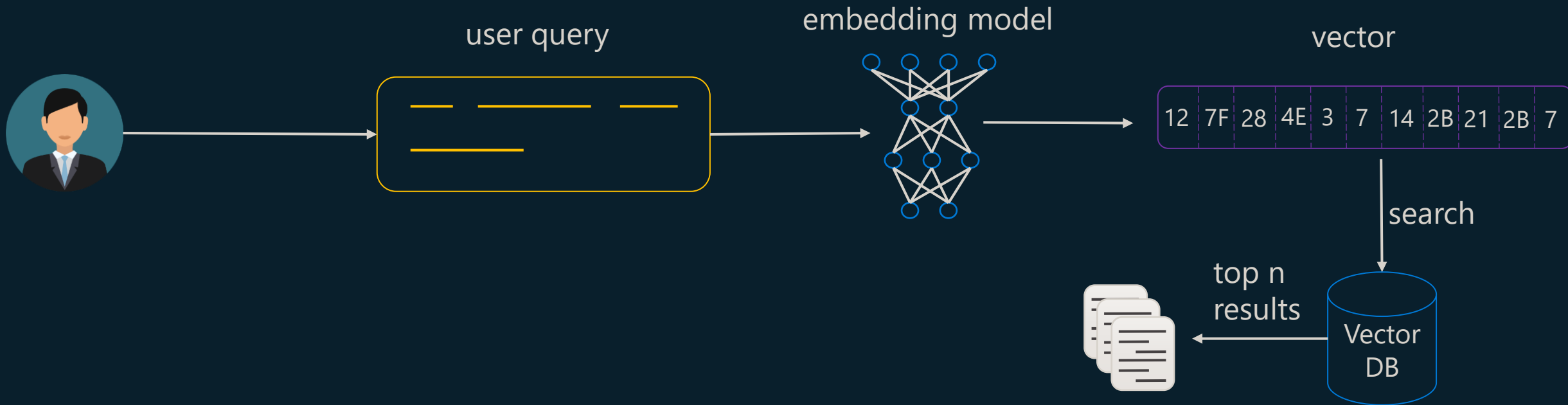
RAG (Retrieval-Augmented Generation)

- Transform, the user query into a vector



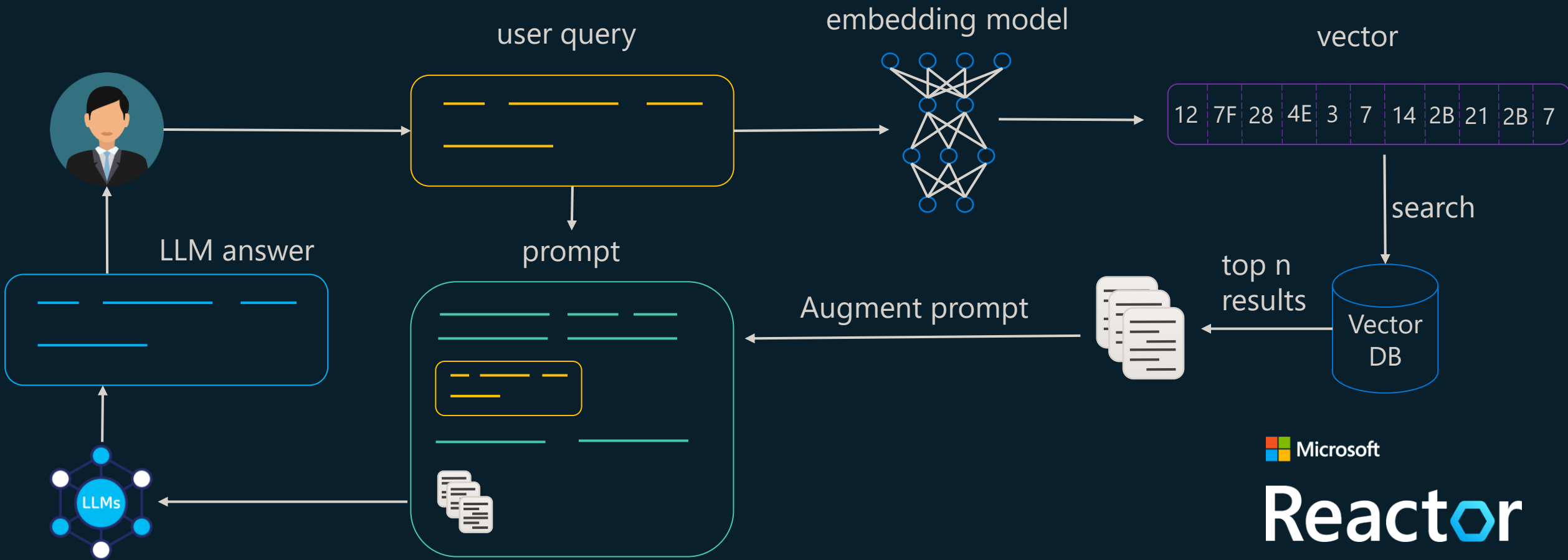
RAG (Retrieval-Augmented Generation)

