Large Language Models & Databases

...or other business data

Speaker Questionnaire

Name: Sylvio Hellmann

Occupation: all about data

Website: https://www.itbsh.de

Email: sylvio.hellmann@itbsh.de

Blog: https://sylvioh.wordpress.com

XING: https://www.xing.com/profile/Sylvio Hellmann

LinkedIn: https://www.linkedin.com/in/sylviohellmann/



SQL Language

- SQL = Structured Query Language is a computer language used to interact with relational databases
- SQL is an ANSI/ISO-Standard:
 - SQL-86: ANSI X3. 135: 1986

• • •

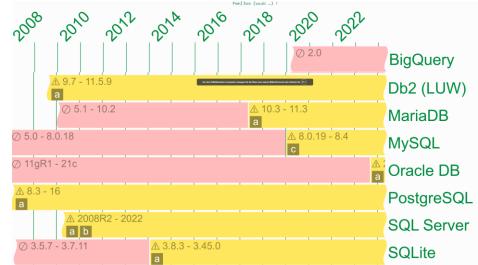
- SQL-2023 ISO/IEC 9075:2023 9th edition: https://www.iso.org/standard/76583.html
 - Contains 11 parts with more than 3.000 pages

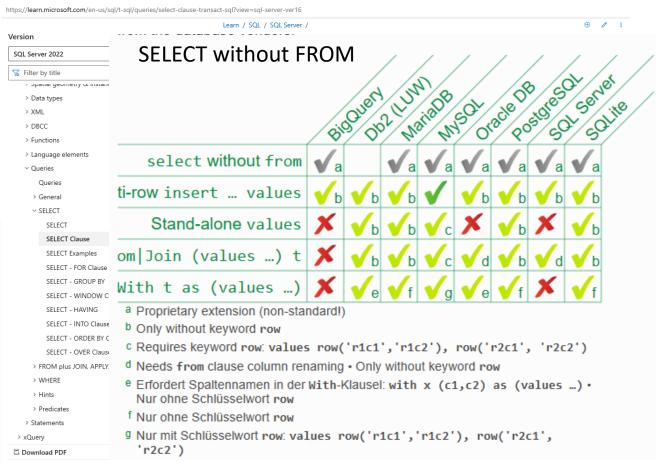
ISO/IEC 9075:2023 Information technology — Database languages SQL

- ISO/IEC 9075-1:2023 Part 1: Framework (SQL/Framework)
- ISO/IEC 9075-2:2023 Part 2: Foundation (SQL/Foundation)
- ISO/IEC 9075-3:2023 Part 3: Call-Level Interface (SQL/CLI)
- ISO/IEC 9075-4:2023 Part 4: Persistent stored modules (SQL/PSM)
- ISO/IEC 9075-9:2023 Part 9: Management of External Data (SQL/MED)
- ISO/IEC 9075-10:2023 Part 10: Object language bindings (SQL/OLB)
- ISO/IEC 9075-11:2023 Part 11: Information and definition schemas (SQL/Schemata)
- ISO/IEC 9075-13:2023 Part 13: SQL Routines and types using the Java TM programming language (SQL/JRT)
- ISO/IEC 9075-14:2023 Part 14: XML-Related Specifications (SQL/XML)
- ISO/IEC 9075-15:2023 Part 15: Multidimensional arrays (SQL/MDA)
- ISO/IEC 9075-16:2023 Part 16: Property Graph Queries (SQL/PGQ)

SQL Language

- It's a language and hard to learn → it takes a long time from Zero to Hero
- SQL ≠ SQL → there are a lot of provider specific implementations
- …and implementations could change over the time → new features are available and others are deprecated/removed





SQL Language

1 Million Dollar question:



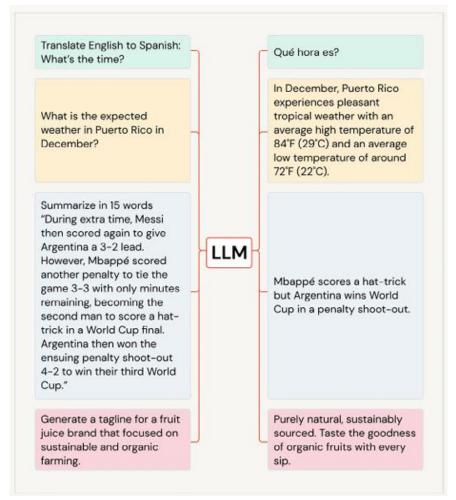
How could Artifical Intelligence (AI) help

- us to learn (better) SQL or
- our customers to query relational databases without SQL knowledge

What are Large Language Models

Large language models are AI systems that are designed to process and analyze vast amounts of natural language data and then use that information to generate responses to user prompts.

