



# EUROPEAN MICROSOFT FABRIC

## Community Conference

STOCKHOLM 24-27 SEPTEMBER 2024

JOIN THE CONVERSATION

#FABCONEUROPE



# Building a custom Q&A bot on your Fabric data with AI Studio



**Marc-Marijn Ossel**

Technology Specialist BI  
Microsoft The Netherlands



**Onno van der Horst**

Partner Solution Architect – ISV  
Microsoft The Netherlands

# Today, you will learn about

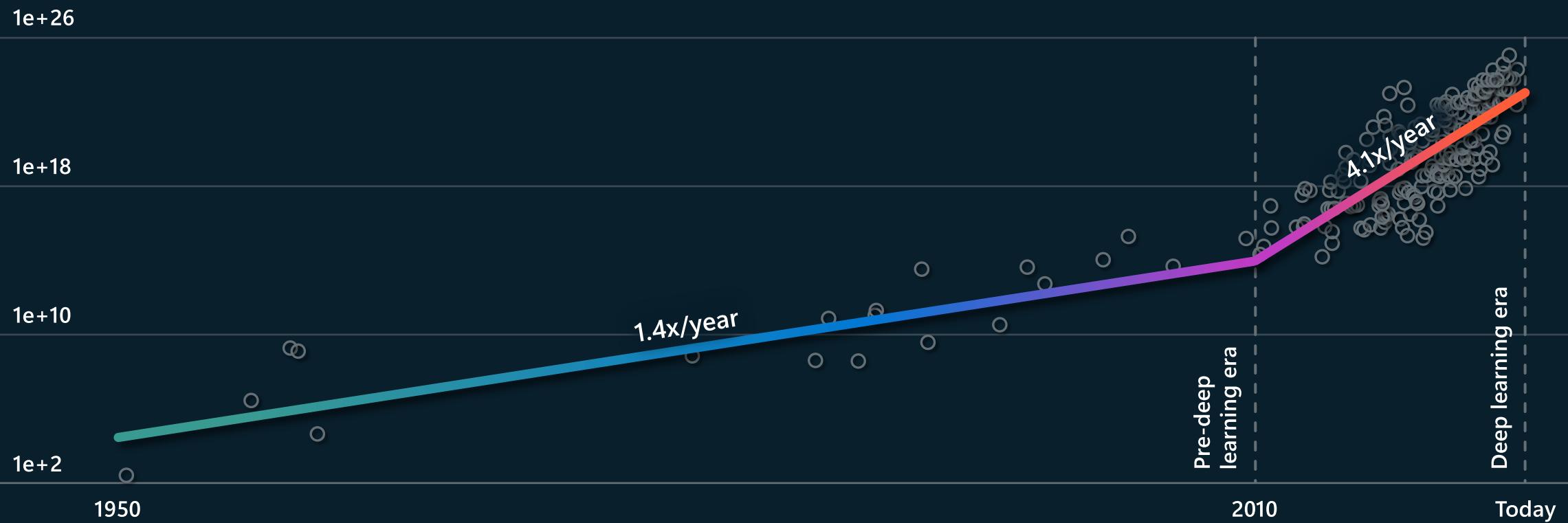
- 01 Current state of GenAI in analytics
- 02 Understanding LLMs and overcoming their limitations
- 03 Case: single chatbot on structured and unstructured data
- 04 Live demo
- 05 Future outlook and key takeaways
- 06 Q&A (of course)

# The history of Q&A on data (2013)



# Scaling laws

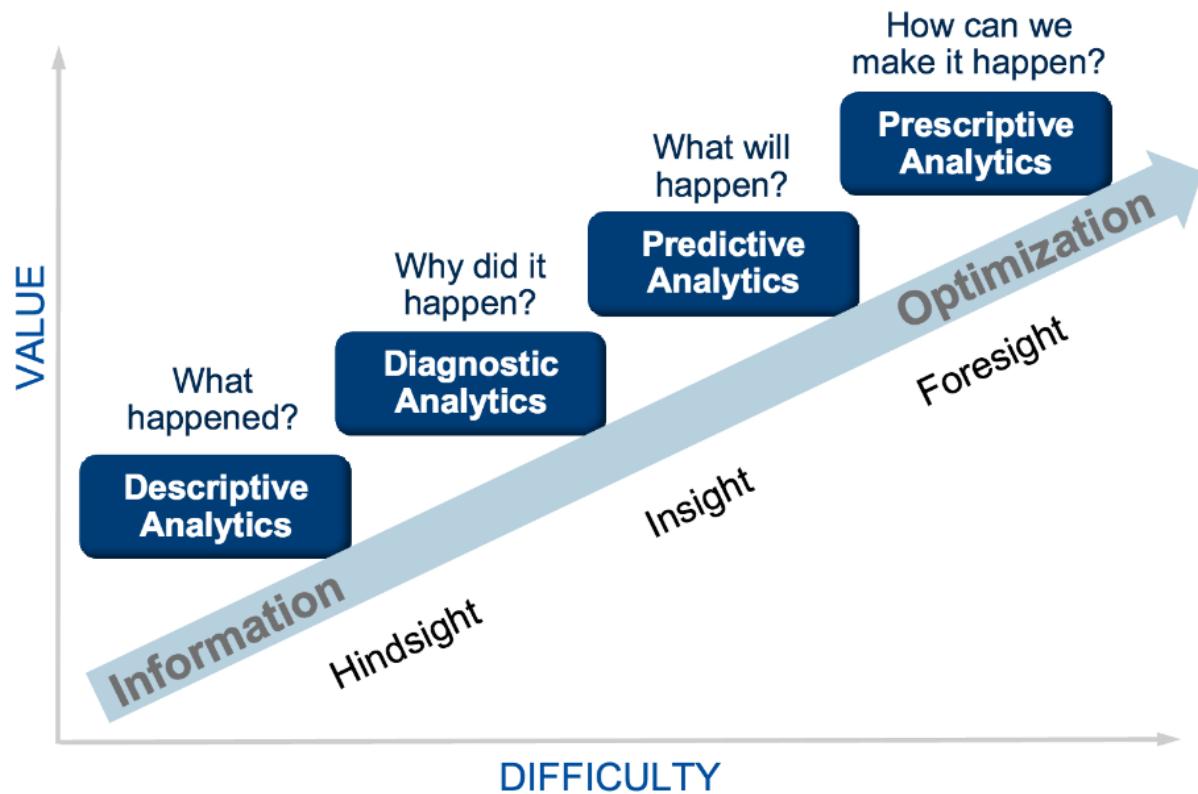
Training compute of notable machine learning systems over time (FLOP)



