

Tag 2



Guten Morgen!

OpenAI und ChatGPT: Neuer Support in der IT!

Thomas Jäkel

brainymotion

Lead Trainer Cloud Infrastructure

Microsoft Certified Trainer since 1999

github.com/www42/openAI

brainymotion

← Physik → IT
FORTRAN
NT 4.0
↓
Azure
PowerShell

Heidelberg



Agenda

1. Was ist OpenAI?
2. Typische Modelle von OpenAI
3. Was ist ChatGPT?
4. Wie kann mich ChatGPT in der IT unterstützen?

Praktische Beispiele

Prompt Engineering

RAG

Beispiel

Contoso
Northwindtraders

Prompting

- Bing = GPT 4 + Internet
- GitHub Copilot

Enterprise Scenario

- Finetuning (Llama → Alpaca)
- Roles Assistant

User
→ System

Data
Source

Was ist OpenAI?



Die Geschichte von OpenAI

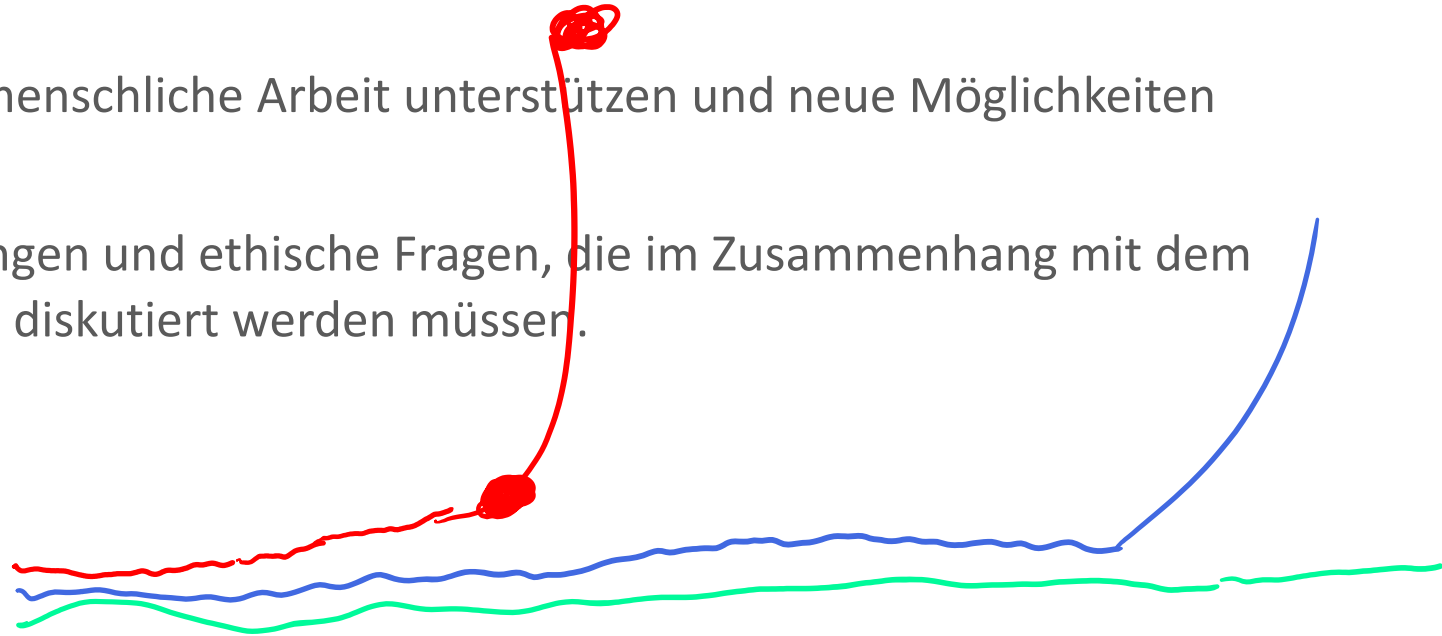
- OpenAI wurde im Jahr 2015 von Elon Musk, Sam Altman und anderen prominenten Persönlichkeiten gegründet.
- Ziel von OpenAI: Förderung und Entwicklung von Künstlicher Intelligenz (KI) für das Wohl der Menschheit.
- OpenAI hat sich zu einem der führenden Unternehmen im Bereich der KI-Forschung und -Entwicklung entwickelt.