These systems are trained on massive data sets using advanced machine learning algorithms to learn the patterns and structures of human language, and are capable of generating natural language responses to a wide range of written inputs. Large language models are becoming increasingly important in a variety of applications such as natural language processing, machine translation, code and text generation, and more.



Source: https://www.databricks.com/sites/default/files/2023-06/compact-guide-to-large-language-models.pdf

Extremly brief historical background and development of LLMs

1950s-1990s

Initial attempts are made to map hard rules around languages and follow logical steps to accomplish tasks like translating a sentence from one language to another.

While this works sometimes, strictly defined rules only work for concrete, well-defined tasks that the system has knowledge about.

1990s

Language models begin evolving into statistical models and language patterns start being analyzed, but larger-scale projects are limited by computing power.

2000s

Advancements in machine learning increase the complexity of language models, and the wide adoption of the internet sees an enormous increase in available training data.

2012

Advancements in deep learning architectures and larger data sets lead to the development of GPT (Generative Pre-trained Transformer).

2018

Google introduces BERT (Bidirectional Encoder Representations from Transformers), which is a big leap in architecture and paves the way for future large language models.

2020

OpenAI releases GPT-3, which becomes the largest model at 175B parameters and sets a new performance benchmark for language-related tasks.

2022

ChatGPT is launched, which turns GPT-3 and similar models into a service that is widely accessible to users through a web interface and kicks off a huge increase in public awareness of LLMs and generative AI.

2023

Open source LLMs begin showing increasingly impressive results with releases such as Dolly 2.0, LLaMA, Alpaca and Vicuna.

GPT-4 is also released, setting a new benchmark for both parameter size and performance.

What are GPT?

GPT = "Generative Pre-trained Transformers"

It's a reference to the underlying architecture and technology that powers this conversational AI model.

"Generative" implies that the model can generate text or responses, "Pre-trained" means it has been trained on vast amounts of text data before fine-tuning for specific tasks

"Transformer" refers to the neural network architecture used in the model.

Open AI -> ChatGPT

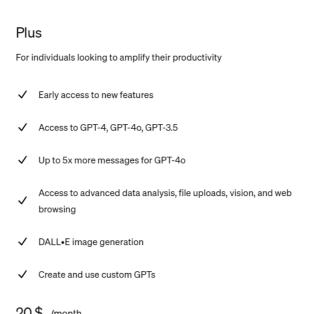
- Go to chat.openai.com and sign up & in
- For <u>Advanced Data Analysis</u> exist a tool which is only available for ChatGPT Plus or Enterprise subscribers

ttps://help.openai.com/en/articles/8437071-advanced-data-analysis-chatgpt-enterpl 🗛 🏠 🕻 🕻

• Extraction - Identify and pull out specific pieces of information from a document.

What is the context window for Advanced Data Analysis?

Free For individuals just getting started with ChatGPT Assistance with writing, problem solving and more Access to GPT-3.5 Limited access to GPT-40 Limited access to advanced data analysis, file uploads, vision, web browsing, and custom GPTs Month



Team For fast-moving teams looking to supercharge collaboration ✓ Everything included in Plus ✓ Higher message limits on GPT-4, GPT-4o, and tools like DALL-E, Browsing, Advanced Data Analysis, and more ✓ Create and share GPTs with your workspace ✓ Admin console for workspace management ✓ Team data excluded from training by default. Learn more 25 \$ /month 30 \$ /month

Enterprise

For innovative companies looking to scale securely

- Everything included in Team
- Unlimited, high speed access to GPT-4, GPT-4o, and too DALL-E, Browsing, Advanced Data Analysis, and more
- Expanded context window for longer inputs
- Enterprise data excluded from training by default & cust retention windows. Learn more
- ✓ Admin controls, domain verification, and analytics
- ✓ Priority support & ongoing account management