- Search in vector DB for relevant documents

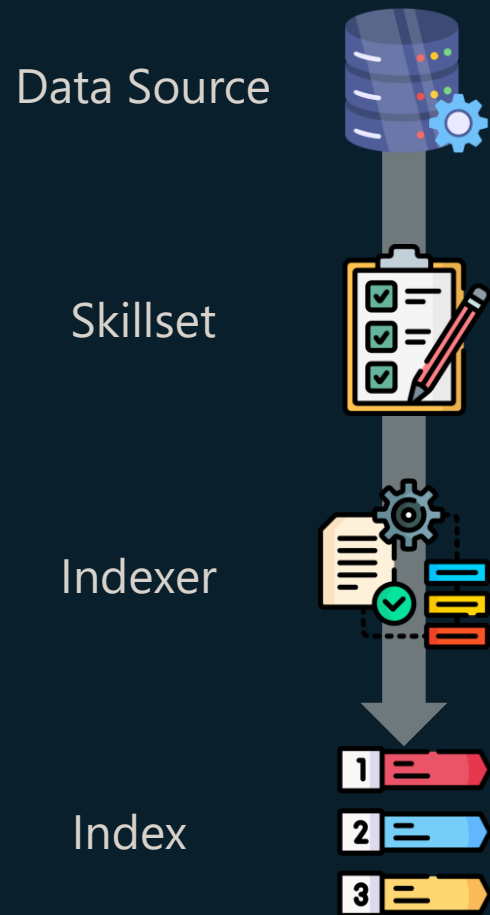


RAG (Retrieval-Augmented Generation)

