



# Azure OpenAI – Revolutionieren Sie Ihr Unternehmen mit KI



**Sophia Cullen**  
Cloud Solution Architect



**Sebastian Henze**  
Architect Strategy & Innovation



“Avocados dancing, drinking, singing and partying at a Hawaiian luau.”

# Agenda

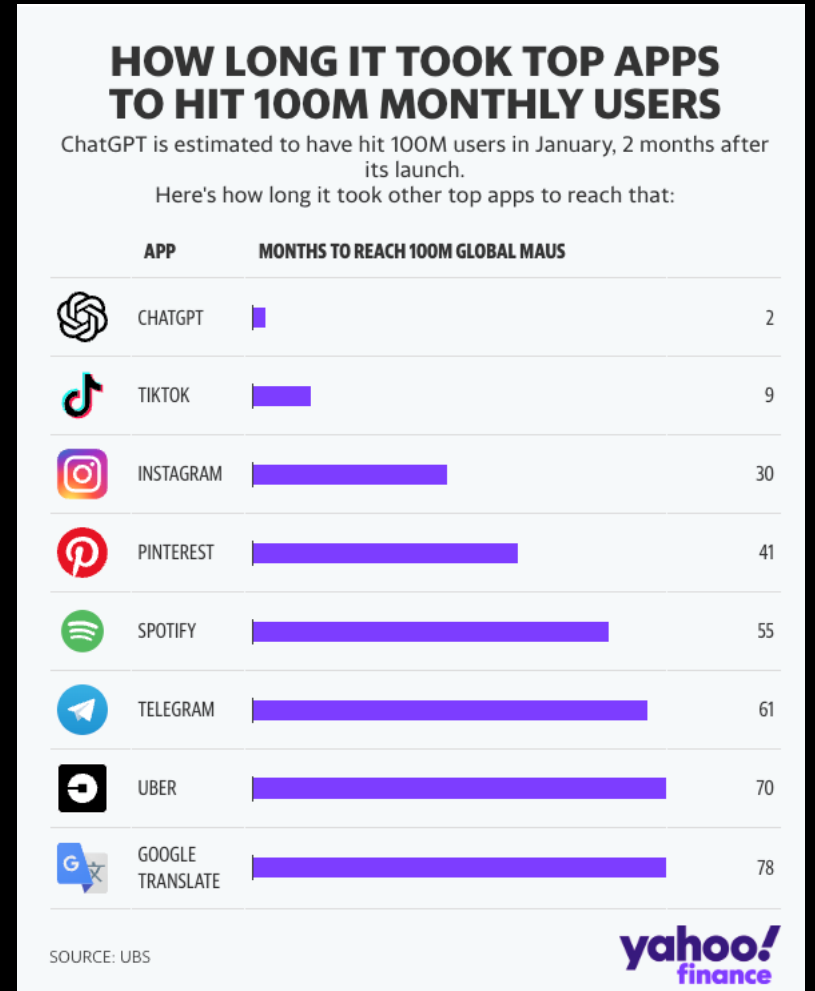
- 💡 Einführung in Large Language Models und Azure OpenAI
- 👤 Technischer Überblick über GPT-3.5, ChatGPT, Codex & DALL-E
- 👛 Wie man diese Modelle auf Azure nutzen kann
- 🔍 Projekterfahrungen und Anwendungsfälle
- 💬 Fragen und Diskussion

