Leveraging Azure AI and Python for Data-Driven Decision Making

Taiob Ali
He/Him/His

Image generated by Microsoft Copilot. This image is for illustrative purposes only and may not represent actual products or services.

Taiob Ali

Database Solutions Manager, GMO LLC

http://sqlworldwide.com/

/sqlworldwide

@sqlworldwide.bsky.social

taiob@sqlworldwide.com

OUR SPONSORS

MICROSOFT
Global Sponsor

IDERA SOFTWARE
Platinum Sponsor

VISCOSITY NA
Gold Sponsor

RED GATE SOFWARE
Bronze Sponsor

DBHAWK
Bronze Sponsor

COZYROC
Bronze Sponsor

OUR VOLUNTEERS

Anuja Nair

Mohit Aragwal

Olga Nesterova

Huyen Vu

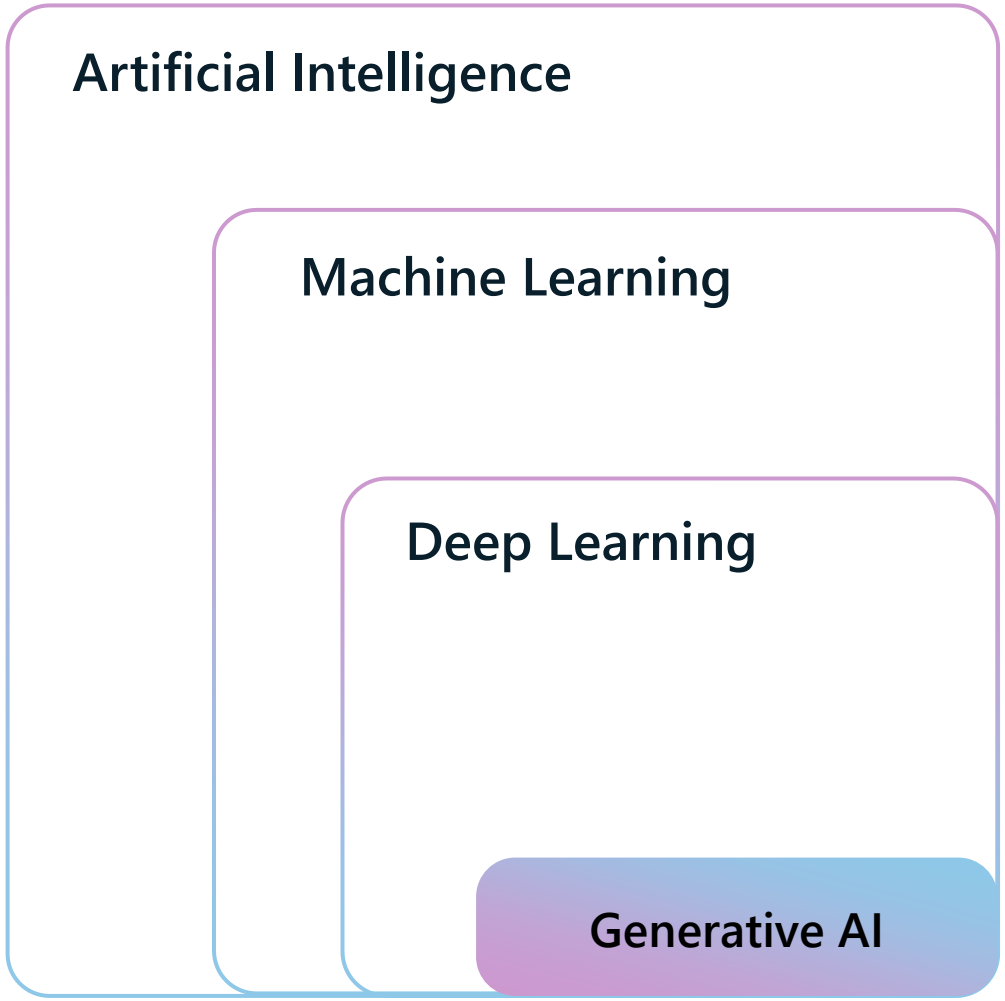
Qin Xu

Senthil Nagarajan

Goals

- Generative AI evolution
- Understand definitions of Generative AI (GenAI) key terms
- Why GenAI and **Your SQL Database?**
- Demo
  - Natural language to SQL in the Azure portal query editor (preview)
  - Microsoft Copilot skills in Azure SQL Database (preview)
  - Keyword search with Large Language Model (LLM)
  - Integrated Vectorization, Semantic ranker with Azure SQL Database
  - Vector Search with Vector Functions
  - Utilizing large language models and Python notebooks against Azure SQL Database