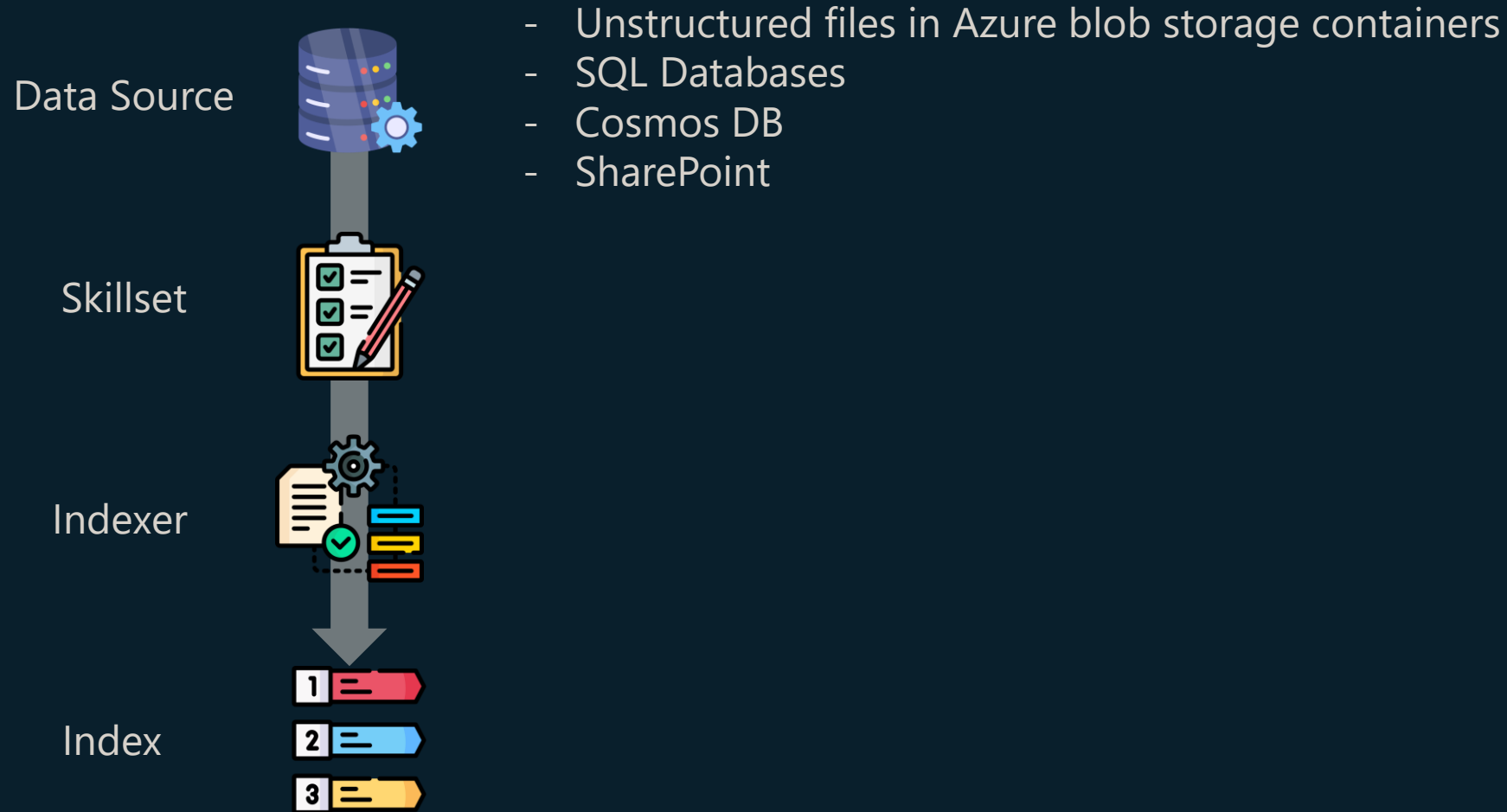
- Add found documents to the context



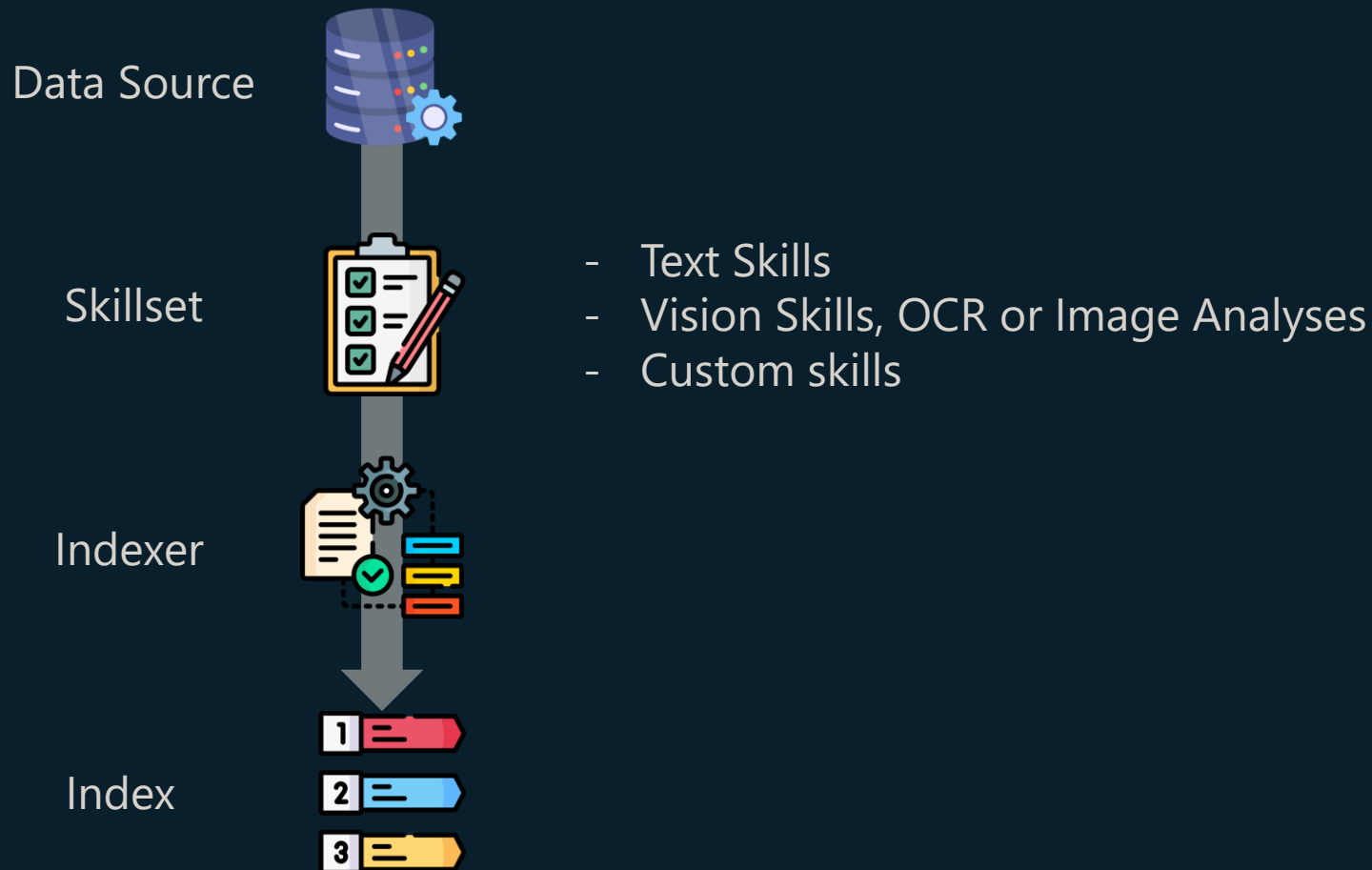