**Einführung**

# This is something special



+



Available since November 30, 2022

## Artificial Intelligence

### Machine Learning

### Deep Learning

### Generative AI



## Artificial Intelligence

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

---



## Machine Learning

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

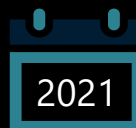
---



## Deep Learning

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

---

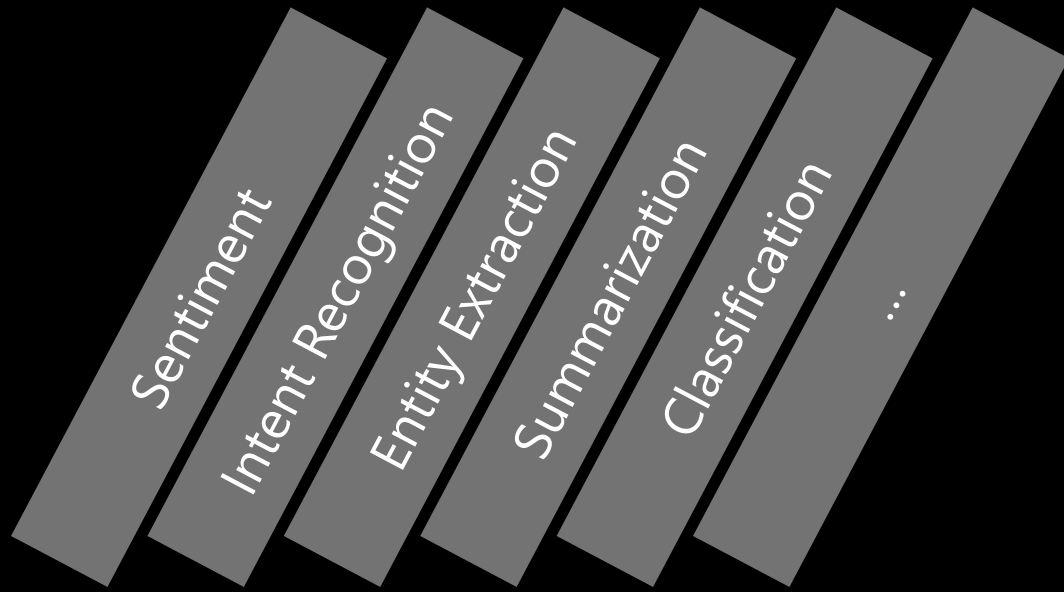


## Generative AI

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

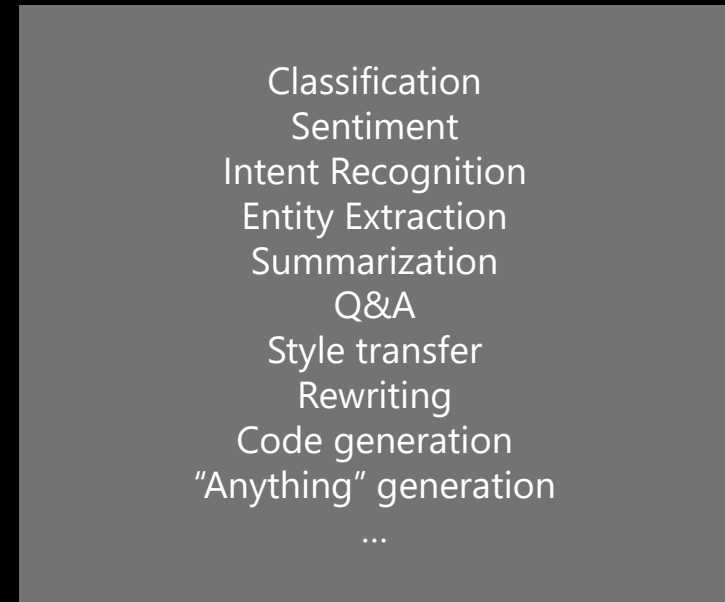
# Large Language Models

## Traditional ML



One model per capability  
Lots of data to train  
Highly optimized for use case

## OpenAI GPT-3



Single large model for all use cases  
Describe in «human language» what you  
want it to do



*Ensure that artificial  
general intelligence (AGI)  
benefits humanity*



*Empower every person and  
organization on the planet  
to achieve more*

---

**GPT-3**

Generate and understand text

**Codex**

Generate and understand code

**DALL·E**

Generate images from text prompts

**ChatGPT**

Generate and understand conversations

# Azure OpenAI Service

GPT-3

Codex

DALL·E *(preview)*

ChatGPT *(preview)*



Deployed within your Azure subscription, secured by you, accessed only by you, and tied to your datasets and applications



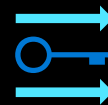
Large, pretrained AI models to unlock new scenarios



Custom AI models fine-tuned with your data and hyperparameters



Built-in responsible AI to detect and mitigate harmful use



Enterprise-grade security with role-based access control (RBAC) and private networks



## Generative AI

### GPT-3

Prompt:

Summarize this personal finance article in two sentences for my tax-filing customers:

<https://www.fidelity.com/learning-center/personal-finance/tax-pitfalls>

Response:

Tax pitfalls can be costly, so it's important to be aware of them. Some common tax pitfalls include failing to file or pay taxes on time, not reporting all income, claiming too many deductions, and not keeping accurate records.

### Codex

Prompt:

```
Table customers, columns =  
[CustomerId, FirstName,  
LastName, Company,  
Address, City, State,  
Country, PostalCode]
```

Create a SQL query for all customers in Texas named Jane

query =

Response:

```
SELECT *  
FROM customers  
WHERE State = 'TX' AND  
FirstName = 'Jane'
```

### DALL·E

Prompt:

A ball of fire with vibrant colors to show the speed of innovation at our media and entertainment company

Response:



### ChatGPT

Prompt:

I'm having trouble getting my Xbox to turn on.

Response:

There are a few things you can try to troubleshoot this issue, First, ... Next, check the power cord ...

Prompt:

Thanks! That worked. What games do you recommend for my 14-year-old?

Response:

Here are a few games that you might consider:  
(bulleted list) ...