This is a new moment for AI



1950s

Artificial Intelligence

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

1959

Machine Learning

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

2017

Deep Learning

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

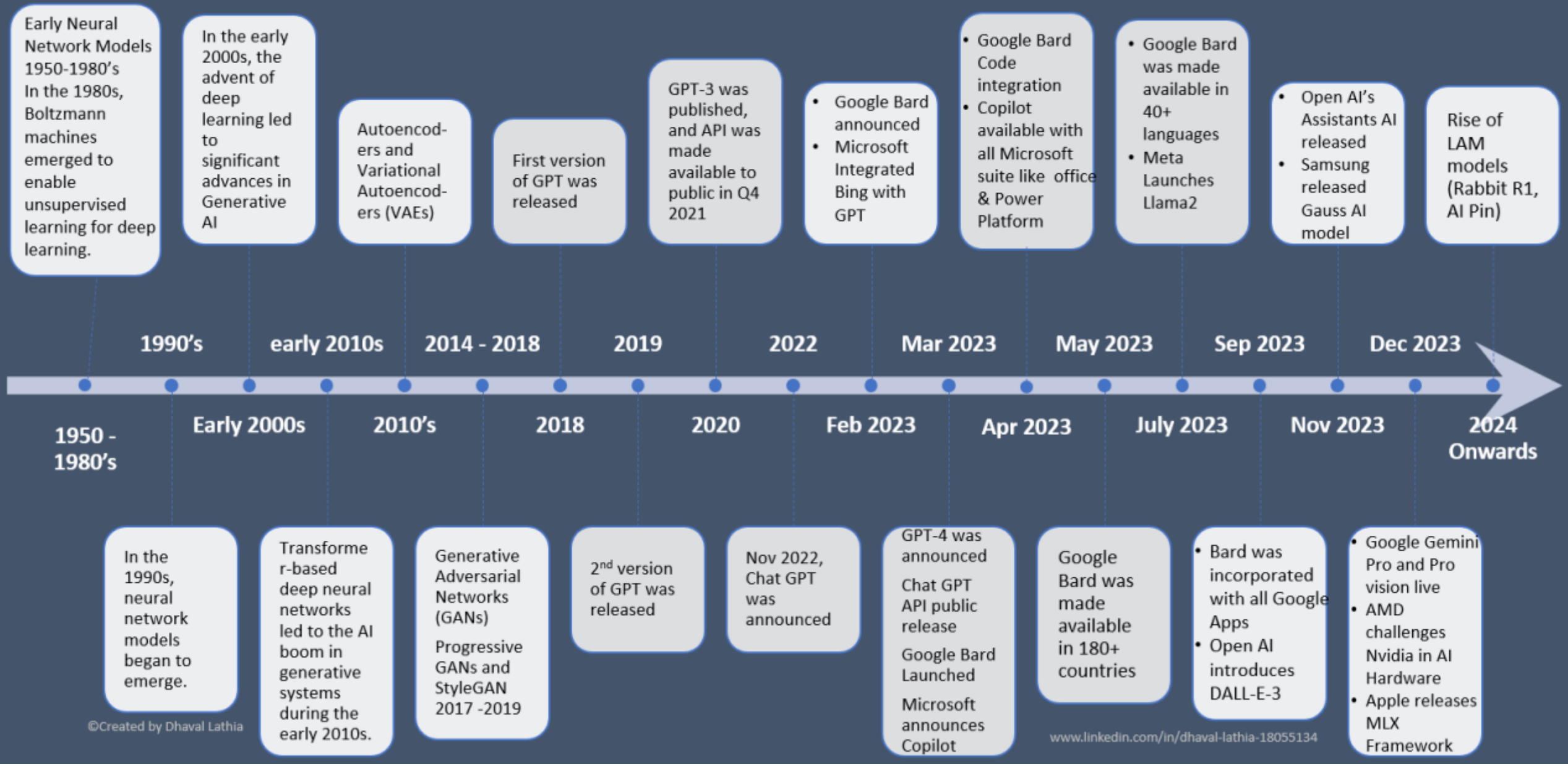
2021

Generative AI

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

This slide is copied from:
https://github.com/Azure-Samples/SQLAIWorkshop/tree/main/slides

Evolution of Generative AI



https://www.linkedin.com/pulse/evolution-generative-ai-dhaval-lathia-cfnie/

PROMPT

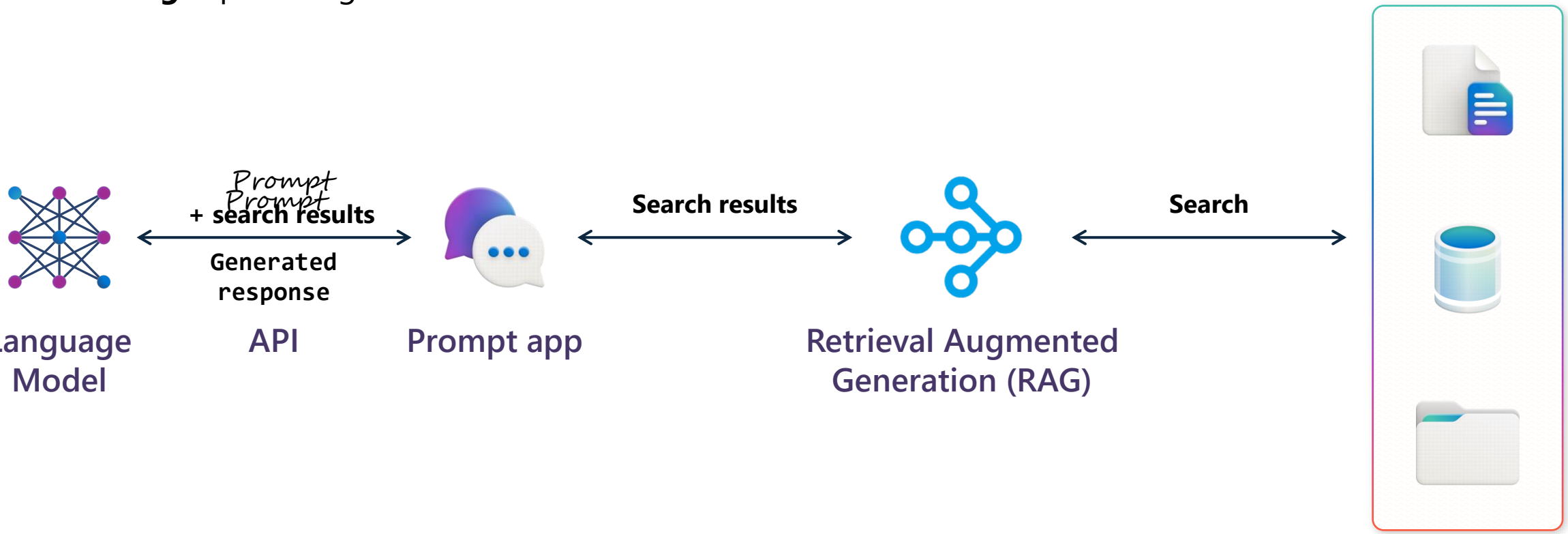
Prompt Type	Definition	Example
Zero-Shot Prompts	Tells the model: "Do the thing I want."	Create a recipe for chocolate chip cookies. List all ingredients and steps.
Few-Shot Prompts	Tells the model: "Do the thing I want, here are some examples."	Create a recipe for chocolate chip cookies. Here's an example recipe for oatmeal raisin cookies. Now, create the chocolate chip cookie recipe.
Chain-of-Thought (CoT) Prompts	Directs the model to reason before answering.	Imagine you are creating a recipe. Think through the ingredients and steps before finalizing the recipe. Now, write the complete recipe for chocolate chip cookies.
Multi-Turn Prompts	Involves a back-and-forth interaction where the model builds on previous responses.	User: How do I make chocolate chip cookies? AI: First, gather your ingredients. User: Great, what's the first step? AI: Preheat your oven to 350°F.
Retrieval-Augmented Generation (RAG) Prompts	Involves retrieving information from a database or document to help generate a response.	Using the provided document on baking techniques, answer: What are the best practices for making chewy chocolate chip cookies?