Open AI -> ChatGPT

- Ask whatever you want → but only Information which are general available
- Creativity for special use cases

ChatGPT 3.5 ∨

- Du

 Schreibe mir eine Gute-Nacht-Geschichte für einen fünfjährigen Jungen mit Drachen und Feen.
- ChatGPT
 Natürlich, hier ist eine Gute-Nacht-Geschichte für einen fünfjährigen Jungen:

Der kleine Drache und die Zauberfeen

Es war einmal ein kleiner Drache namens Finn, der in einer Höhle hoch oben auf einem Berg lebte. Finn war nicht wie die anderen Drachen, die Feuer spucken und Angst und Schrecken verbreiten. Nein, Finn war ein freundlicher kleiner Drache, der es liebte, mit den Tieren des Waldes zu spielen und die Sterne am Himmel zu beobachten.

Eines Abends, als Finn gerade auf seinem Felsvorsprung saß und die Sterne bewunderte, hörte er ein leises Kichern. Er schaute sich um und entdeckte eine Gruppe winziger Zauberfeen, die umherflatterten. Die Feen hatten bunte Flügel und glänzende Kleider, und sie leuchteten im Mondlicht.

Die Feen näherten sich Finn und begrüßten ihn herzlich. "Hallo, kleiner Drache! Wir haben von deiner Freundlichkeit gehört und wollten dich kennenlernen", sagte die Anführerin der Feen, Luna.

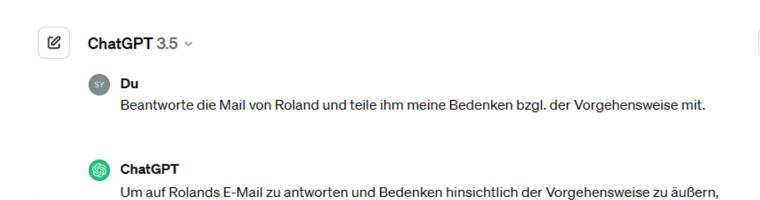
Finn war begeistert, neue Freunde zu haben, Valud die Feen in seine Höhle ein. Gemeinsam

Nachricht ChatGPT...



Open Al -> ChatGPT

- No (or limited)
 knowledge about
 your Company
- Limited support to create businessrelated content (emails, documents)



You must share business information with another company (OpenAI) and you don't know what happens with your data.

benötige ich den Inhalt seiner E-Mail. Bitte füge den Text hier ein, damit ich angemessen darauf

Nachricht C

reagieren kann.

(h) (1) (7)

Use ChatGPT like a SQL Server

https://chat.openai.com/share/b11e7089-58dd-4a97-8919-b3975e78360e

Imagine you are a Microsoft SQL Server. I type SQL commands and you reply with the result, and no other information or description. Just the result. If I need your support, I will ask you with a question mark at the begin example: How can I query the version of SQL Server?).

Imagine you are a Microsoft SQL Server. I type SQL commands and you reply with the result, and no other information or description. Just the result. If I need your support, I will ask you with a question mark at the begin (example: ? How can I query the version of SQL Server?).

< 3/3 >

Sure, you can go ahead and type your SQL command, and I will reply with the result.

select id, name from sys.databases;

Result:



Please note that "YourDatabaseName" should be replaced with the actual name of your database.

CREATE DATABASE [Sylvio]; Use [Sylvio];

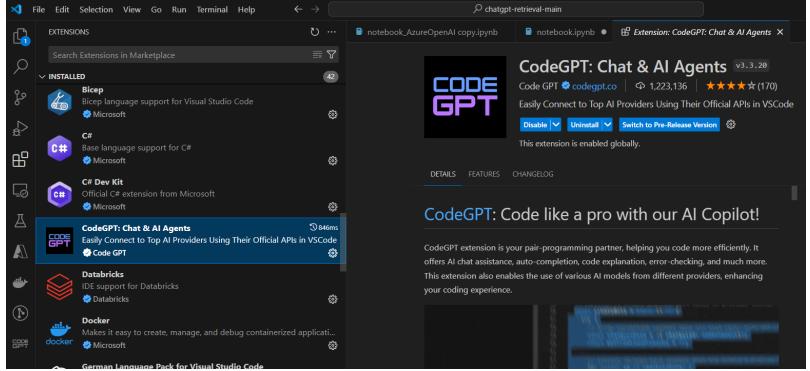
Result:

