

Microsoft Semantic Kernel & Azure OpenAI



Dr. Damir Dobric

Lead Software Architect daenet GmbH / ACP Digital
Microsoft Regional Director,
Most Valuable Professional: Azure + IoT

Damir Dobric Microsoft Semantic Kernel & Azure OpenAI



#ittage



IT-Tage
2023



GENERATIVE AI

What is a Model?

A *model* refers to a specific instance or version of an LLM AI

Available GPT Models

- Ada
 - basic natural language understanding
 - classification, sentiment analysis, summarization, and simple conversation.
- Babbage
 - more complex natural language tasks, such as reasoning, logic, arithmetic, and word analogy.
- Curie
 - advanced natural language tasks, such as text-to-speech, speech-to-text, translation, paraphrasing, and question answering.
- Davinci
 - Handles almost any natural language task,
 - as well as some multimodal tasks, such as image captioning, style transfer, and visual reasoning.
 - It can also generate coherent and creative texts on any topic, with a high level of fluency, consistency, and diversity.

What is a token?

- 1 token \approx 4 chars in English
- 1 token \approx $\frac{3}{4}$ words
- 100 tokens \approx 75 words
- Byte Pare Encoding (Gage,1994): [Wikipedia](#)
- Tokenizer: <https://platform.openai.com/tokenizer>
- [What are tokens and how to count them?](#)
- Token Pricing: [Pricing \(openai.com\)](#)

Open AI

<https://platform.openai.com/docs/models>

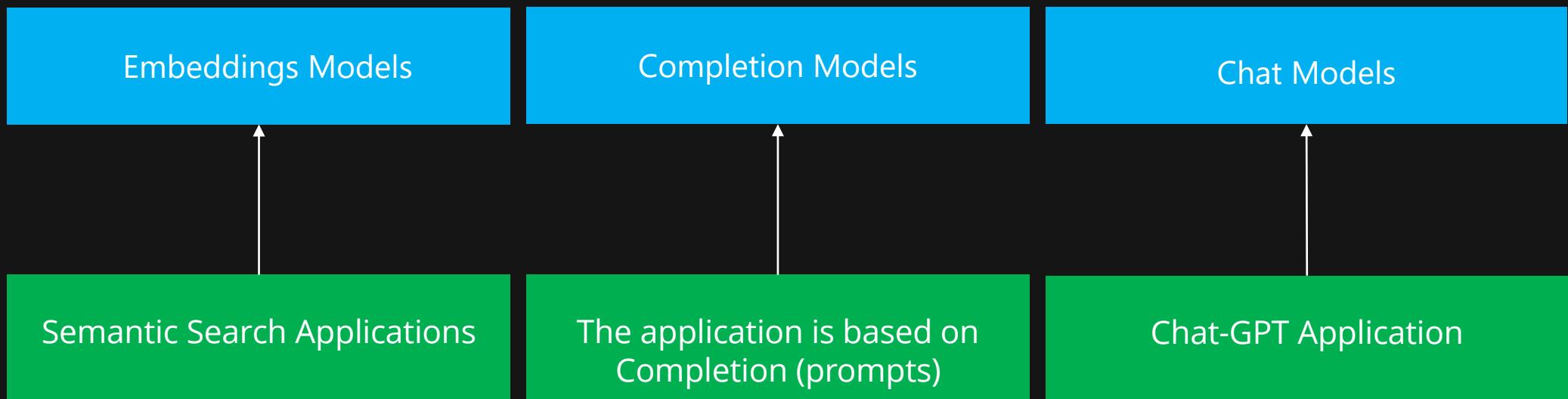
Azure Open AI

https://oai.azure.com/portal/***/models

DEMO

Open AI Playground Azure Open AI Playground

Generative Model Types



Completion Models

Summarize following scientific abstract, using a user-friendly language.