Azure AI Search - Big Picture



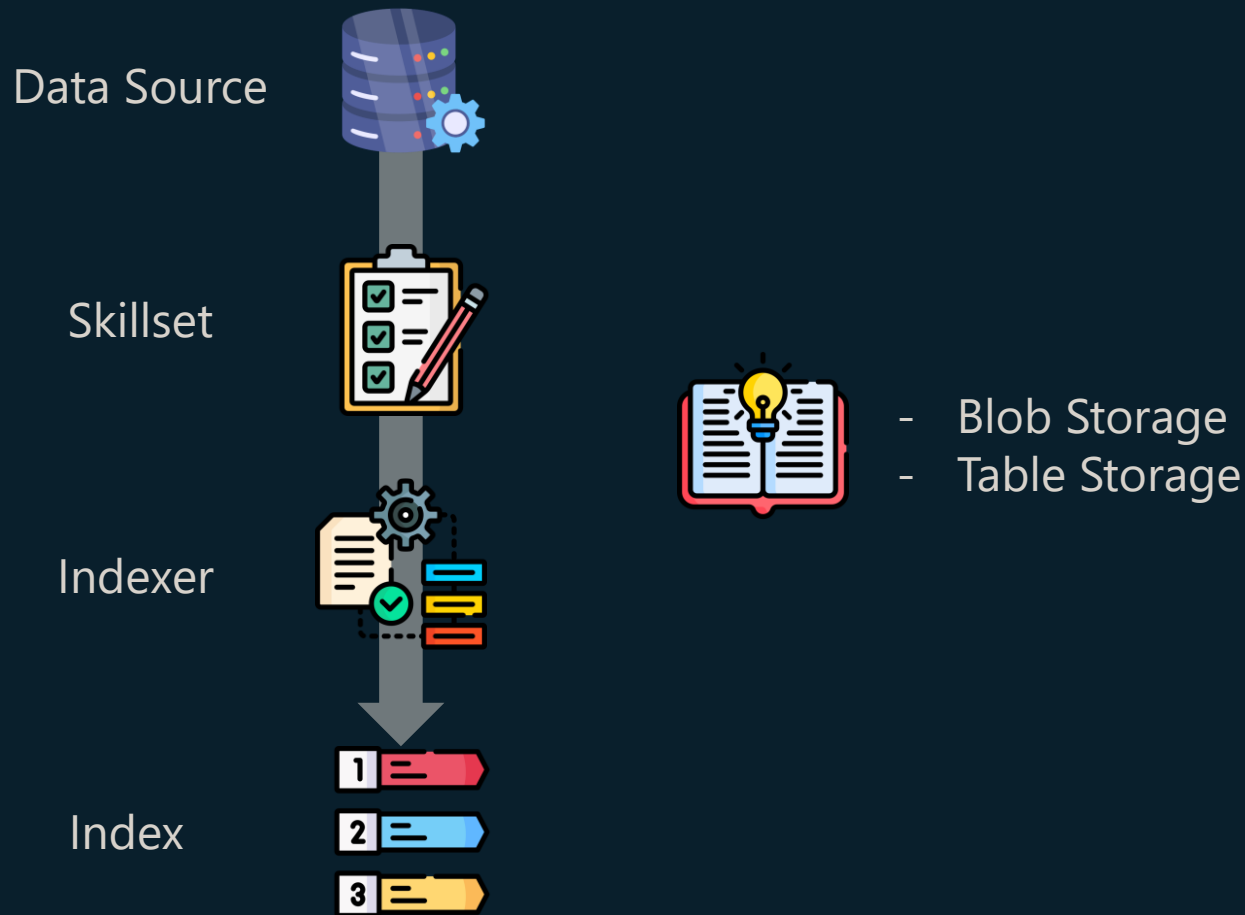
Azure AI Search - Data Source