Source: Epoch, 'Parameter, Compute and Data Trends in Machine Learning'

# Are LLMs useful for analytics?

## Gartner Analytic Ascendancy Model

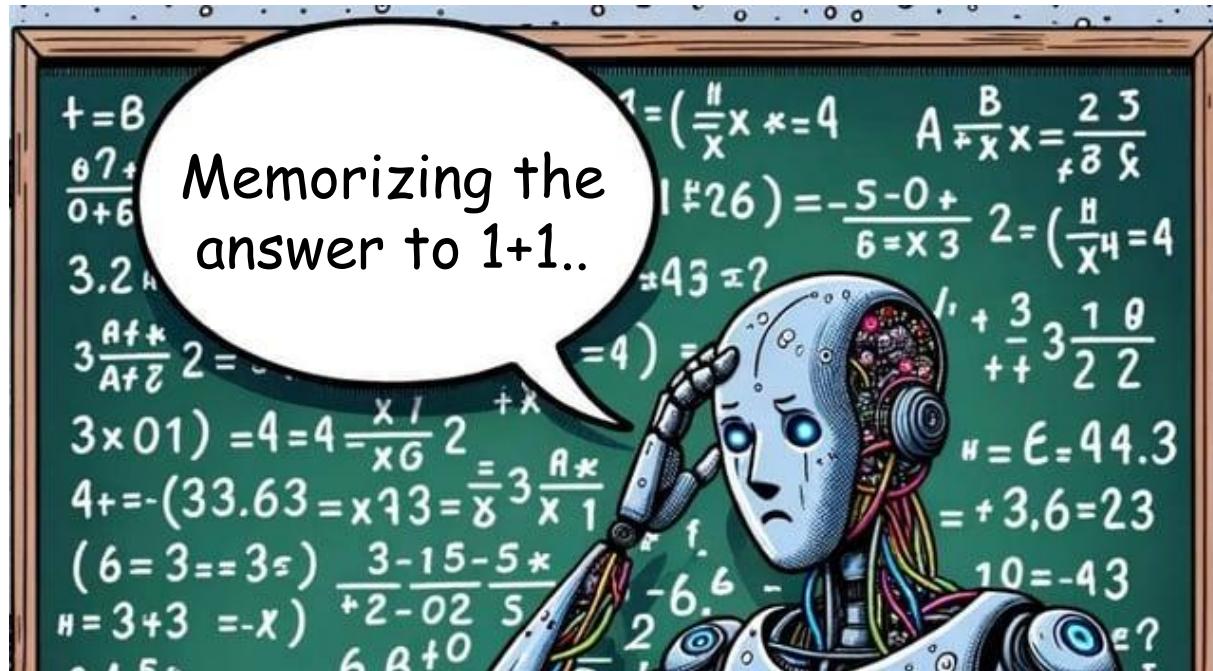


## When Generative AI Is and Is Not Effective

Use-case family	Generative models' current usefulness	Example use cases
Prediction/forecasting	Low	Risk prediction, customer churn prediction, sales/demand forecasting
Decision intelligence	Low	Decision support, augmentation, automation
Segmentation/classification	Medium	Clustering, customer segmentation, object classification
Recommendation systems	Medium	Recommendation engine, personalized advice, next best action
Content generation	High	Text generation, image and video generation, synthetic data
Conversational user interfaces	High	Virtual assistant, chatbot, digital worker

Sources: Gartner (2012) & (2024)

# Understanding LLMs and their limitations in simple math



Source: Generated by Daniel Litt, using Dall-E 3

- Treats numbers as **words**
- Pattern recognition, not **calculation**
- **Finite** set of training data