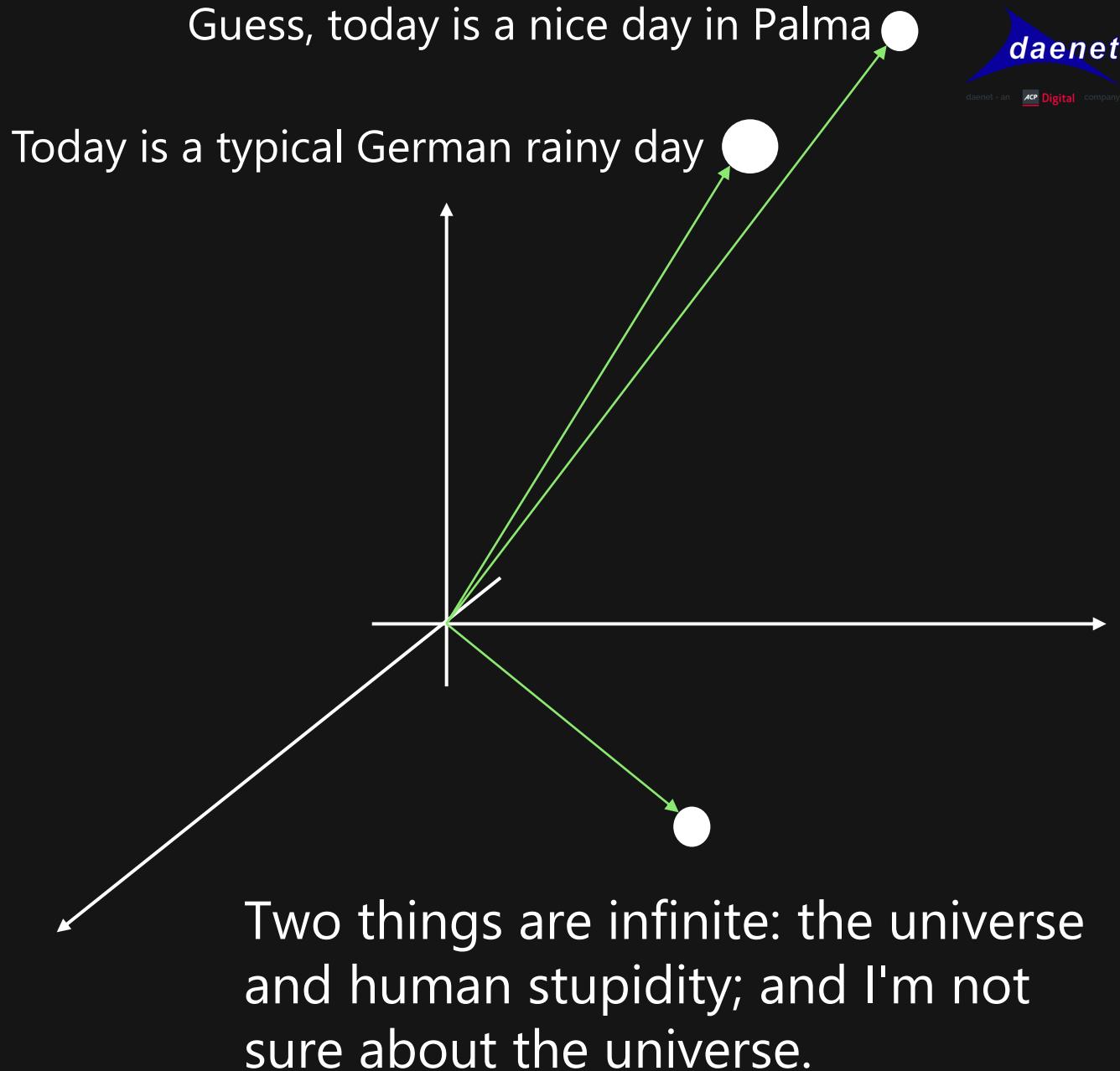
[Abstract]

Sparse representation has attracted much attention from researchers in fields of signal processing, image processing, computer vision, and pattern recognition. Sparse representation also has a good reputation in both theoretical research . . .

Chat Models

```
Bot: How can I help you?  
User: What's the weather like today?  
Bot: Where are you located?  
User: I'm in Frankfurt.  
Bot: It's 22 degrees and sunny in Frankfurt today.  
User: Thanks! Can you explain scientific abstracts?  
Bot: Yes. Please provide abstract.
```

Embedding Models



Similarity Between Multidimensional Vectors

- Dot Product
- The Norm
- Cosine Similarity

$$A \cdot B = a_1 \cdot b_1 + a_2 \cdot b_2 + \dots + a_n \cdot b_n$$

$$\|\mathbf{A}\| = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}$$

$$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \|\mathbf{B}\| \cos(\theta)$$

When to use Embeddings?