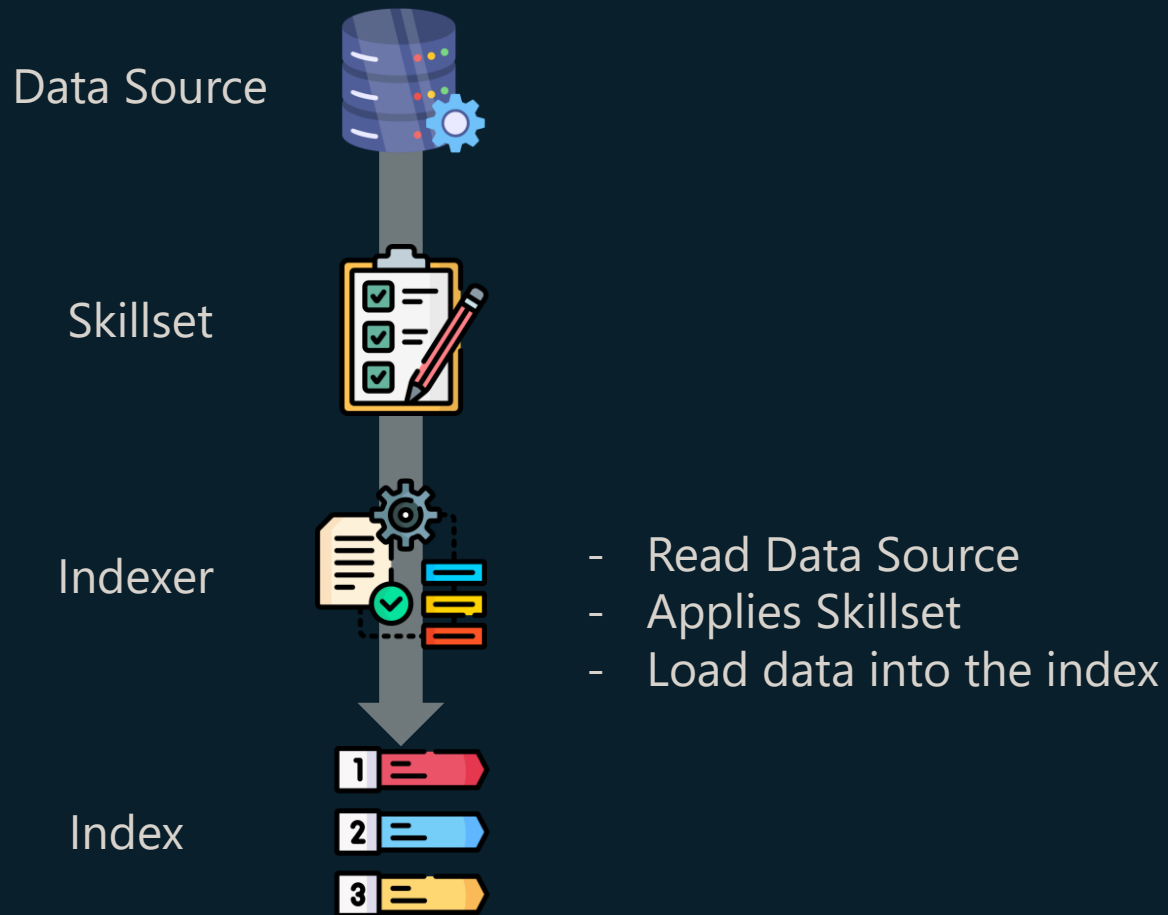
Azure AI Search - Skillset to define enrichment pipeline