ChatGPT Auto ▾



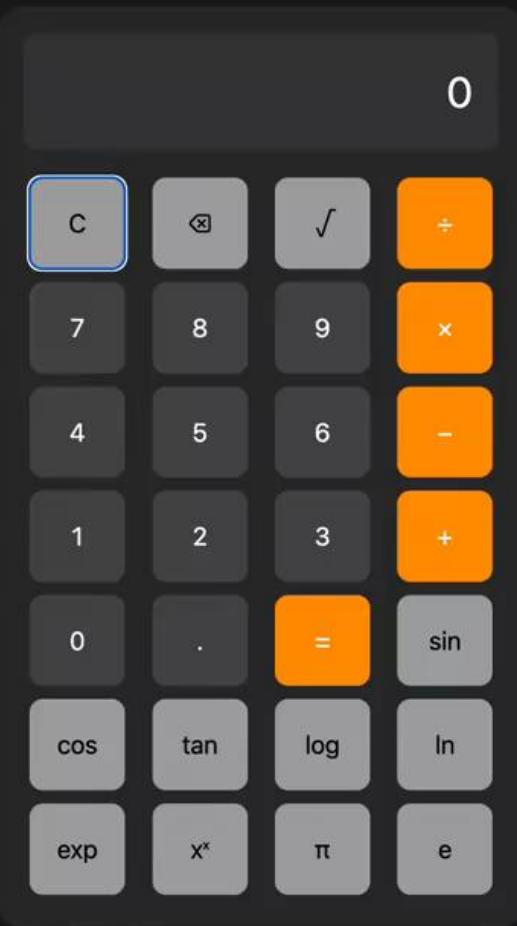
M

9.11 or 9.9 which is bigger



9.11 is bigger than 9.9. This is because 9.11 is equivalent to 9.110, and when compared to 9.900, 9.900 is larger in the hundredths and thousandths places.

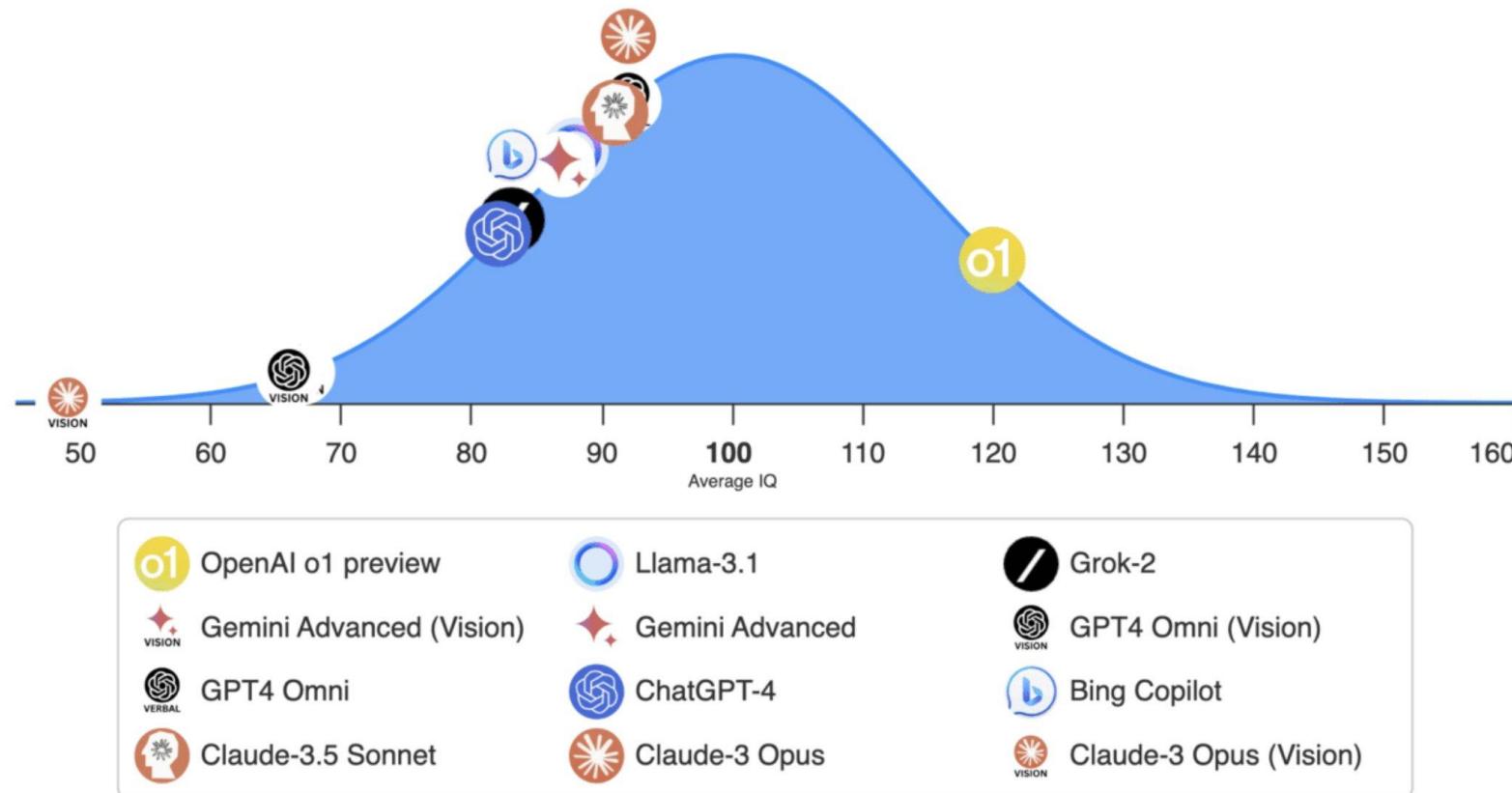




Source: [GitHub - Workflows with OpenAI o1-mini model](#)