# Azure AI

## Applications



Partner Solutions

## Application Platform

AI Builder



Power BI



Power Apps



Power Automate



Power Virtual Agents

## Scenario-Based Services

Applied AI Services



Bot Service



Cognitive Search



Form Recognizer



Video Indexer



Metrics Advisor



Immersive Reader

## Customizable AI Models

Cognitive Services



Vision



Speech



Language



Decision

Azure OpenAI Service

## ML Platform



Azure Machine Learning



Business Users



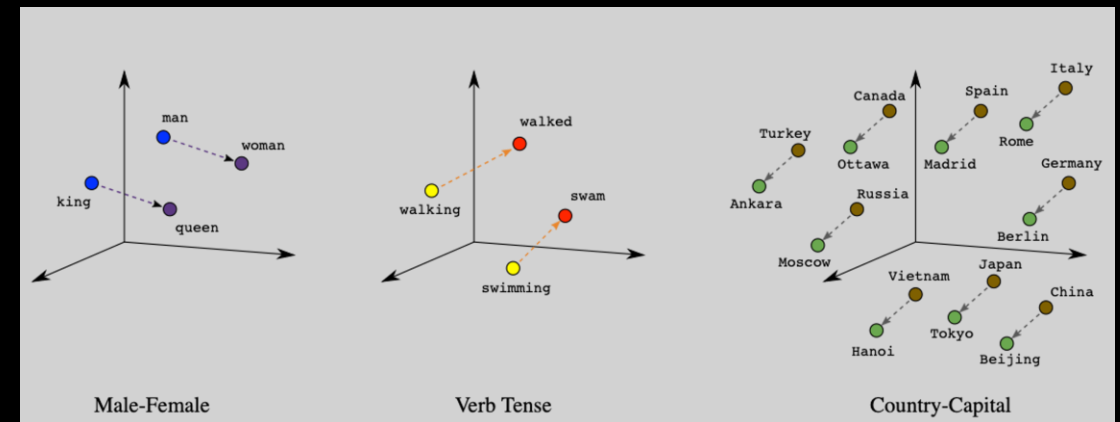
Developers & Data Scientists

# Technischer Überblick

# What is GPT-3?

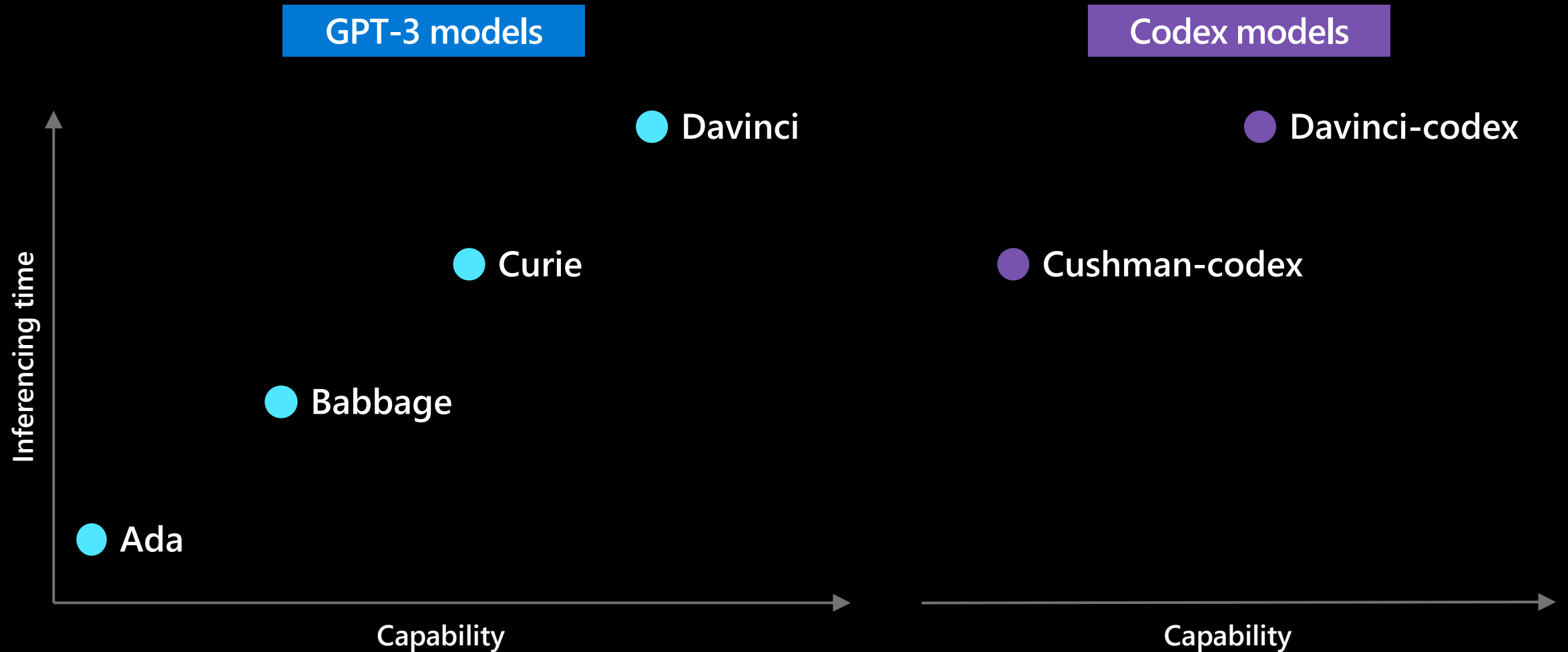
## Generative Pre-Trained Transformer

- Large language prediction model that uses deep learning neural nets to predict the next token.
- Generative in nature i.e., it will continue to produce text from the generated token.
- Pre-Trained with 175B+ parameters with data from web scraping, Books, Wikipedia etc., extensible via embeddings
- First use case is NLP – generating human understandable text, in multiple languages



Embedding Examples

# Azure OpenAI | Family of GPT-3 Models



# Azure OpenAI | Prompt Design

Extract the mailing address from this email:

Hi John Doe,

It was great to meet up at Build earlier this week. I thought the AI platform talk was great and I really enjoyed it.

I appreciate the offer for the book. If you are OK, you can mail it to me at home, or 123 Microsoft Way, Bellevue WA 92004.

Regards,

Chris Hoder

**Prompt**—Text input that provides some context to the engine on what is expecting.

**Completion**—Output that GPT-3 generates based on the prompt.

# Azure OpenAI | GPT-3 Ideate, Experiment and Fine-Tune

Iterate on ideas with a general-purpose text-in/text-out interface

## Prompt

Summarize game commentary into highlights:

Shey Peddy is applying ball pressure at the top against Sabrina Ionescu. At 7:48 remaining in the quarter; Peddy

What are the main highlights of the game so far?

## Sample response

The game has been close with Phoenix leading New York 7-5. Shey Peddy has been key for Phoenix.

Refine with examples ('few shot learning') with a simple UX

## Prompt

Turn game commentary into highlights:

Commentary: What a pickup she has  
Main highlights: New York has domina  
###

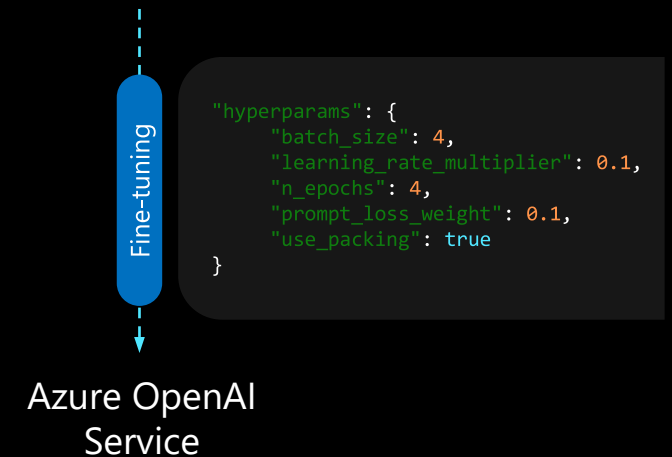
Commentary:  
1. Turner is so important defensively to  
2. Griner pulled way out, Hartley with  
3. At 1:54 remaining in the quarter, Pho

## Sample response

Main highlights:  
1. New York has had a strong run in the  
2. Phoenix leading by 1 point, 24-23  
3. New York Liberty's comeback has be