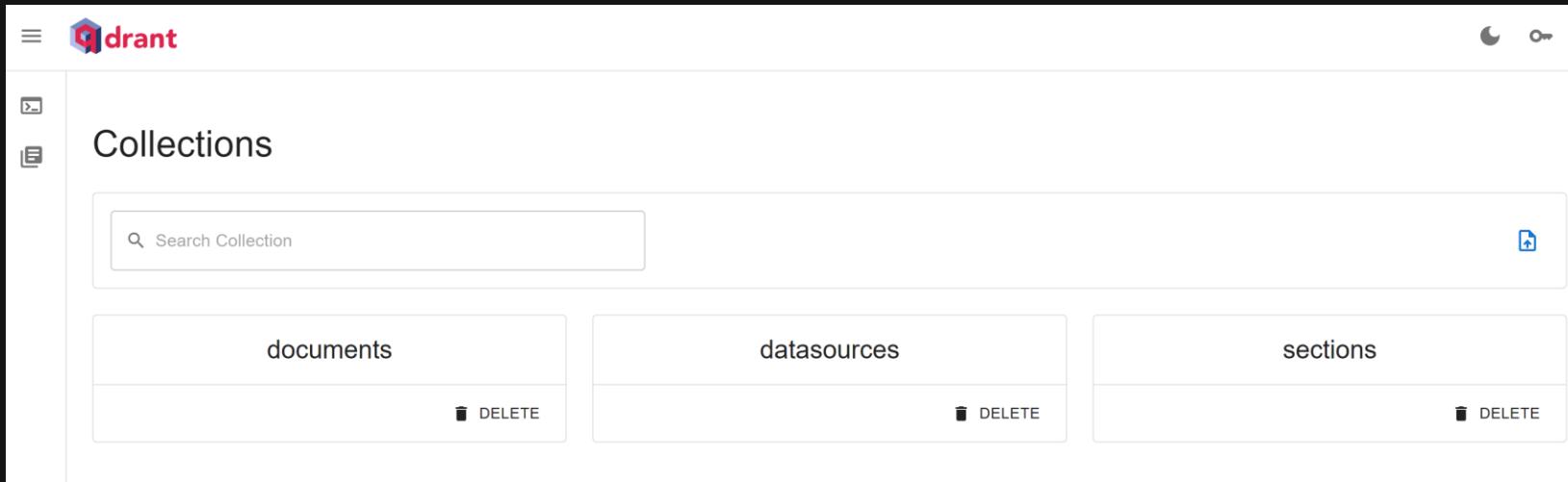
- Semantic Search
- Classification
- Grounding

DEMOs

- Consuming AI Service as a REST Endpoint
 - Using Completions
 - Generating Embeddings
- Using OpenAPClient
 - Chat models and more

qDrant DB

Vector Search



The screenshot shows the qDrant web interface. At the top left is the qDrant logo. On the far right are a dark mode toggle and a user profile icon. The main area is titled "Collections". It features a search bar labeled "Search Collection" with a magnifying glass icon. Below the search bar are three cards: "documents", "datasources", and "sections". Each card has a "DELETE" button at the bottom. To the right of the cards is a blue square icon with a white "d".

Collection Type	Action
documents	DELETE
datasources	DELETE
sections	DELETE

Software V2

Eine neue Generation von Software

Today

Syntax Driven Languages

“Tomorrow”

AI & Large Language Model
Integration

“After tomorrow”

Quantum Computing
Software

Semantic Kernel is an open-source SDK that lets you easily combine Generative AI Models syntactic programming languages like C# and Python.

Currently supported languages: C#, Python

Currently supported AI Services: OpenAI, Azure OpenAI, and Hugging Face

[Introduction](#) |  [Langchain](#)

Two types of functions

Native Functions

Semantic Functions



Wenn meine Investition von 2130,23 Dollar um 23% gestiegen ist, wie viel hätte ich, nachdem ich 5 Dollar für einen Latte ausgegeben habe?



Nach einer Erhöhung um 23% hättest du einen Betrag von etwa 2618,79 Dollar. Nachdem du 5 Dollar für einen Latte ausgegeben hast, würdest du immer noch einen Betrag von etwa 2613,79 Dollar haben.



How about something not solvable?