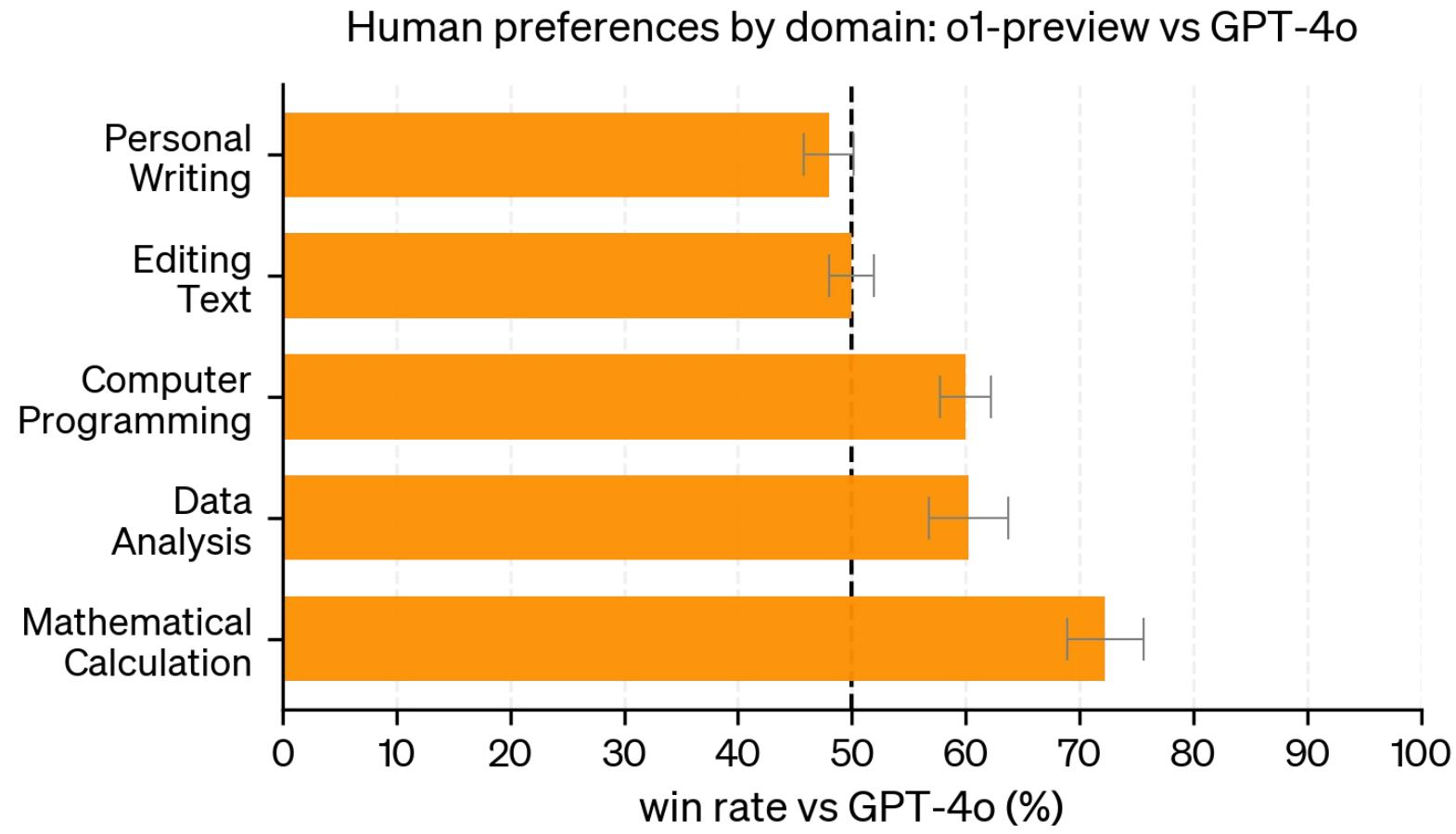


# The era of AI reasoning with OpenAI o1-preview



Source: <https://trackingai.org/IQ>

# Where does OpenAI o1-preview perform better?



Source: OpenAI

# Foundational LLM capabilities are improving, but..



# LLMs on structured data for descriptive analytics



# LLMs on structured data for descriptive analytics



# AI Skills for structured data



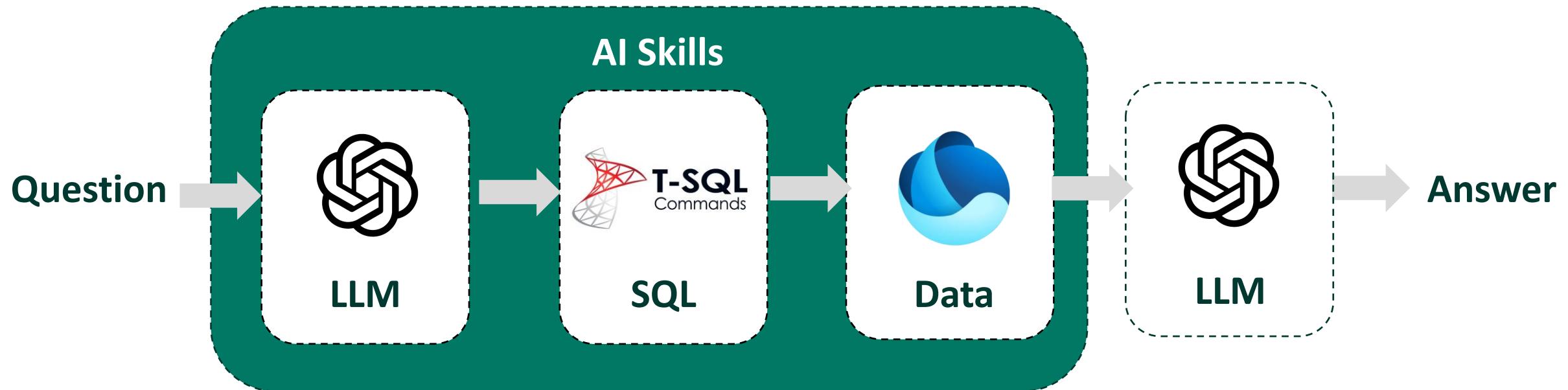
AI Skills in  
Microsoft Fabric  
Now in public preview