Meilensteine von OpenAI

- 2016: OpenAI veröffentlicht das erste Paper "Unsupervised Pretraining for Sequence to Sequence Learning".
- 2017: OpenAI veröffentlicht "Deep Reinforcement Learning from Human Feedback", ein bahnbrechendes Paper zur KI-gesteuerten Spieleentwicklung.
- 2018: OpenAI entwickelt die dritte Iteration des Sprachmodells GPT (Generative Pre-trained Transformer).
- 2019: OpenAI startet das Projekt "GPT-2", ein noch leistungsfähigeres Sprachmodell, das für Aufsehen sorgt.
- 2020: OpenAI stellt GPT-3 vor, das bis dato größte und fortschrittlichste Sprachmodell.

Fortschritte in der KI-Forschung

- Die KI-Technologien von OpenAI haben das Potenzial, verschiedene Bereiche der Gesellschaft zu revolutionieren.
- Sie könnten die Produktivität steigern, menschliche Arbeit unterstützen und neue Möglichkeiten für personalisierte Dienste schaffen.
- Gleichzeitig gibt es auch Herausforderungen und ethische Fragen, die im Zusammenhang mit dem Einsatz von KI-Technologien von OpenAI diskutiert werden müssen.





OpenAI's Zukunftsvision

- OpenAI strebt an, KI-Technologien zu entwickeln, die sicher, transparent und allgemein zugänglich sind.
- OpenAI setzt sich dafür ein, dass die Vorteile von KI gerecht auf die gesamte Menschheit verteilt werden.
- Die Zukunftsvision von OpenAI beinhaltet die Zusammenarbeit mit anderen Organisationen und die Förderung von Forschung und Entwicklung im Bereich der KI, um eine positive Zukunft zu gestalten.



OpenAI & Microsoft

\$ $10 \cdot 10^9$



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



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

GPT-3.5 and GPT-4

Text

ChatGPT

Conversation

Codex

Code

DALL·E 2

Images

What is Azure OpenAI Service?

Azure Resource Manager
ARM

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

Cutting-edge generative AI models from OpenAI
+
Scalability, interoperability, and data protection
of Azure