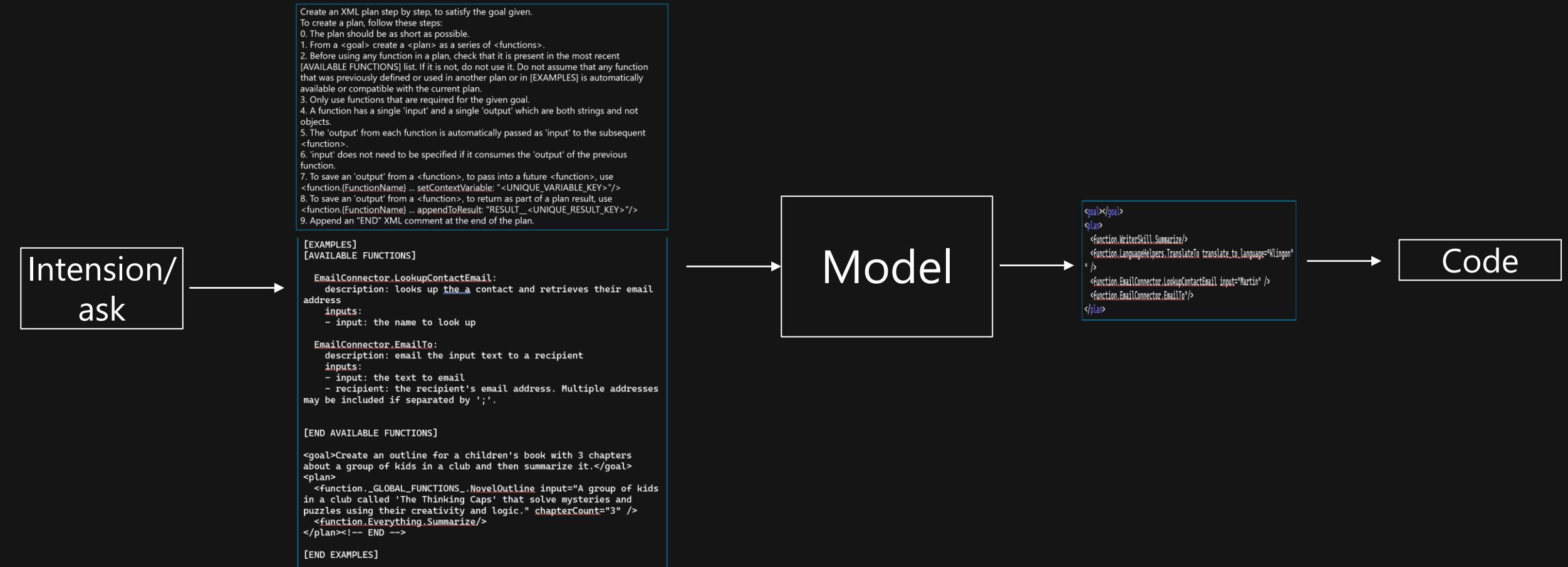
Please calculate the fiction between the stone and alpha centaury with the contraction jumping of 150 blas.



...

It seems like your request contains a mix of unrelated concepts and terms that don't have any clear meaning. To better assist you, please provide more context or clarify your question.

Mapping the Goal to Functions



SK creates the plan for a given goal

```
<goal>Summarize the input, then translate to Klingon and email it to Damir</goal>
<plan>
    <function.WriterSkill.Summarize/>
    <function.LanguageHelpers.TranslateTo translate_to_language="Klingon"
        setContextVariable="TRANSLATED_TEXT" />
    <function.EmailConnector.LookupContactEmail input="Martin"
        setContextVariable="CONTACT_RESULT" />
    <function.EmailConnector.EmailTo input="$TRANSLATED_TEXT"
        recipient="$CONTACT_RESULT"/>
</plan>
```



References

[Azure/aistudio-copilot-sample at oct-refresh \(github.com\)](#)

[Microsoft Copilot Studio](#)

[Azure AI Studio](#)

[Semantic Kernel](#)

[AI-Generated 3D Worlds](#)

[Imagine AI Art Generator](#)

[Midjourney](#)

[Azure AI SDK - Azure AI Studio](#)

[BOOK OF NEWS - Microsoft Ignite](#)

Microsoft Semantic Kernel & Azure OpenAI

THANK YOU ☺

Dr. Damir Dobric

Lead Software Architect daenet GmbH / ACP Digital
Microsoft Regional Director,
Most Valuable Professional: Azure + IoT





daenet - an  **ACP Digital** company