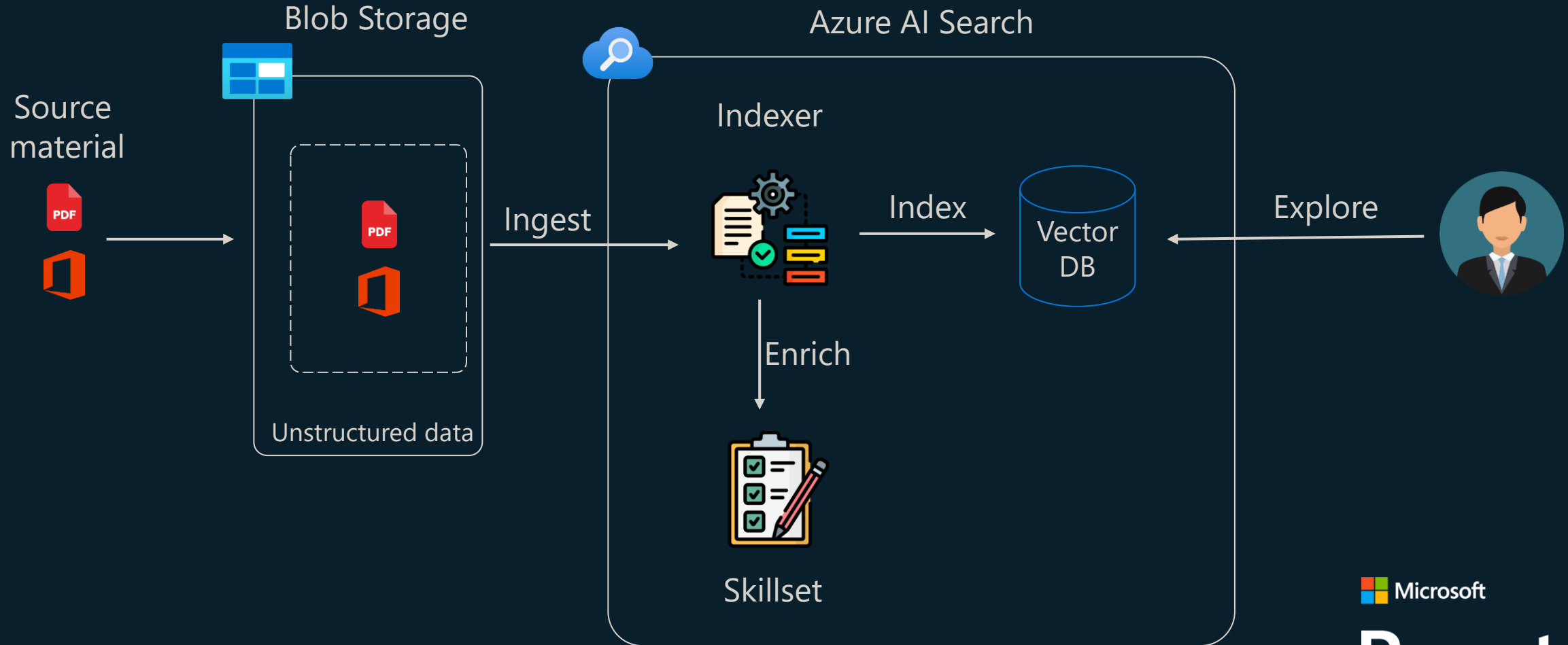
Azure AI Search - Persist enriched data in a knowledge store



Azure AI Search - Indexer



Azure AI Search



Azure AI Search - Definition

- **Data Source:** Persists connection information to source data, including credentials. A data source object is used exclusively with indexers.
- **Skillset:** A complete set of instructions for manipulating, transforming, and shaping content, including from image files. Except for very simple and limited structures, it includes a reference to an Azure AI services resource that provides enrichment.
- **Knowledge store:** Stores output from an AI enrichment pipeline in tables and blobs in Azure Storage for independent analysis or downstream processing.
- **Indexer:** A configuration object specifying a data source, target index, an optional AI skillset, optional schedule, and optional configuration.
- **Index:** Physical data structure used for search queries.