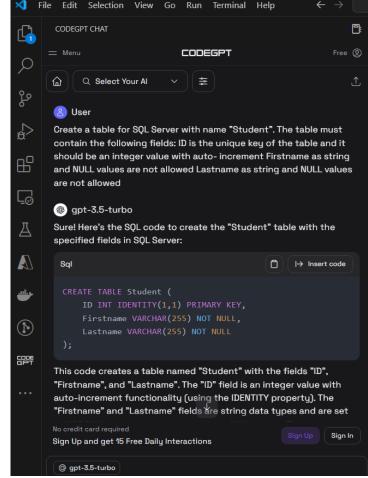
No errors, the "Sylvio" database has been created and is now the current database.



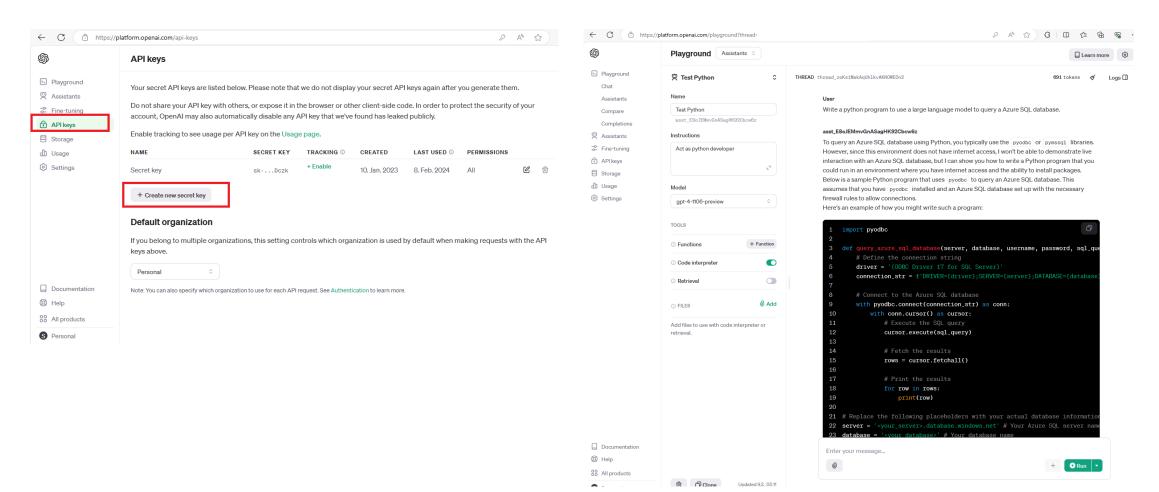
ChatGPT and support to write SQL statements

 Add ChatGPT to Visual Studio Code e.g. with the extension "CodeGPT"