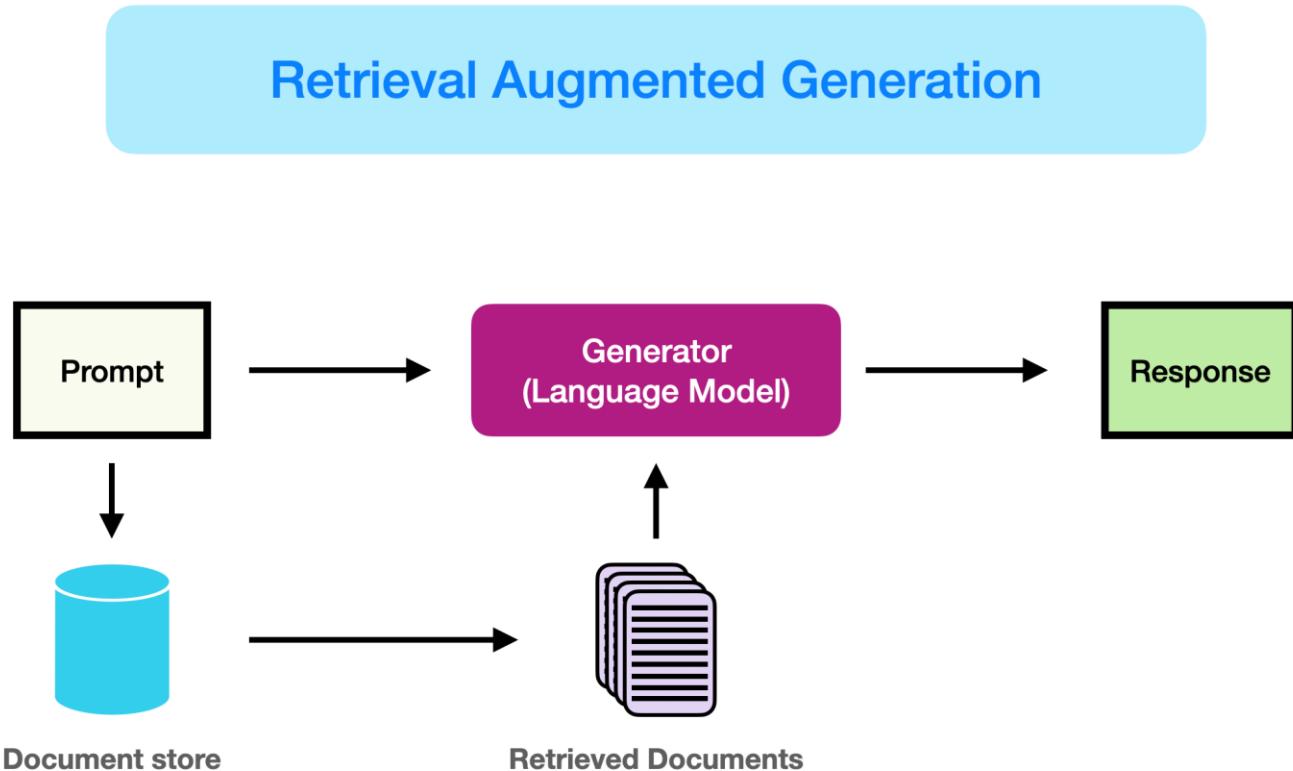
# LLMs on structured data for descriptive analytics