Optimize accuracy and latency to validate proof of concept fast

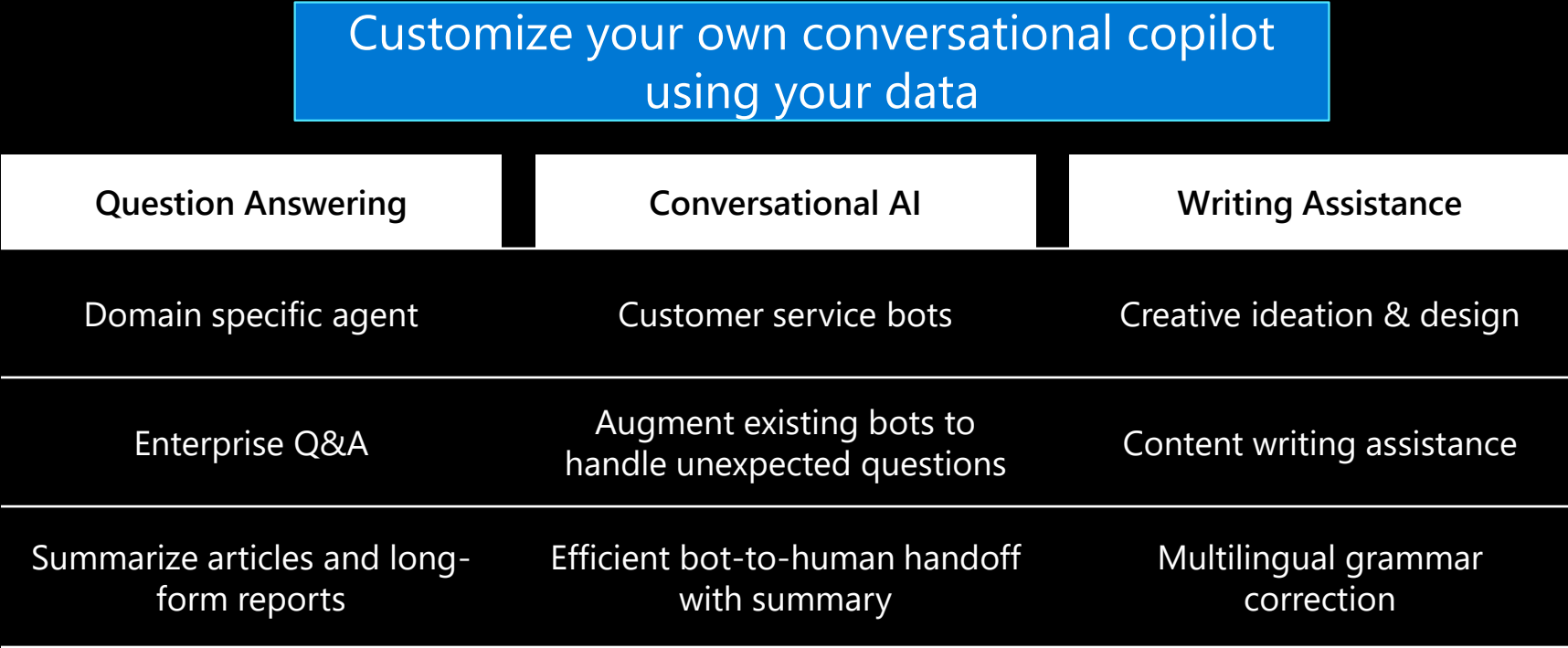
## Prompt and completion examples



# What is ChatGPT?

The ChatGPT model is a variant of GPT-3 specifically fine-tuned for conversational AI applications, such as dialogue systems and chatbots. The model can be used to build the next-generation of conversational experiences with the broader enterprise-wide context and awareness of the world around us.

**[NEW]** ChatGPT is available in Azure OpenAI Service, starting March 9, 2023.





GPT-4

# GPT-4 Example

User Can you explain why this is funny. Think about it step-by-step.



GPT-4 The comic is satirizing the difference in approaches to improving model performance between statistical learning and neural networks.

In statistical learning, the character is shown to be concerned with overfitting and suggests a series of complex and technical solutions, such as minimizing structural risk, reworking the loss function, and using a soft margin.

In contrast, the neural networks character simply suggests adding more layers to the model. This is often seen as a common solution to improving performance in neural networks, but it's also considered a simplistic and brute-force approach.

The humor comes from the contrast between the complexity and specificity of the statistical learning approach and the simplicity and generality of the neural network approach. The "But unironically" comment adds to the humor by implying that, despite being simplistic, the "stack more layers" approach is often effective in practice.

# Key facts about GPT-4

- Large multimodal model that accepts image and text inputs and emits text outputs
- It exhibits human-level performance on various professional and academic benchmarks
- GPT-4 exhibits similar capabilities on text-only inputs and documents with text and photographs, diagrams, or screenshots
- GPT-4 is more reliable, creative, and able to handle much more nuanced instructions than GPT-3.5
- GPT-4's image inputs are still a research preview and not publicly available
- Two model versions announced: gpt-4 and gpt-4-32k (32,768–context (about 50 pages of text))
- No details about the availability in the Azure OpenAI Service
- OpenAI Evals, OpenAI's framework for automated evaluation of AI model performance, is open-sourced

**Nutzung von Azure  
OpenAI**

# Azure AI

## Applications



Partner Solutions

## Application Platform

AI Builder



Power BI



Power Apps



Power Automate



Power Virtual Agents

## Scenario-Based Services

Applied AI Services



Bot Service



Cognitive Search



Form Recognizer



Video Indexer



Metrics Advisor



Immersive Reader

## Customizable AI Models

Cognitive Services



Vision



Speech



Language



Decision

Azure OpenAI Service

## ML Platform



Azure Machine Learning



Business Users



Developers & Data Scientists

# Microsoft Teams

## Meeting Recaps

The screenshot shows a Microsoft Teams meeting recap for a meeting titled "Sales Analysis Review". The meeting is currently "Tentative". The recap includes an agenda, meeting notes, and follow-up tasks.