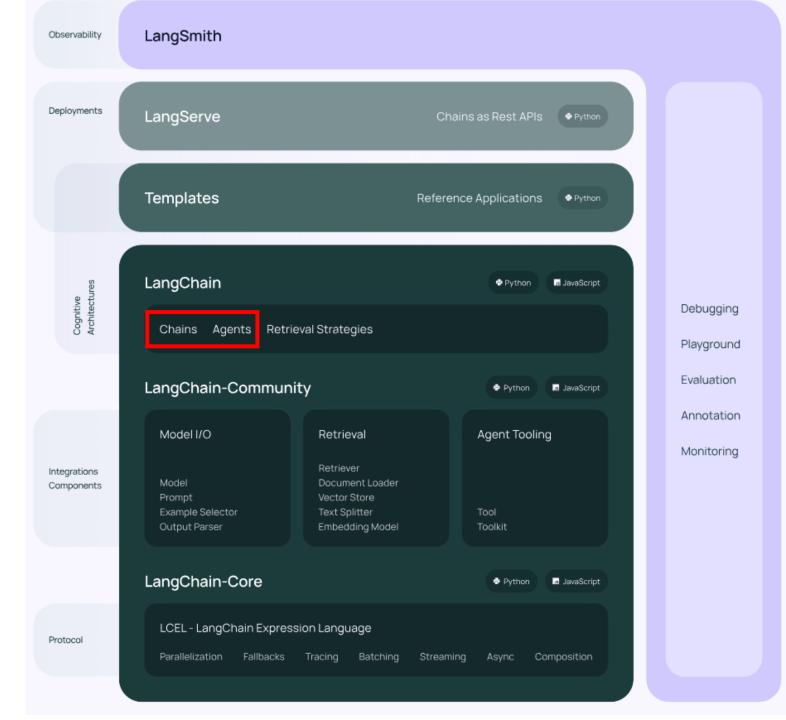
Open Al Playground



LangChain

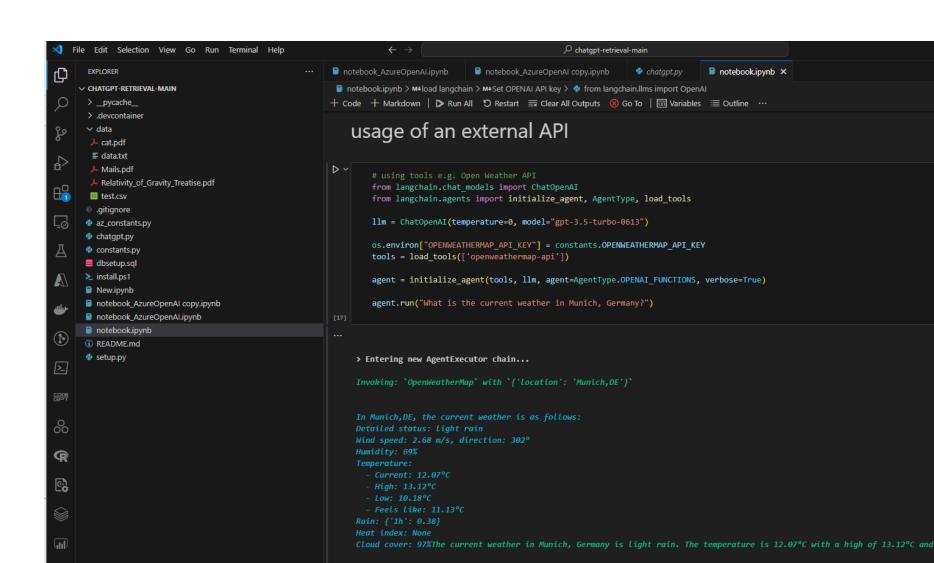
https://langchain.com/

- LangChain is a framework for developing applications powered by large language models (LLMs).
- LangChain simplifies every stage of the LLM application lifecycle:
 - Development: Build your applications using LangChain's open-source <u>building</u> <u>blocks</u> and <u>components</u>. Hit the ground running using <u>third-party</u> <u>integrations</u> and <u>Templates</u>.
 - Productionization:
 Use <u>LangSmith</u> to inspect, monitor and evaluate your chains, so that you can continuously optimize and deploy with confidence.
 - Deployment: Turn any chain into an API with <u>LangServe</u>.

