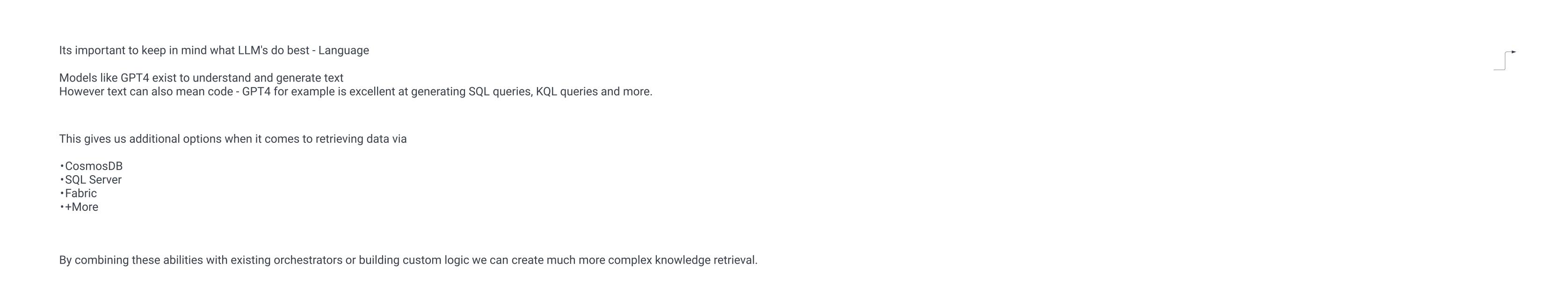
https://platform.openai.com/tokenizer

CONTEXT IS EVERYTHING - INTEGRATION IS HOW WE GET CONTEXT

• Integration patterns are critical for maintaining conversation context with Azure OpenAI by enabling stateful services or event-driven architectures to store and persist session data across interactions. • Utilizing integration with databases and APIs allows the model to retrieve relevant data dynamically, ensuring conversations are personalized and contextual. • These patterns help deliver a continuous, engaging, and coherent conversation flow, enhancing the user experience in AI-powered systems.

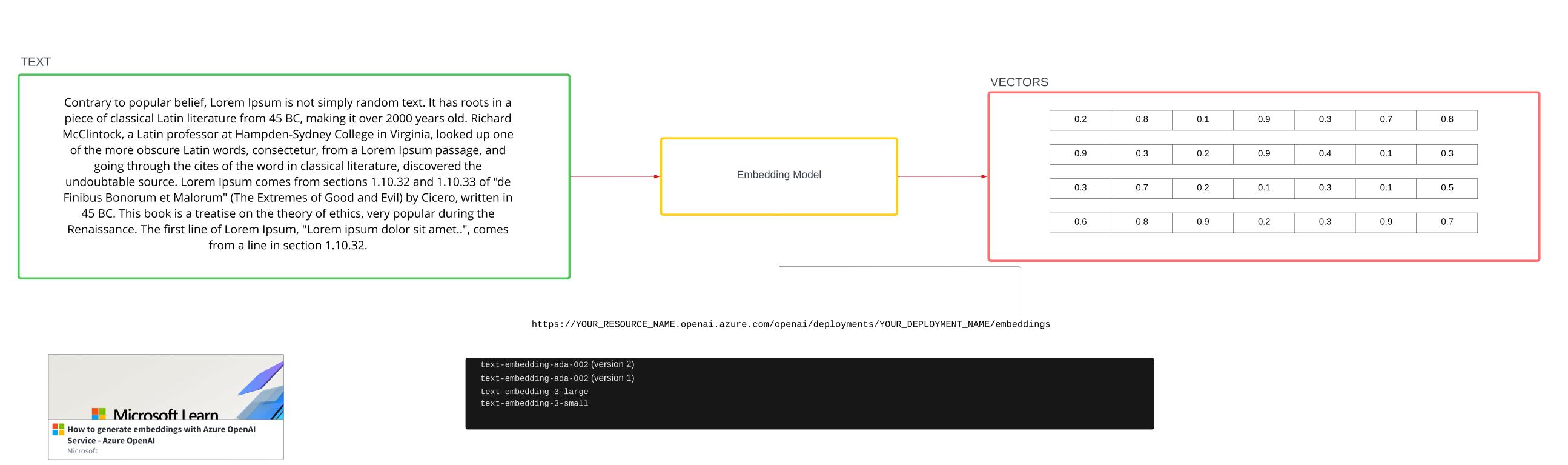
CHAT / COMPLETIONS TOOLS OpenAl tools are a set of APIs and functionalities designed to integrate artificial intelligence models into applications, enabling features like natural language understanding, code generation, and content creation. "type": "function", "function": { "function": { "name": "get_good_emails", "description": "Get good and high quality email examples that have been used in the past", "name": "get_stock_price", "description": "Get the current stock information for a given stock symbol. Only Stock symbol supported is IBM", "parameters": { "type": "object", "parameters": { "type": "object", "properties": { "howMany": { "properties": { "symbol": { "type": "number", "description": "how many good emails to return. Maximum is 5", "type": "string", "description": "The stock symbol for which to get the information. Example: IBM", "required": ["howMany"], "required": ["symbol"], **User Question** 1 question text What is the stock price of IBM? Response: tool call Original Question + Tool Result APPLICATION LOGIC (actually executing the function) answer text Stock price of IBM is 399\$ LLM Analyzes user question + Azure Search TEXT results to formulate an answer