# LLMs on structured data for descriptive analytics

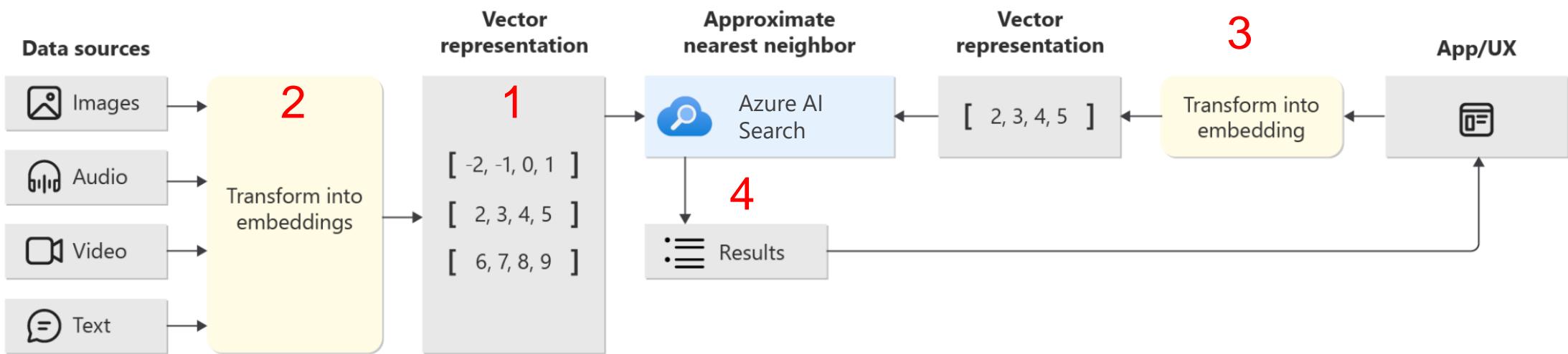


# Unstructured data: Retrieval Augmented Generation



Source: Omar Sar

# Azure AI Search: vector search



Source: Microsoft Docs

# Bringing it all together

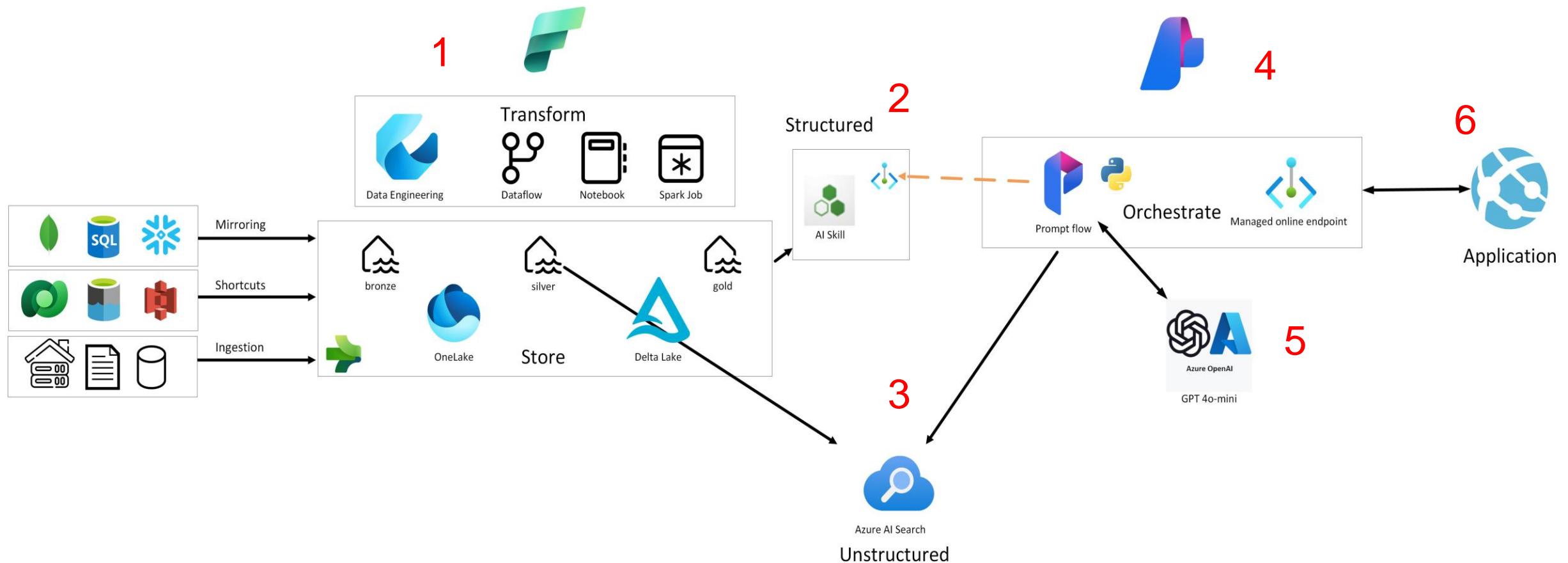


Azure AI Studio



Prompt Flow

# Solution architecture of this demo



# Use case of this demo



Sales employees



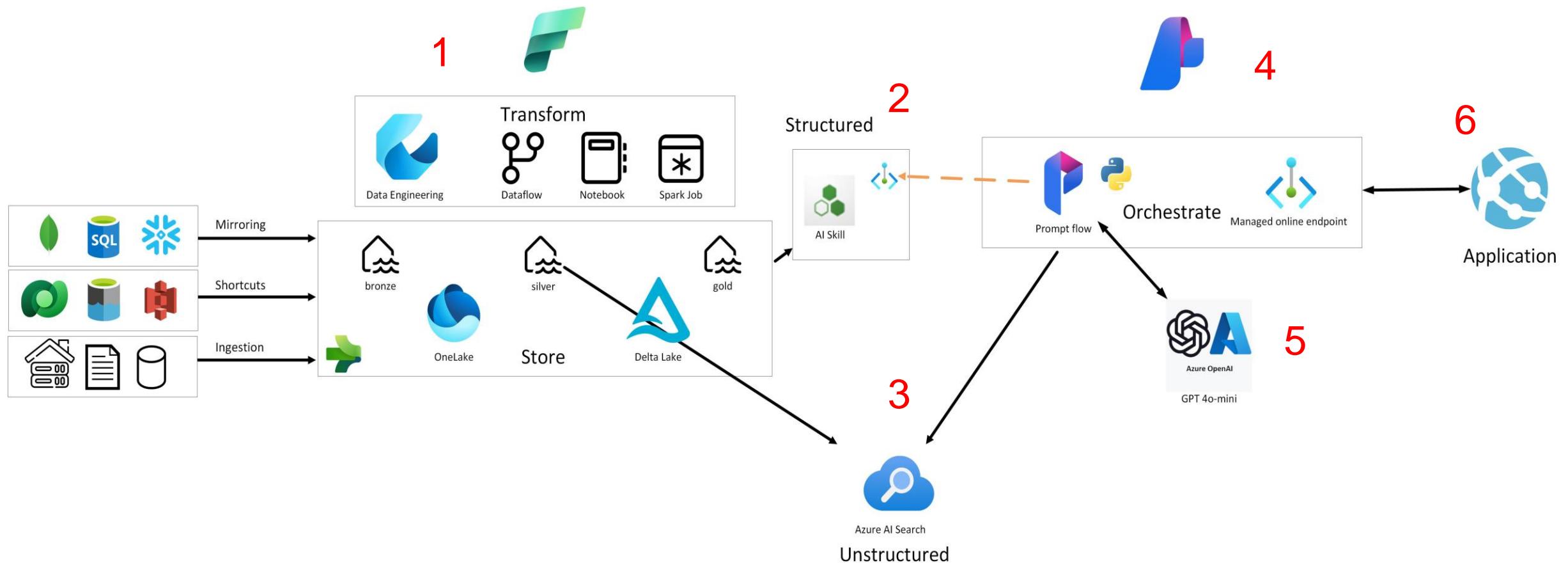
Business application

*“Improve customer satisfaction”*

# Live Demo



# Solution architecture of this demo



# Getting started



[aka.ms/fabcon-azureai-demo](https://aka.ms/fabcon-azureai-demo)