Business
Users

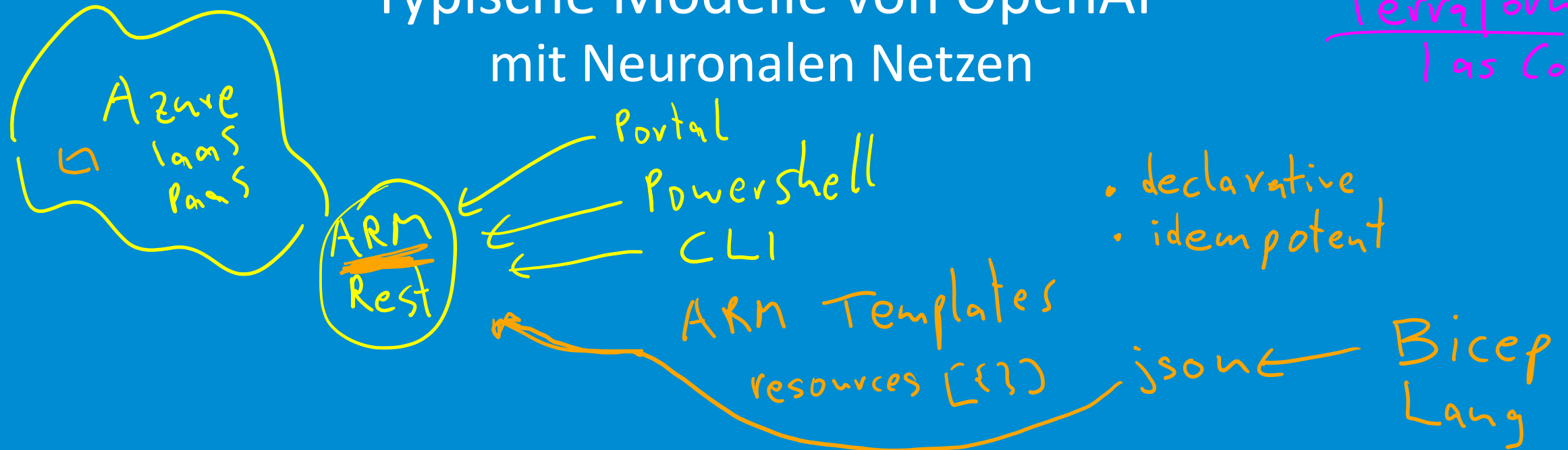


Developers &
Data Scientists

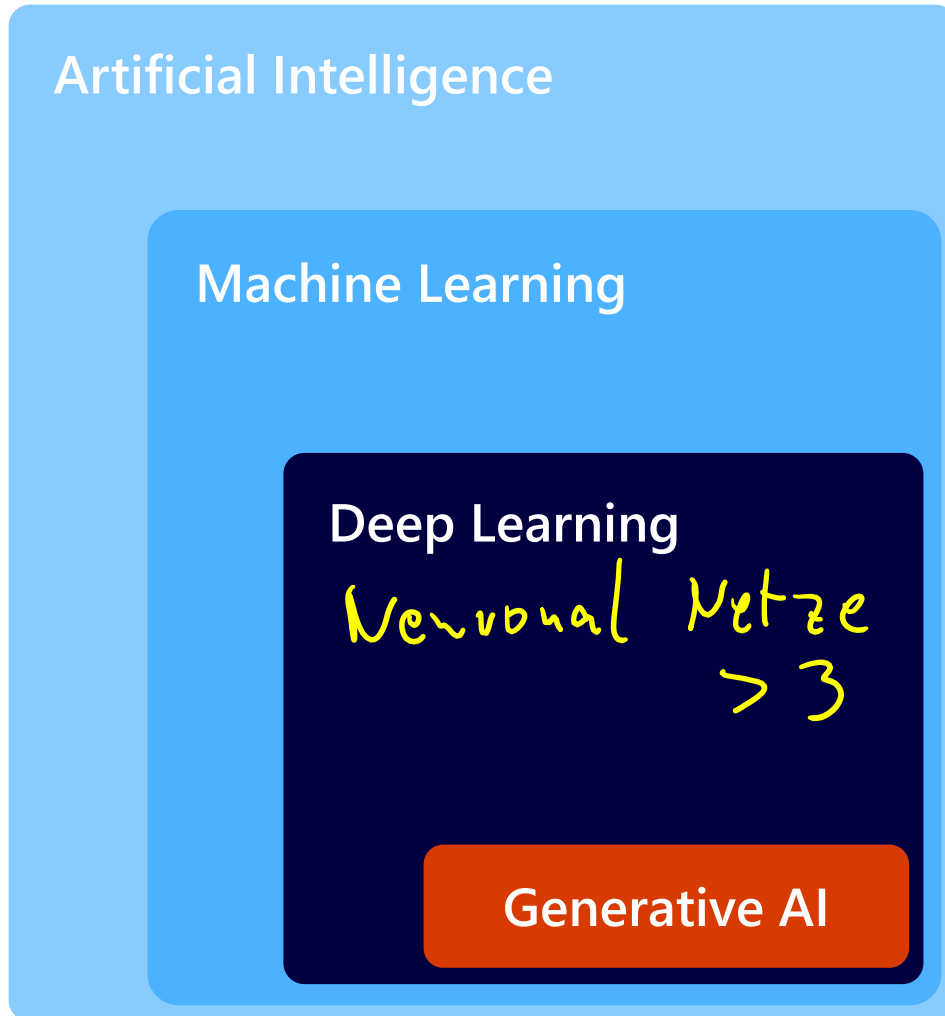
Semper Idem!

Typische Modelle von OpenAI mit Neuronalen Netzen

Terraform
I as Code



What is generative AI?



Norbert Wiener

1956

Artificial Intelligence

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

1997

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.