**Meeting Details:** Sales Analysis Review (Tentative)

**Agenda:**

- Setup and introductions, @Daniela Mander, 5min
- CXE Updates from Tues meeting, @Bruno Zhao, 40min

**Meeting notes:**

- Leverage agile frameworks to provide a robust synopsis for high level overviews.
- Show more demos about the new features in next sync meeting.
- Refer to the last year's Q4 plan for comparison
- Add meeting notes

**Follow-up tasks:**

Task	Assigned to	Due date
<input type="checkbox"/> Follow up with Dave on files	@Daniela Mander	Tue, Oct 11 (1d left)
<input type="checkbox"/> Aadi Kapoor will check on incoming shipments	@Aadi Kapoor	Tue, Oct 11 (1d left)
<input type="checkbox"/> Daniela Mander will show demos about new features	Add assignee	Select date

**Tracking:**

- Daniela Mander (Organiser)
- Aadi Kapoor (Accepted)
- Ray Tanaka (Accepted)
- Mona Kane (Tentative)



# GitHub Copilot

Your AI pair programmer

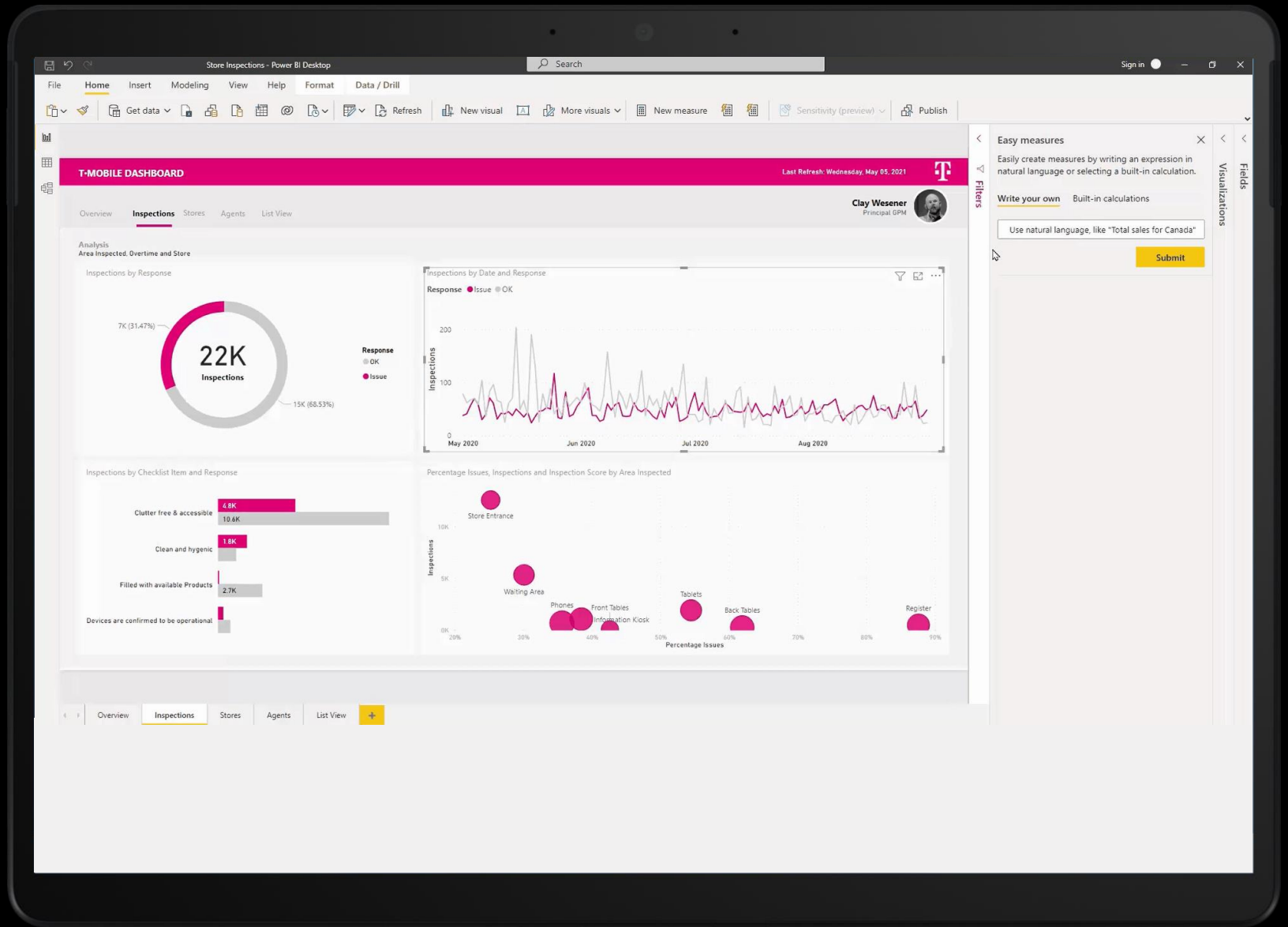
(powered by Codex)

```
19     screenshot = imageGrab.grab()
20     # Convert to text
21     text = image_to_string(screenshot)
22     # Parse text for email addresses
23     emails = re.findall(r'[\w\.-]+@[\w\.-]+', text)
24     return emails
25
26 def validate(addresses) :
27     ...
28
```

Generate & understand Code

# Power BI

Generating DAX expressions with natural language





Generate & understand Code



Let's start with your content


< Generate an image using a description ⓘ

Ombre cake decorated with flowers and fall foliage 



Tip  
Explore different camera angles  
"An aerial still of a seascape, Brazil"