Azure AI Search - Full Text Search Params

- **search:** A search expression that includes the terms to be found.
- **queryType:** The Lucene syntax to be evaluated (simple or full).
- **searchFields:** The index fields to be searched.
- **select:** The fields to be included in the results.
- **searchMode:** Controls the default operator for Boolean queries: any (default) or all.

Azure AI Search - Search request

POST /indexes/hotels/docs/search?api-version=2024-07-01

```
{  
  "search": "Spacious, air-condition* +\"Ocean view\"",  
  "searchFields": "description, title",  
  "searchMode": "any",  
  "filter": "price ge 60 and price lt 300",  
  "orderby": "geo.distance(location, geography'POINT(-159.476235 22.227659)')",  
  "queryType": "full"  
}
```

AI Agent Service in Action

Step 1
Create an Agent

Step 2
Create a Thread

Step 3
Run the Agent

Step 4
Check the Run status

Step 5
Display the
Agent's Response

 **Agent**
Microsoft Sales Agent

Instruction
You are an advanced sales analyses agent for Microsoft, specializing in assisting users with sales data inquiries

Model


Your data (optional)
 Azure AI Search
 Files (local or Azure Blob)

Tools (optional)
File Search
Code Interpreter
Function Calling
Bing Search
Microsoft SharePoint
Microsoft Fabric
Azure Logic Apps
Azure Functions
OpenAI 3.0 specified tools