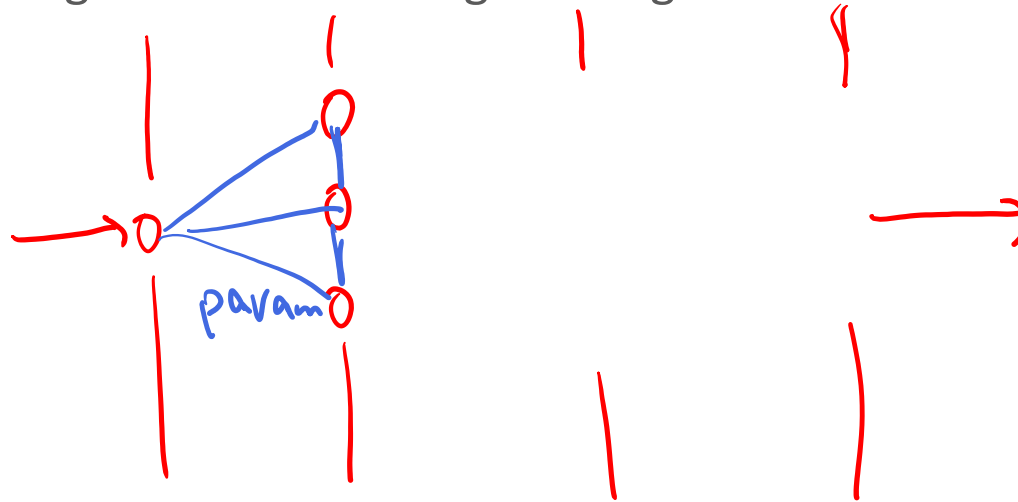
Azure OpenAI model capabilities

- The service include multiple models, optimized for specific tasks
- Models generate responses based on natural language *prompts*

	Language Generation	Code Generation	Image Generation
Prompt:	Write a haiku about marmalade	Write a Python function to add two numbers	Paint a pink fox in a field in the style of Monet
Output:	<i>Orange sunrise, sweet Spread on toast with morning tea A marmalade treat</i>	<pre>def add_two_numbers(a, b): return a + b</pre>	

Einführung in Neuronale Netze

- Neuronale Netze sind ein Kernkonzept in der Künstlichen Intelligenz.
- Sie bestehen aus künstlichen Neuronen, die miteinander verbunden sind und Informationen verarbeiten.
- Neuronale Netze werden verwendet, um komplexe Aufgaben wie Bilderkennung, Sprachverarbeitung und Entscheidungsfindung zu lösen.





GPT (Generative Pre-trained Transformer)

- GPT ist eine Modellreihe von OpenAI, die auf dem Transformer-Modell basiert.
- Das Modell wird mit großen Textdatensätzen trainiert und kann darauf basierend Texte generieren.
- GPT-4, die neueste Version, ist besonders bekannt für ihre Fähigkeit, qualitativ hochwertige und kohärente Texte zu erzeugen.

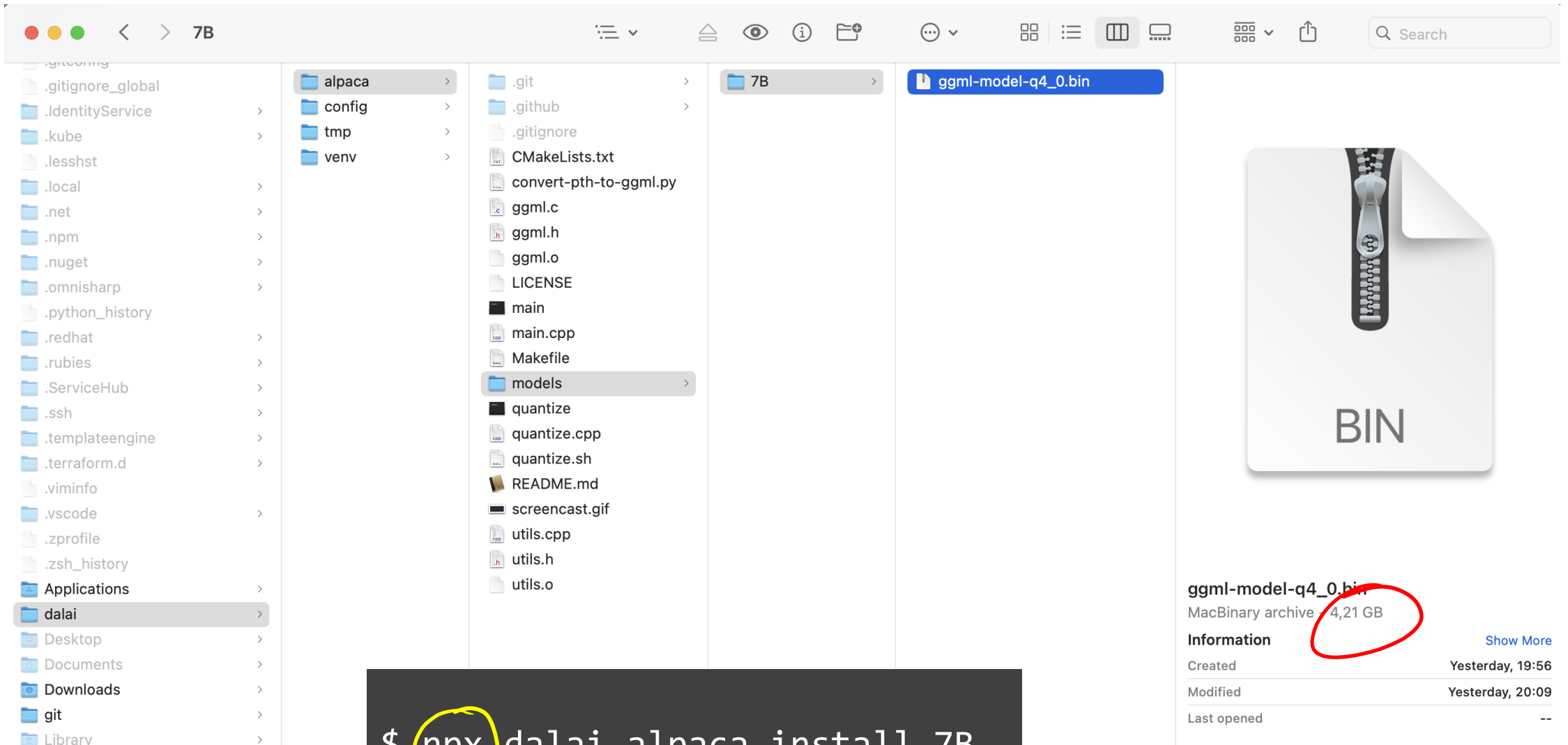
GPT (Generative Pre-trained Transformer)