AI Skills: F64 or higher

# Considerations and limitations

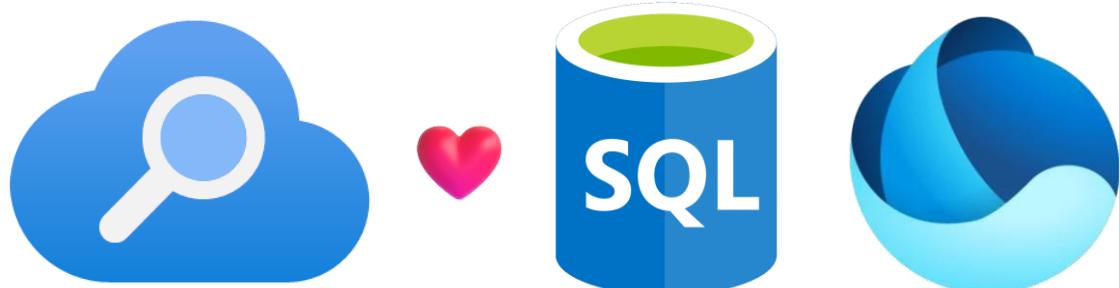


Invoke AI Skills endpoint from Azure AI Studio is not (yet) available

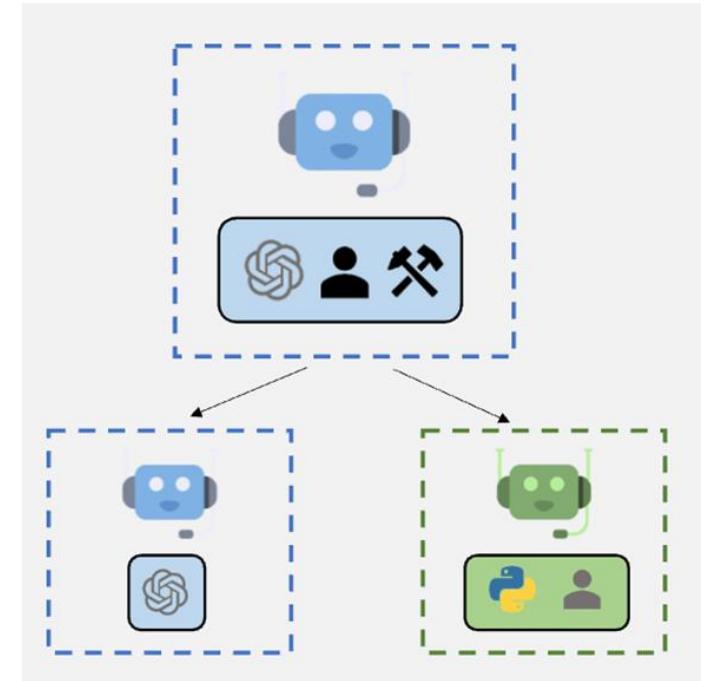


For unstructured data: access is managed in Azure AI Search

# Future outlook



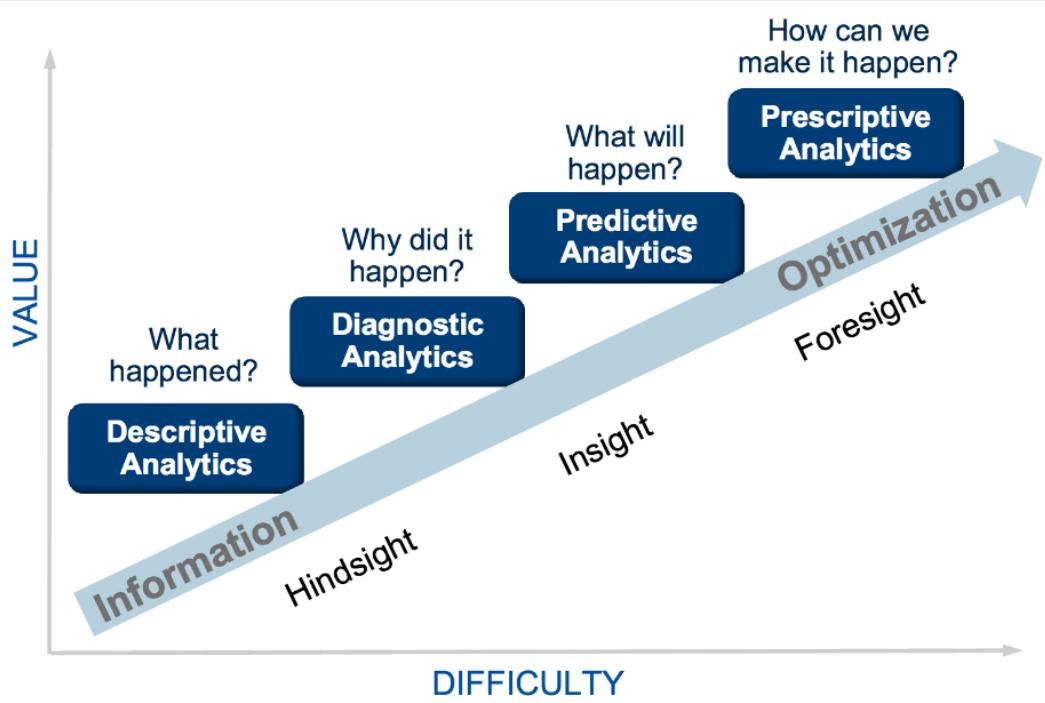
Vector indexing for structured data



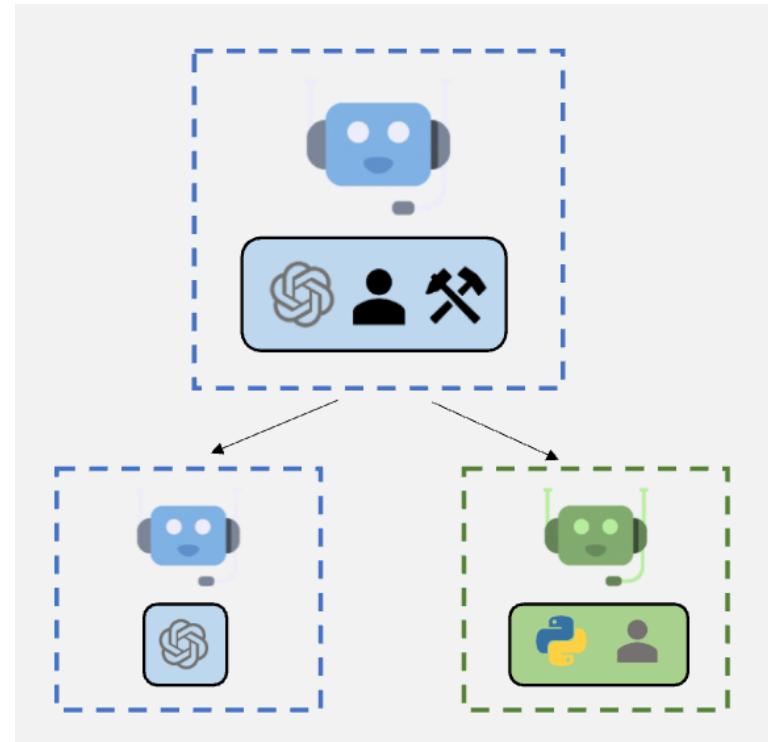
Autonomous agents

# Future outlook

## Gartner Analytic Ascendancy Model



Source: Gartner (2012)



Source: Microsoft AutoGen Whitepaper (2024)

# Key takeaways

- 01 Foundational models are getting better at data analysis through reasoning
- 02 LLMs still depend on invoking external tools for mathematical calculations