Color ▾ Size ▾

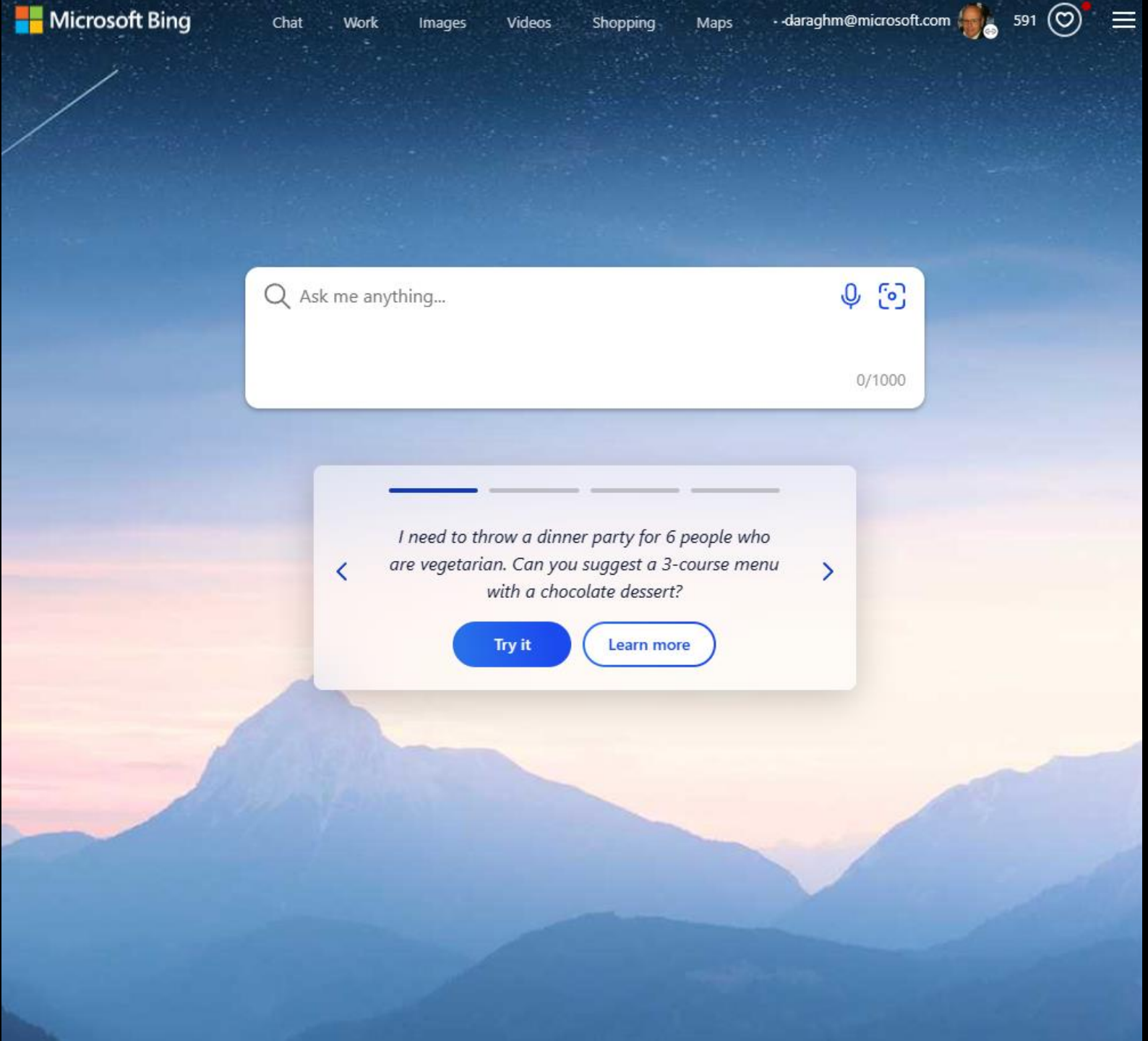


Continue editing Share

# Microsoft Bing

• Your co-pilot for the web

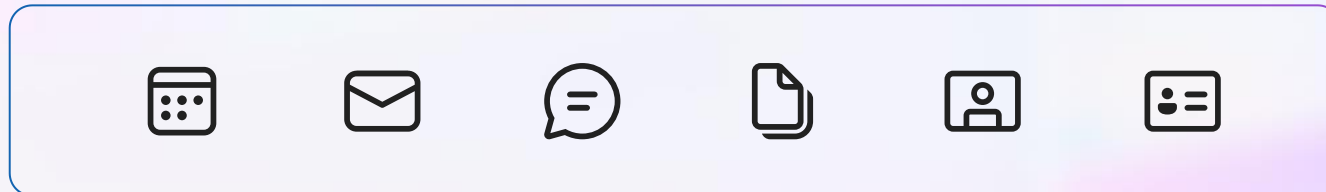
[Link](#)