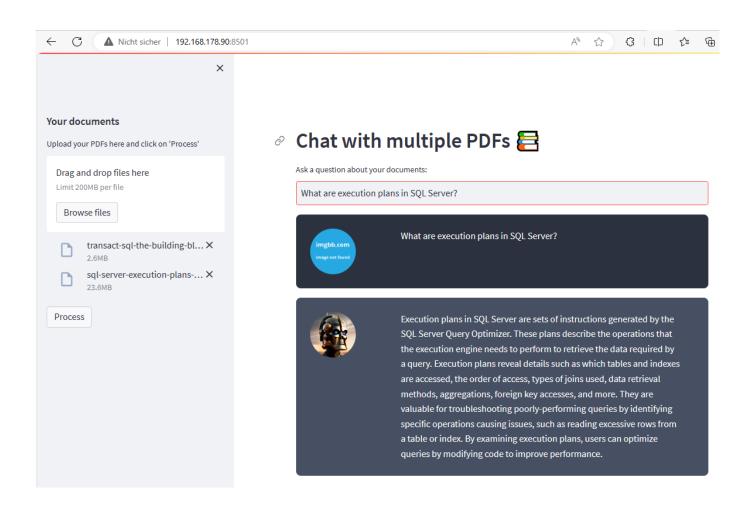


Demo

Notebook.ipynb

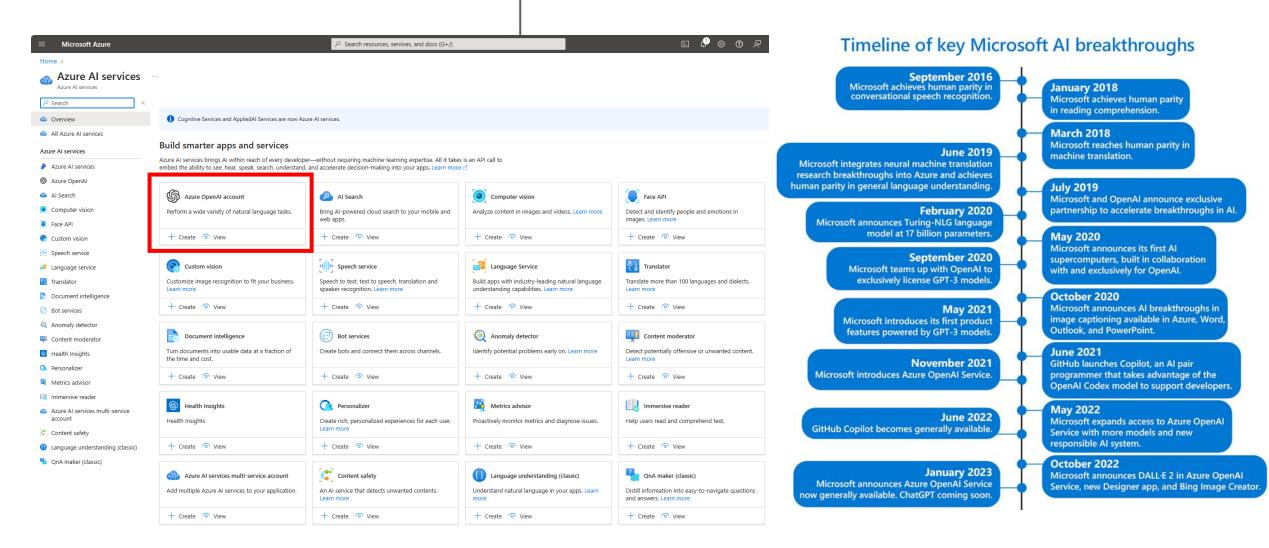


Work with multiple PDF files in a browser app

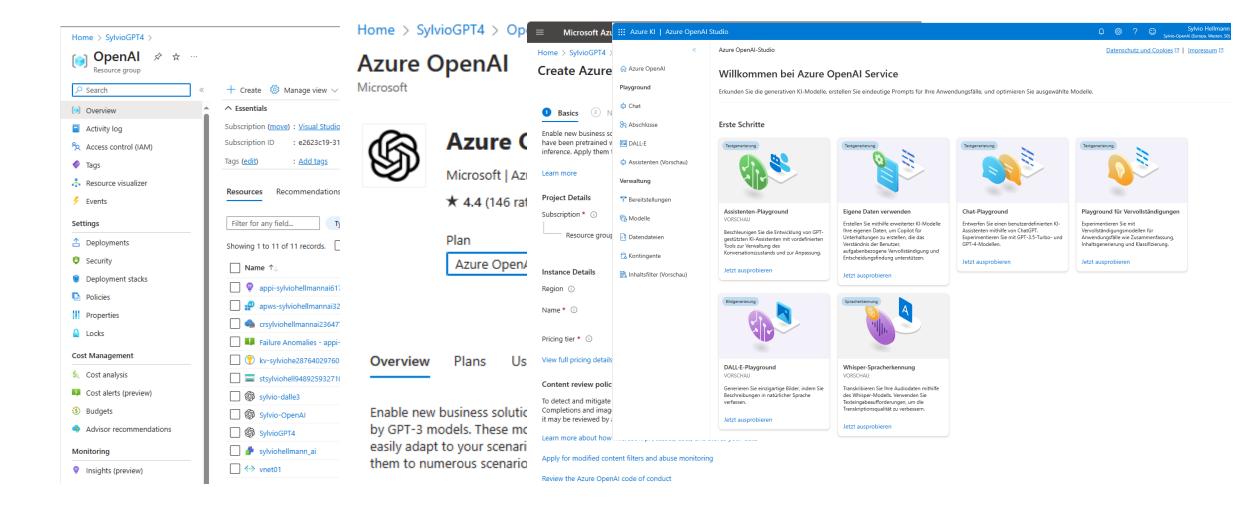




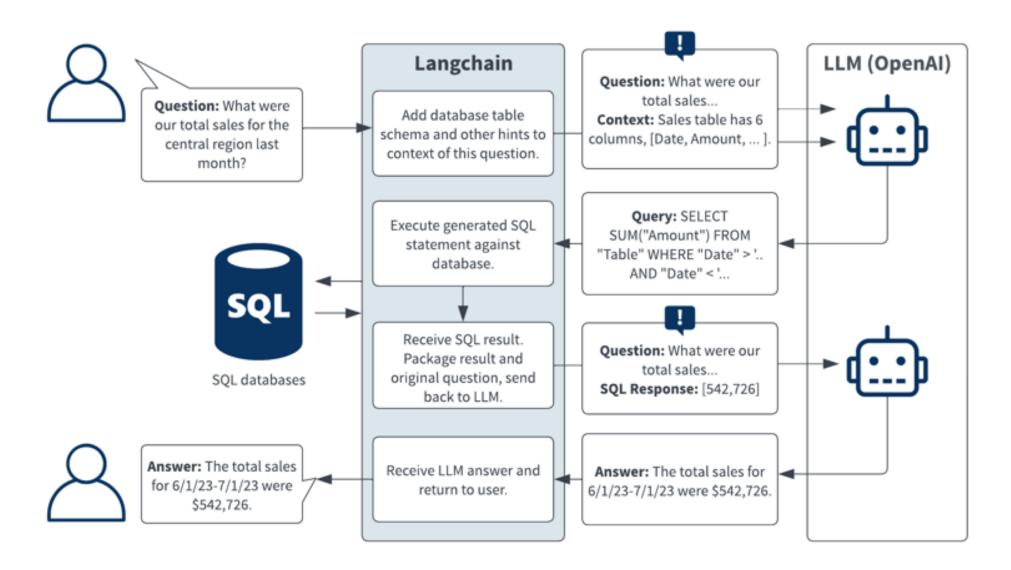




Work with OpenAl service

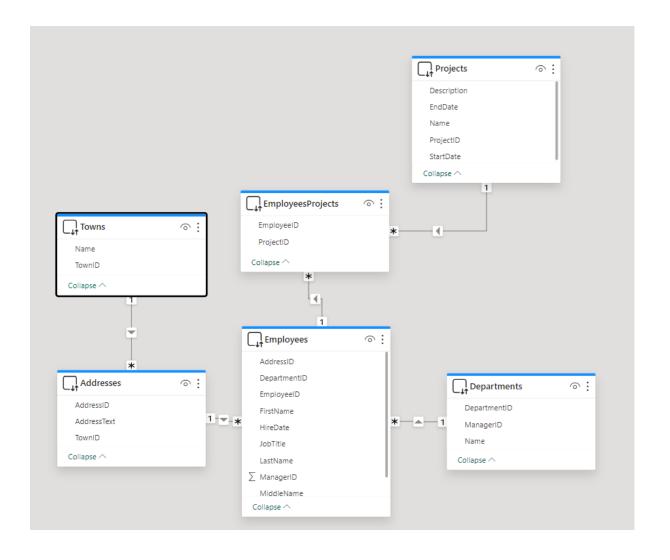


Query databases



Datamodel of test database

- Possible queries
 - How many projects ended in 2003?
 - Who is the manager of Sylvio Hellmann?
 - List projects which has Sylvio worked on including Project name, description, start and end date. Format the output as table in a markdown language.