PROMPT
+
GROUNDING

RETRIEVAL AUGMENTED GENERATION (RAG)

Let’s get grounded on prompts

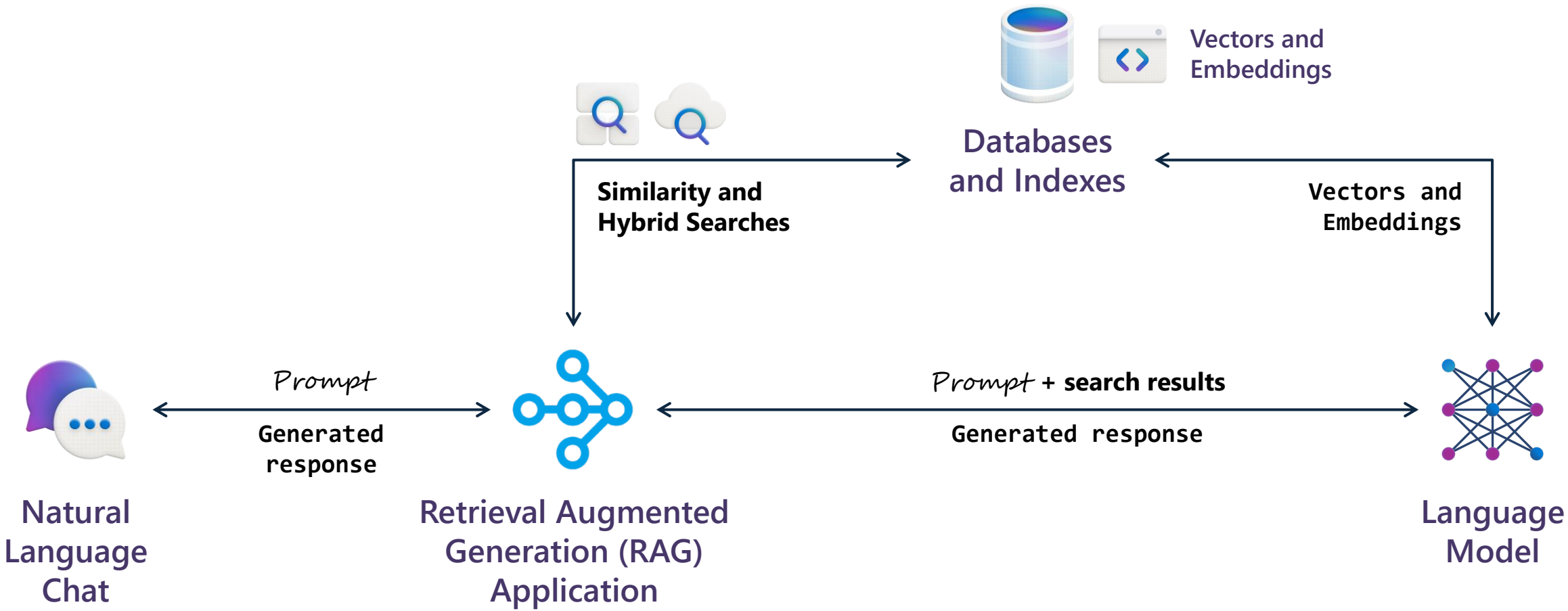
Fine tuning = customizing a trained model
Grounding = providing additional context to a trained model