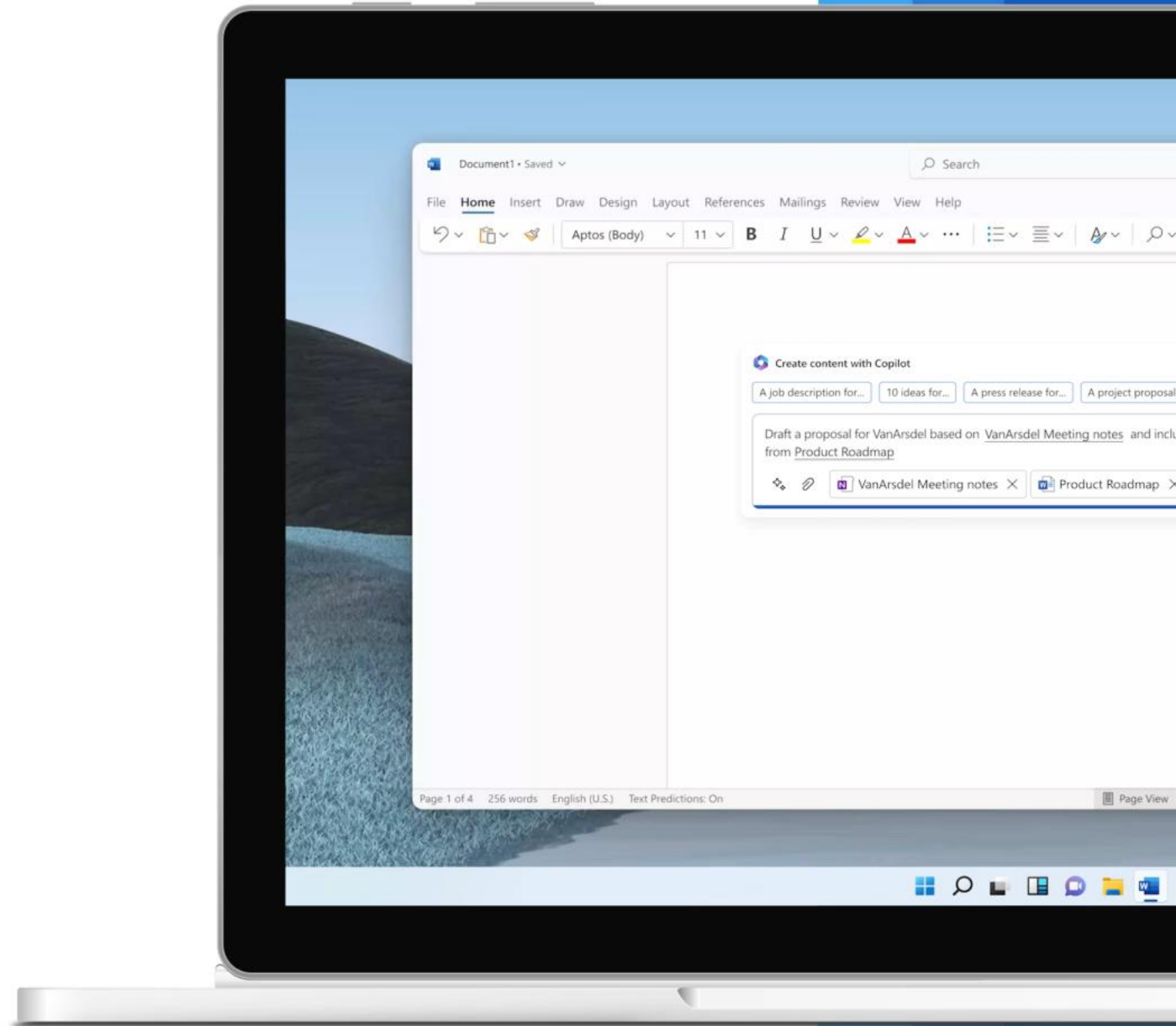
# Microsoft 365 Copilot

Access to your business **content** and **context**



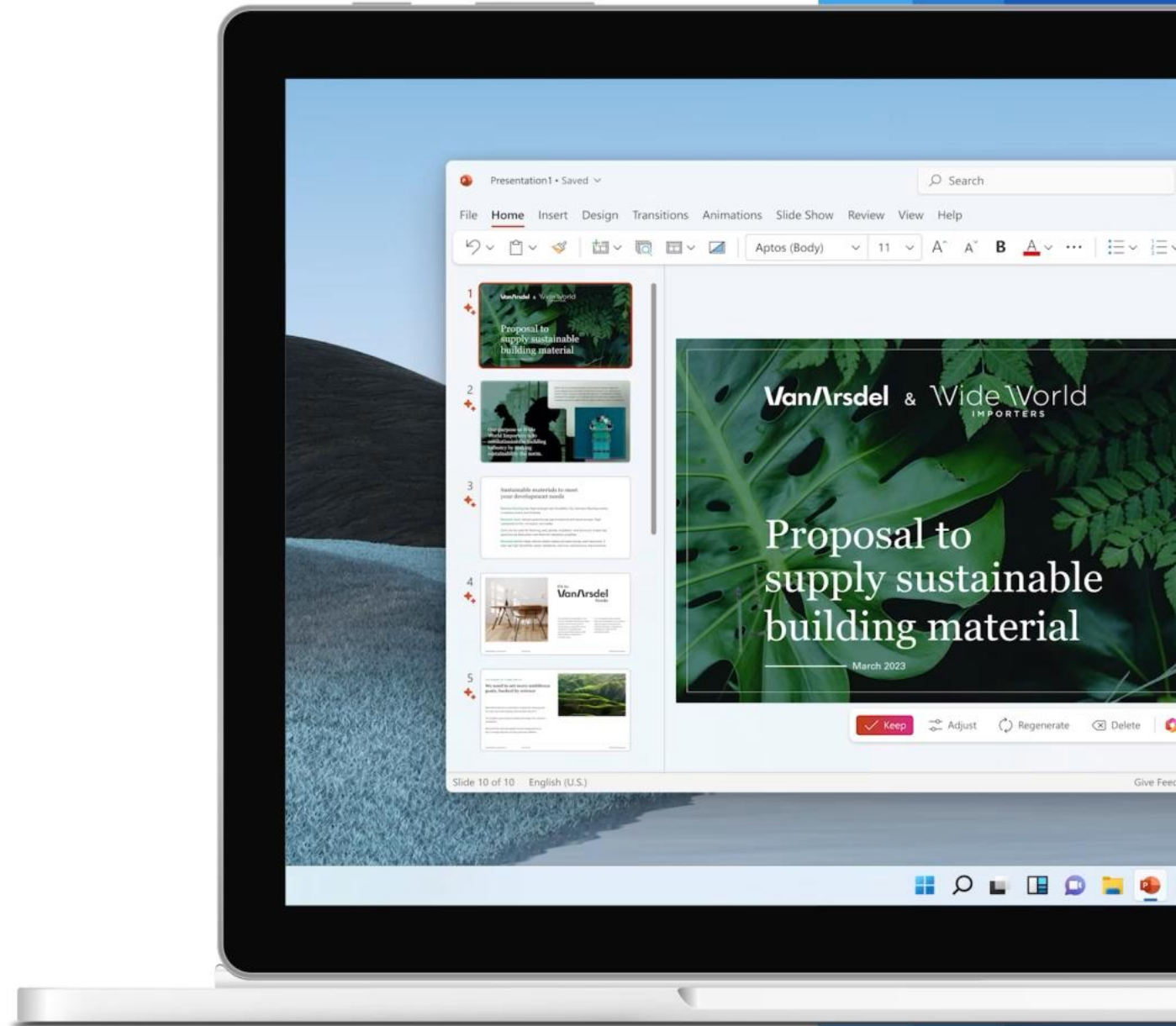
Microsoft 365

# Copilot in Word



Microsoft 365

# Copilot in PowerPoint



Demos

# Projekterfahrungen und Use Cases

# Azure OpenAI | Top 4 Capabilities & Use Cases



## Content Generation

Call Center Analytics: Automatically generate responses to customer inquiries

Generate personalised UI for your website



## Summarization

Call Center Analytics: Summary of customer support conversation logs

Subject Matter Expert Document Summarization (e.g. financial reporting, analyst articles)