• GPT 2	2019	1.5	$\times 10^9$	Parameters
• <u>GPT 3</u>	2020	17	$\times 10^9$	
• <u>GPT 4</u>	2023	120?	$\times 10^9$	



Fine tuning

Vergleich: LLaMA → Alpaca
Meta Stanford

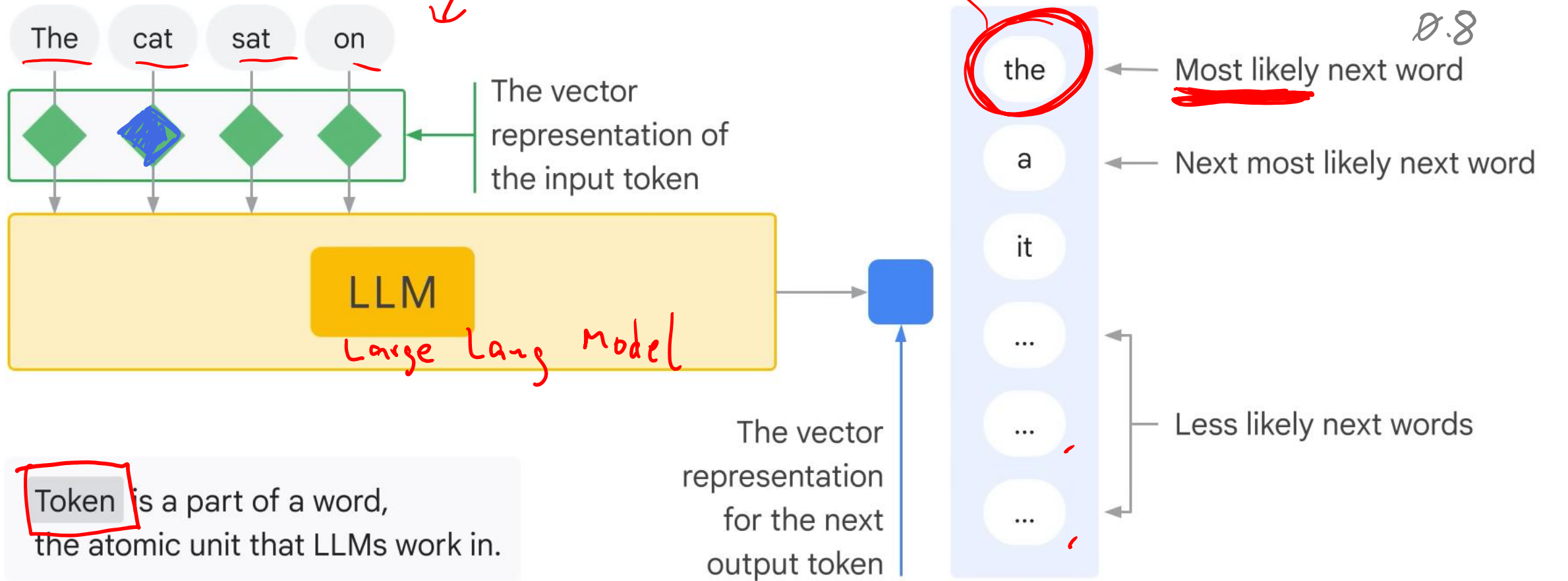


```
$ npx dalai alpaca install 7B
```

```
$ npx dalai serve
```