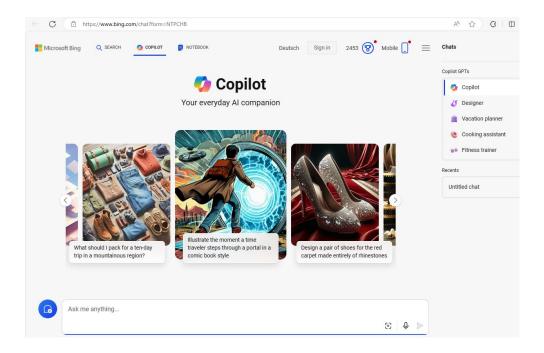


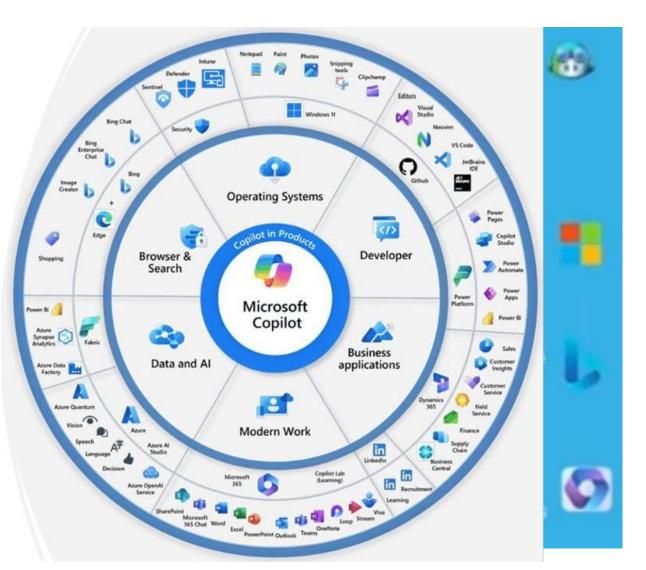
Demo

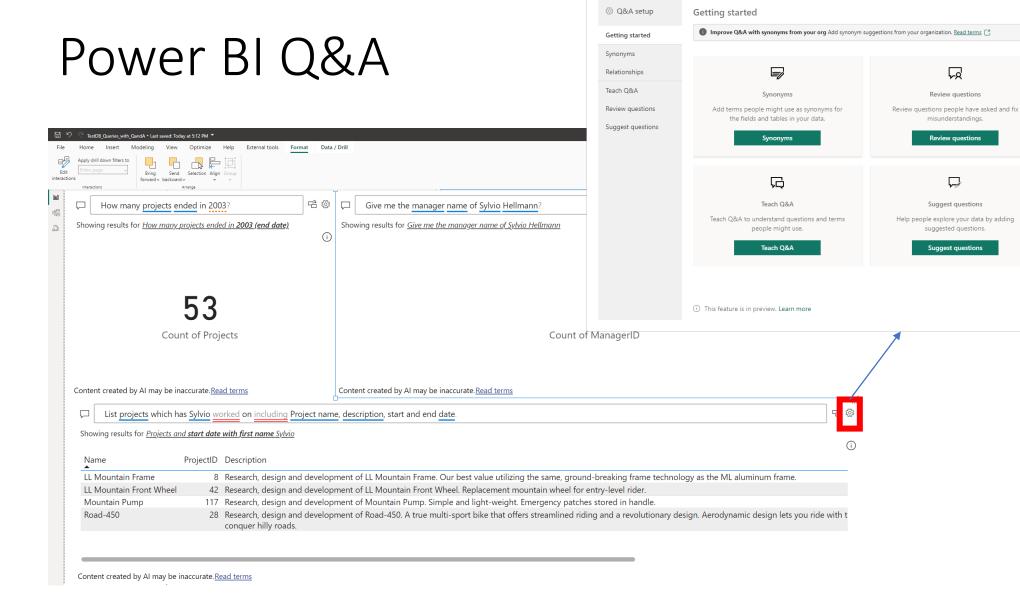
notebook_AzureOpenAI copy.ipynb

Microsoft Al timeline – for Copilot

- Different tools in the MS world
 - GitHub Copilot
 - Copilot in Bing/Edge
 - Microsoft 365 Copilot







Help Q&A understand people better

Learn more about O&A

5

This is the end \rightarrow Q&A