Prompt Engineering
Better prompt = better responses

This slide is copied from:
https://github.com/Azure-Samples/SQLAIWorkshop/tree/main/slides

Building Generative AI applications with databases



This slide is copied from:
https://github.com/Azure-Samples/SQLAIWorkshop/tree/main/slides

VECTOR

EMBEDDING

Vectors and Embeddings

Feature Vector

Ordered array of numbers typically created by a human to train a model

[Height, Weight, Age, Fur Length, Energy Level]

"dog"

Embedding

Vector generated by a model that has semantic meaning

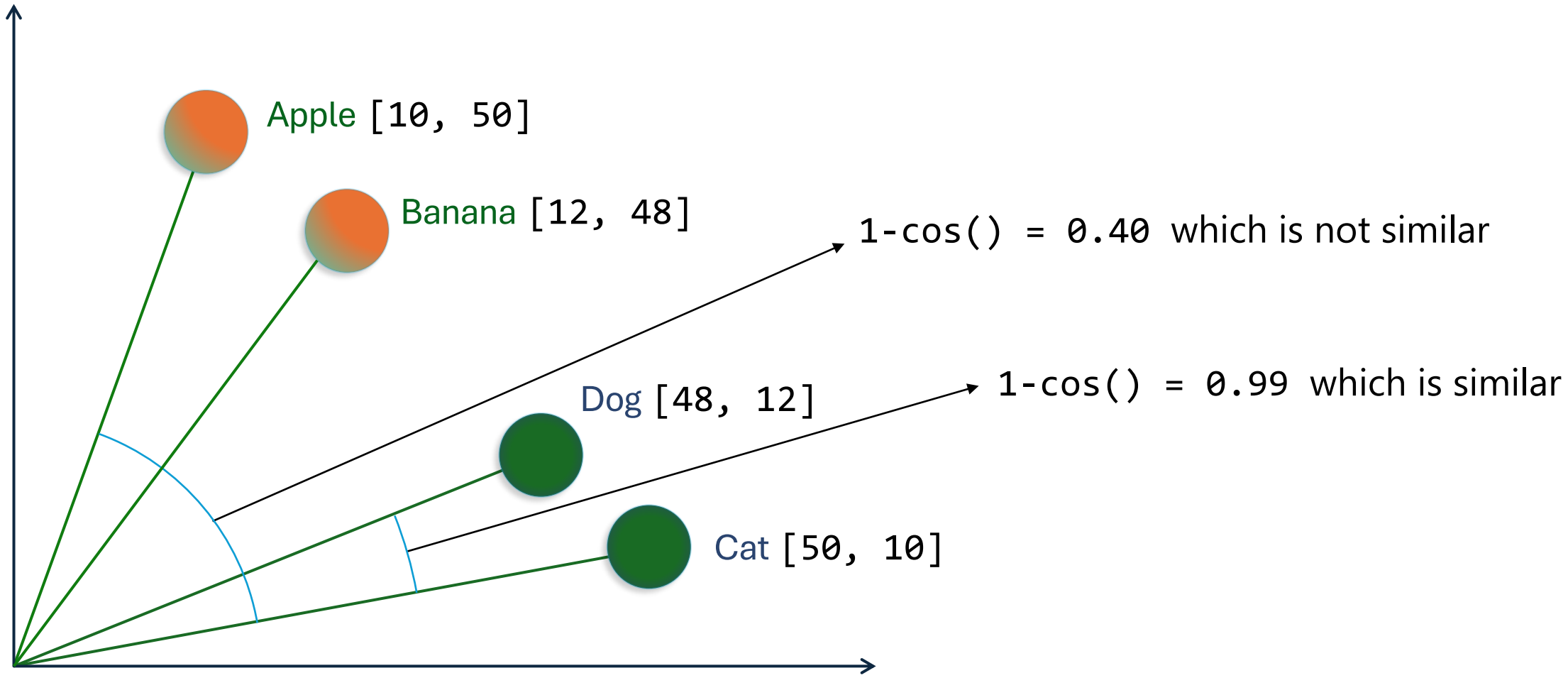
```
dog: [0.9, 0.3, 0.2, ...]
Image of a dog: [0.9, 0.3, 0.2, ...]
puppy: [0.88, 0.33, 0.21, ...]

"I took the dog for a walk": [1.5, -0.8, 2.1, ...]
```