Thread
Sales analysis

 **User's message**
Tell me the total sales by region

 **Agent's message**
Here is the sales:
Europe: \$1923
America: \$1776

 **User's message**
Show a pie chart

 **Agent's message**


Run 1

1 Function Calling Tool
Query SQLite DB

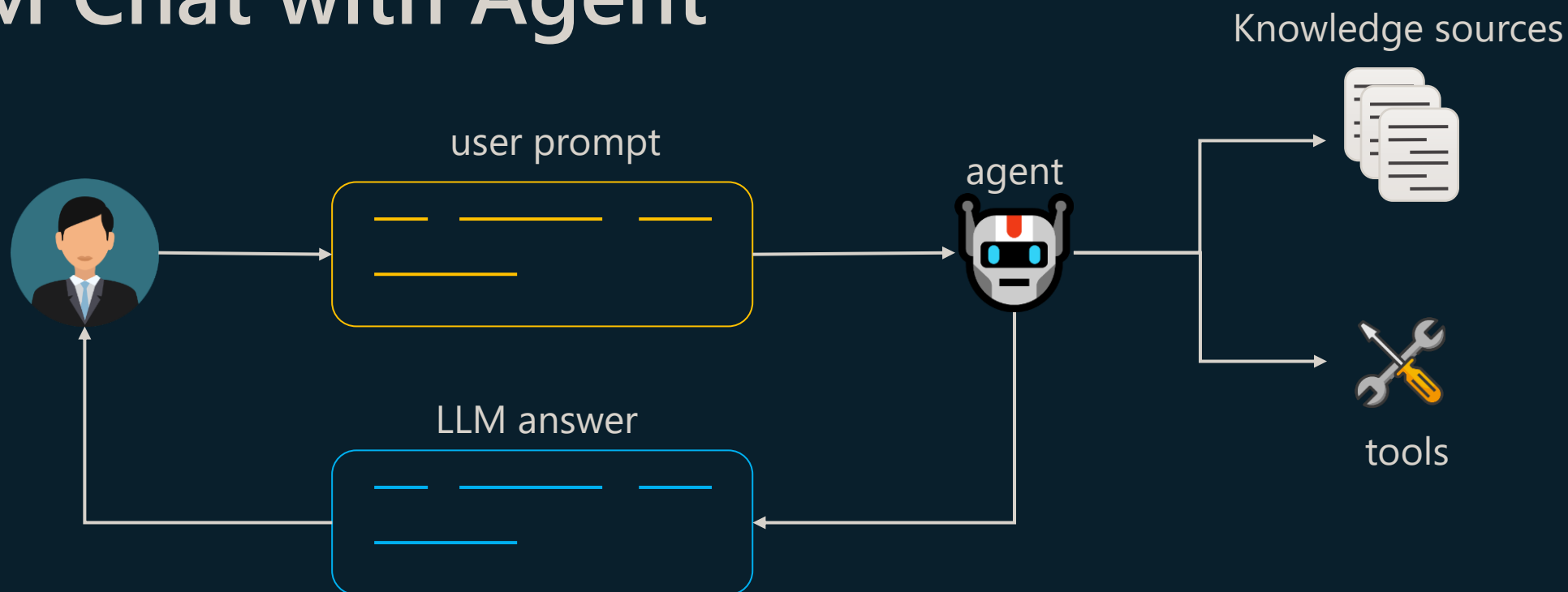
2 Create message

Run 2

1 Code Interpreter Tool
Create a pie chart

2 Create message

LLM Chat with Agent

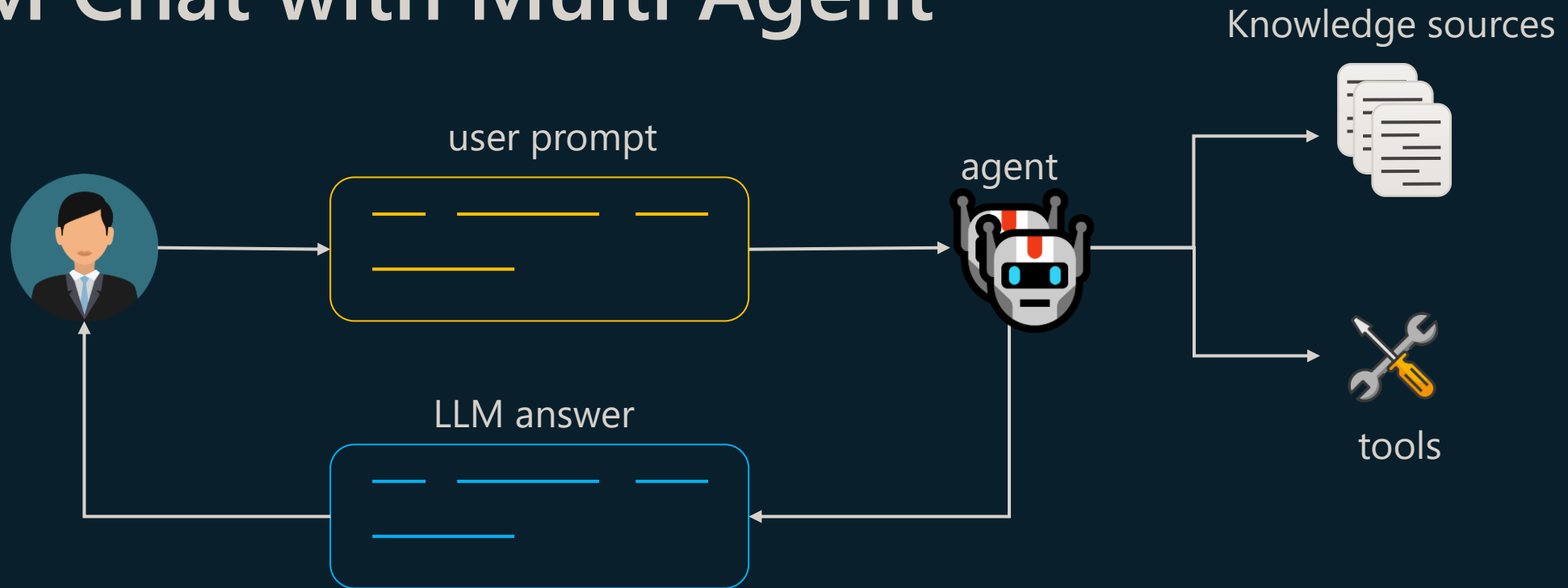


Agent perform complex tasks

Microsoft

Reactor

LLM Chat with Multi Agent



Specific assigned task agent

Microsoft

Reactor



Reactor

Demo

RAG Agent



SDK

Language	SDK
Java	Azure SDK for Java, Semantic Kernel, Spring AI, LangChain4j
Python	Azure SDK for Python, Semantic Kernel, AutoGen, LangChain
C#	Azure SDK for .NET, Semantic Kernel
JS	Azure SDK for Js, LangChainJs, Vercel AI SDK

Resources

- <https://github.com/Azure-Samples/>
- <https://github.com/Azure/GPT-RAG>
- <https://aka.ms/ragchat>
- <https://aka.ms/June23AzureAISearchNew1>
- <https://aka.ms/June23RAGIndex1>



Reactor

Thank You

Linkedin: @cihancinar

Github: @cihancinar

X: @Cihan_Cinar_

<https://blog.cihan-cinar.com/>

Cihan Cinar

Software & AI Architect