- 03 AI Skills in its current state (NL2SQL) is great for descriptive analytics
- 04 Manage the full development cycle of GenAI applications in Azure AI Studio
- 05 Prompt flow templates get you up to speed for RAG
- 06 Recent innovations with agent capabilities are promising!

# Connect with us



## Marc-Marijn Ossel

Technical Specialist @ Microsoft  
The Netherlands

LinkedIn: <https://aka.ms/marc-marijn>



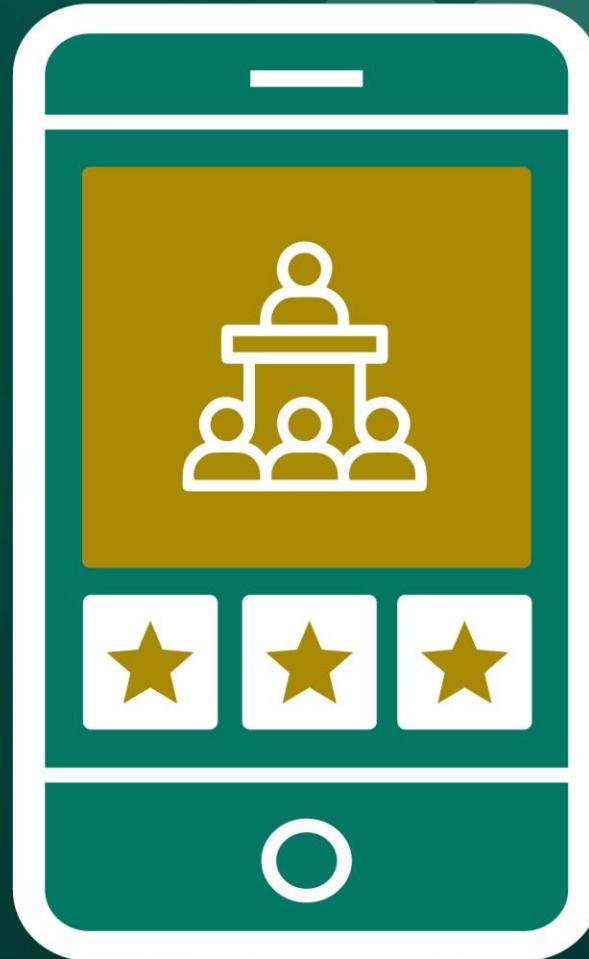
## Onno van der Horst

Partner Solution Architect – ISV @  
Microsoft The Netherlands

LinkedIn: <https://aka.ms/onno>



Please rate  
this session  
on the app



cvent

GET IT ON  
Google Play

Download on the  
App Store