 "dog"

This slide is copied from:
https://github.com/Azure-Samples/SQLAIWorkshop/tree/main/slides

Similarity Searching using Cosine



This slide is copied from:
https://github.com/Azure-Samples/SQLAIWorkshop/tree/main/slides

TOKEN

Search option	Retrieval type	Additional pricing?	Benefits
keyword	Keyword search	No additional pricing.	Performs fast and flexible query parsing and matching over searchable fields, using terms or phrases in any supported language, with or without operators.
semantic	Semantic search	Additional pricing for semantic search usage.	Improves the precision and relevance of search results by using a reranker (with AI models) to understand the semantic meaning of query terms and documents returned by the initial search ranker
vector	Vector search	Additional pricing on your Azure OpenAI account from calling the embedding model.	Enables you to find documents that are similar to a given query input based on the vector embeddings of the content.
hybrid (vector + keyword)	A hybrid of vector search and keyword search	Additional pricing on your Azure OpenAI account from calling the embedding model.	Performs similarity search over vector fields using vector embeddings, while also supporting flexible query parsing and full text search over alphanumeric fields using term queries.
hybrid (vector + keyword) + semantic	A hybrid of vector search, semantic search, and keyword search.	Additional pricing on your Azure OpenAI account from calling the embedding model, and additional pricing for semantic search usage.	Uses vector embeddings, language understanding, and flexible query parsing to create rich search experiences and generative AI apps that can handle complex and diverse information retrieval scenarios.

https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/use-your-data?tabs=ai-search%2Ccopilot#search-types

Why use LLM over built-in SQL Server Semantic Search?

- **Natural Language Understanding**: Excels at interpreting context and nuances in user queries.
- **Flexibility and Adaptability**: Can be fine-tuned for specific tasks and domains, offering high accuracy.
- **Enhanced Data Integration**: Processes data from multiple sources, including unstructured data.
- **Advanced Analytics**: Performs sentiment analysis, entity recognition, summarization, and more.
- **Interactive and Conversational**: User-friendly for non-technical users, enabling intuitive interactions.
- **Scalability**: Efficiently handles large volumes of data and queries, suitable for enterprise-level applications.

DEMO

Azure Portal
Azure Data Studio
Visual studio Code

Reference

- SQL AI Workshop
- Intelligent applications
- SQL-AI-samples
- azure-sql-db-vector-search
- Vector Search with Azure SQL Database by Muazma Zahid
- Integrated data chunking and embedding in Azure AI Search
- Vector Similarity Search with Azure SQL database and OpenAI by Davide Mauri
- VECTOR_DISTANCE (Transact-SQL) (Preview)
- Indexers in Azure AI Search
- Announcing the Public Preview of Integrated Vectorization in Azure AI Search by Gia Mondragon

linkedin.com/in/sqlworldwide

sqlworldwide.com

taiob@sqlworldwide.com

@sqlworldwide.bsky.social


THANK
YOU