Social Media Trends Summarization



## Code generation

Convert Natural Language to SQL (or vice versa) for telemetry data

Convert Natural Language to Query Proprietary Data Models

Code Documentation



## Semantic Search

Search reviews for a specific product / service

Information Discovery and Knowledge Mining

## Examples of Multiple Model Use Cases

End to End Call Center Analytics: Classification, Sentiment, Entity Extraction, Summarization and Email Generation

Customer 360: Hyper-personalisation using timely Summarization of customer queries & trends, Search, and Content Generation

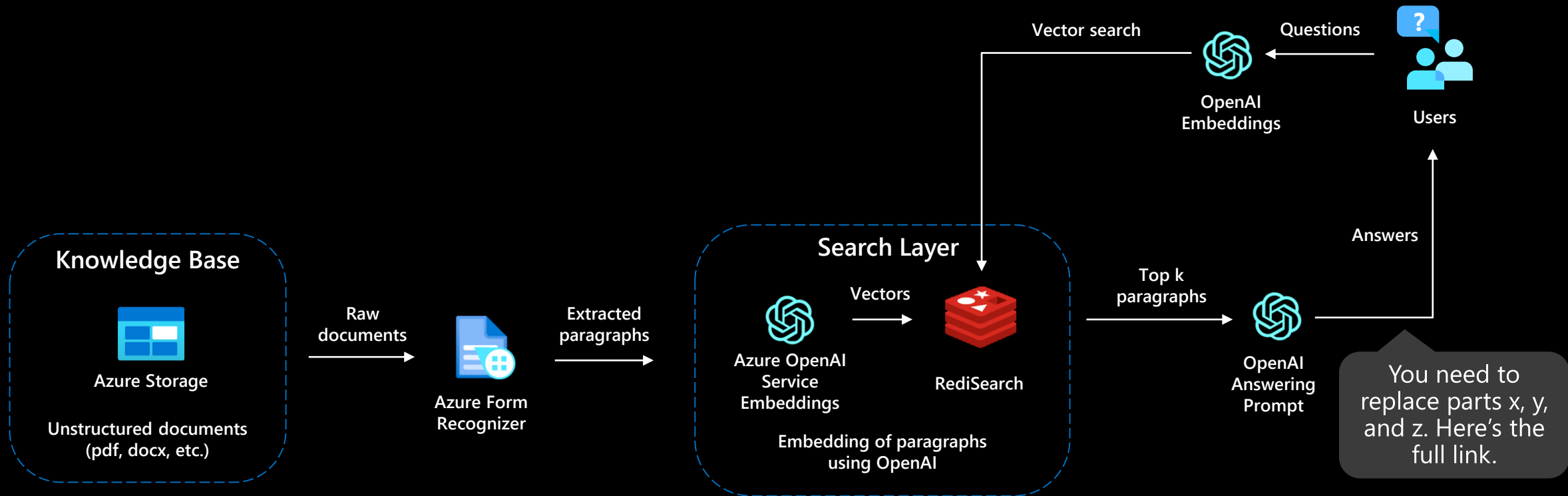
Business Process Automation: Search through structured & unstructured documentation, Generate Code to query data models, Content Generation



# Q&A with Semantic Answering with Azure OpenAI Service

Extract precise answers from your unstructured documents similar to ChatGPT

The xyz is broken on machine 5243, which part do I need to replace?



# Project Learnings


- 1 An OpenAI solution architecture consists of several components built around OpenAI
- 2 OpenAI Projects are Data Science Projects
- 3 Data sources must be understood (GPT-3 vs Codex)
- 4 There is a small risk for wrong answers (hallucinations),  
but there are ways to keep them to a minimum
- 5 A functioning ranking is crucial
- 6 The (automated) testing of generative responses is challenging

# How to improve the model output (Prompt Engineering)


1. Give clearer instructions




2. Split complex tasks into simpler subtasks



3. Structure the instruction to keep the model on task




4. Prompt the model to explain before answering



5. Ask for justifications of many possible answers, and then synthesize



6. Generate many outputs, and then use the model to pick the best one








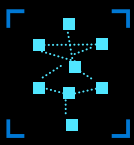
7. Fine-tune custom models to maximize performance



Zusammenfassung

# | Azure OpenAI Service | Is it right for you?

	<input checked="" type="checkbox"/>	<b>I need a general-purpose model that can handle multiple tasks.</b> e.g., translation+entity recognition+sentiment analysis
	<input checked="" type="checkbox"/>	<b>I need to generate human-like content, whilst preserving data privacy and security</b> e.g., abstractive summarization, content writing, paraphrasing, code
	<input checked="" type="checkbox"/>	<b>I need rapid prototyping and quick time to market for many use cases</b>
	<input checked="" type="checkbox"/>	<b>I could use a model with little or no training</b>
	<input checked="" type="checkbox"/>	<b>I want to explore solutions / use cases that have been described previously</b>



Vision



Speech

Azure OpenAI  
Service



Language



Decision

Azure AI Cognitive Services

## Next steps

- To learn more, please visit [Azure OpenAI Service](#)
- Have a use case? Speak to your Specialist / AE and sign up for access here: [aka.ms/oai/access](https://aka.ms/oai/access)
- [MS Learn "Explore Azure OpenAI"](#) training
- Watch [Pablo Castro show Azure OpenAI in action](#)

## Resources

- [Azure AI](#) solutions
- Microsoft's [Responsible AI principles](#)
- Our [approach to AI](#)



# Microsoft AI