Data Retrieval for LLM's

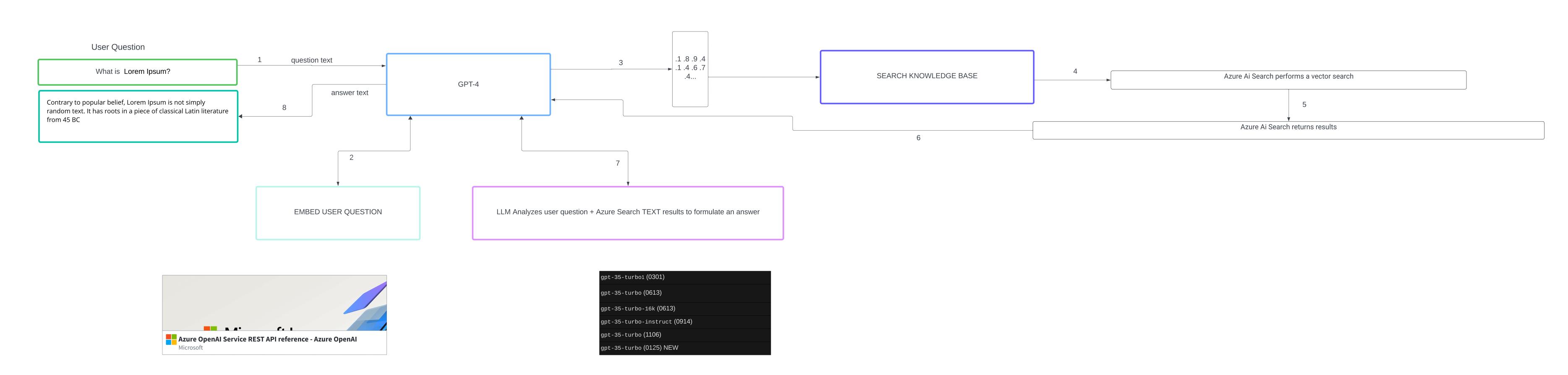


EMBEDDING

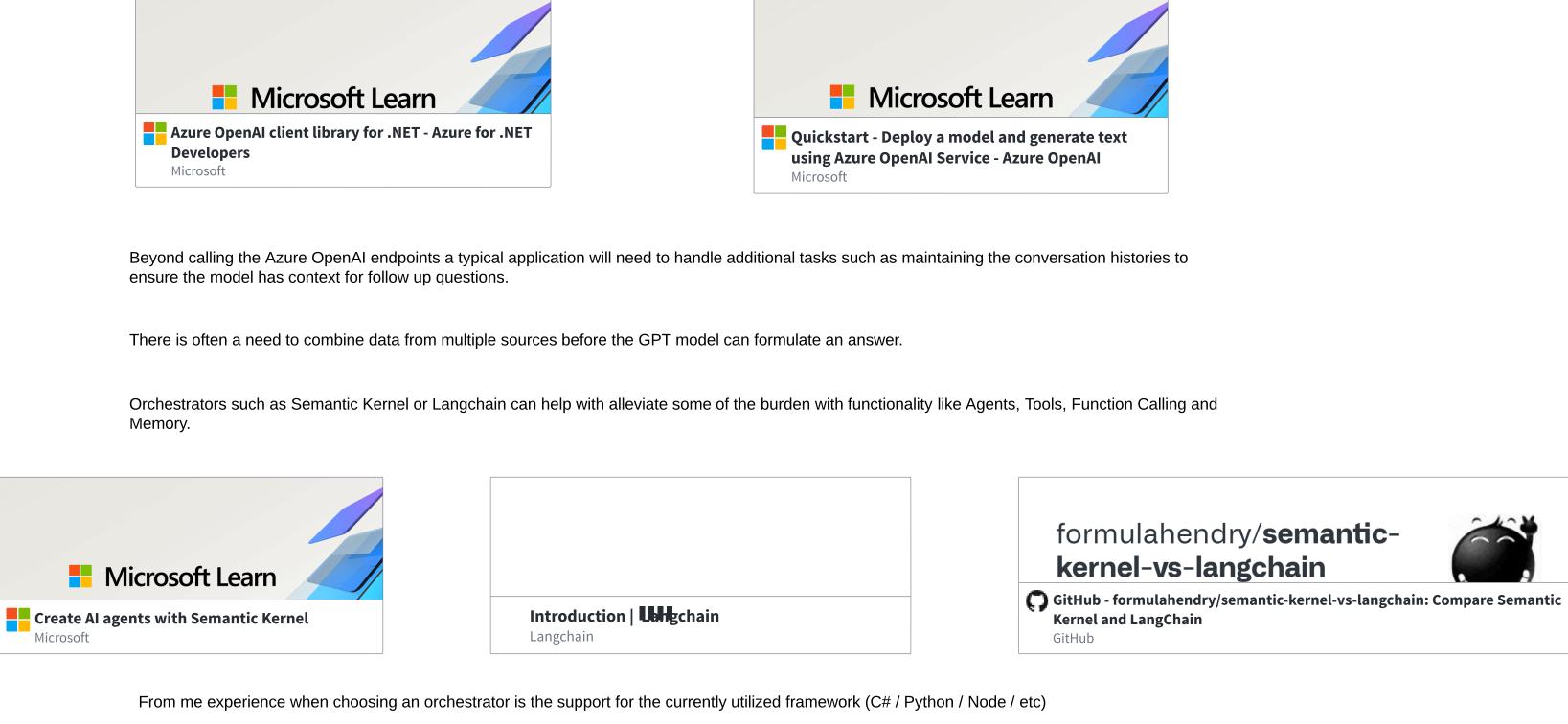
Embeddings are numerical representations of text, images, or other data, where similar items are positioned closely in a high-dimensional space. They capture underlying patterns or features, enabling tasks like similarity search, classification, and recommendations.



GROUNDED COMPLETIONS / CHAT Text completion models take a list of messages as input and return a model-generated message as output.



ORCHESTRATORS Any of the above models can be interacted with via API calls to the Azure OpenAI service or through provided SDK's.



Models like GPT4 exist to understand and generate text

This gives us additional options when it comes to retrieving data via

Both Semantic Kernel and LangChain support Python.

Out of the two, Semantic Kernel is the clear choice for dotnet / c# environments.

CosmosDBSQL ServerSnowflake

• +More

Data Retrieval for LLM's Its important to keep in mind what LLM's do best - Language

However text can also mean code - GPT4 for example is excellent at generating SQL queries, KQL queries and more.

User Question question text SEMANTIC KERNEL What are the top 10 purchased products that in their reviews mention the customer being thrilled? Tools/Plugins/Extensions/Functions Tool Description: "Ability to retrieve data from sql" Tool Function: Execute query on SQL database Useful for question Based on available tools, I will The top purchased products where users described feeling Tool Description: "Ability to retrieve user reviews" answer text 1. Create and execute query that will retrieve top sold products thrilled are ... Tool Function: Execute query on AI Search with filters as parameters to the function Azure Ai Search 3.I will perform a vector query on Azure AiSearch agains the vector of 'thrilled' and pass product ID's as the search filter Useful for question 4.1 will return results to the user Tool Description: "Provides Current Weather give zip" Tool Function: Execute query on AI Search with filters as parameters to the function Weather API

By combining these abilities with existing orchestrators or building custom logic we can create much more complex knowledge retrieval.