Attention

Generic language model - A next word predictor...





Tokenizer

The GPT family of models process text using **tokens**, which are common sequences of characters found in text. The models understand the statistical relationships between these tokens, and excel at producing the next token in a sequence of tokens.

You can use the tool below to understand how a piece of text would be tokenized by the API, and the total count of tokens in that piece of text.

GPT-3 Codex

Zum Kampf der Wagen und Gesänge,
Der auf Corinthus Landesenge
Der Griechen Stämme froh vereint,
Zog Ibycus, der Götterfreund.
5

Clear

Show example

Tokens

Characters

54

128

Zum Kampf der Wagen und Gesänge,
Der auf Corinthus Landesenge
Der Griechen Stämme froh vereint,
Zog Ibycus, der Götterfreund.
5

TEXT

TOKEN IDS

A helpful rule of thumb is that one token generally corresponds to ~4 characters of text for common English text. This translates to roughly $\frac{3}{4}$ of a word (so 100 tokens \approx 75 words).

If you need a programmatic interface for tokenizing text, check out our [tiktoken](#) package for Python. For JavaScript, the [gpt-3-encoder](#) package for node.js works for most GPT-3 models.

ChatGPT

- ChatGPT ist ein Modell von OpenAI, das entwickelt wurde, um natürliche und interaktive Gespräche mit Benutzern zu führen.
- Es basiert auf der GPT-Architektur und ermöglicht es, Fragen zu beantworten, Ratschläge zu geben und Informationen bereitzustellen.
- ChatGPT nutzt Neuronale Netze, um kontextabhängige und sinnvolle Antworten zu generieren.



DALL·E

- DALL·E ist ein Modell von OpenAI, das auf GPT-3 aufbaut und sich auf die Erzeugung von Bildern spezialisiert hat.
- Es kann aus textuellen Beschreibungen Bilder erstellen und ist in der Lage, kreative und detaillierte Bilder zu generieren.
- DALL·E hat das Potenzial, bei der Erstellung von Inhalten wie Illustrationen und Grafiken große Auswirkungen zu haben.



CLIP (Contrastive Language-Image Pretraining)

- CLIP ist ein Modell, das die Verbindung zwischen Bildern und Texten verstehen kann.
- Es wurde entwickelt, um gemeinsames Lernen von visuellen und sprachlichen Informationen zu ermöglichen.
- CLIP kann beispielsweise Bilder anhand von beschreibenden Texten klassifizieren und interpretieren.

OpenAI Five

- OpenAI Five war ein Projekt von OpenAI, das sich mit künstlicher Intelligenz im Bereich des Spiels Dota 2 beschäftigte.
- Das Modell nutzte Neuronale Netze, um menschliche Spieler herauszufordern und komplexe Spielstrategien zu entwickeln.
- OpenAI Five zeigte, wie Neuronale Netze in kompetitiven Umgebungen eingesetzt werden können.

MuseNet

- MuseNet ist ein Musik-Generierungsmodell von OpenAI.
- Es verwendet Neuronale Netze, um Musik in verschiedenen Stilen und Genres zu komponieren.
- MuseNet hat das Potenzial, Musiker und Komponisten bei der Ideenfindung und der Erstellung neuer Musikstücke zu unterstützen.



Reinforcement Learning Modelle

- OpenAI hat auch Modelle entwickelt, die auf dem Konzept des Reinforcement Learning basieren.
- Diese Modelle lernen, indem sie mit ihrer Umgebung interagieren und Belohnungen erhalten.
- Sie wurden verwendet, um komplexe Aufgaben wie das Spielen von Videospielen oder das Steuern von Robotern zu erlernen.



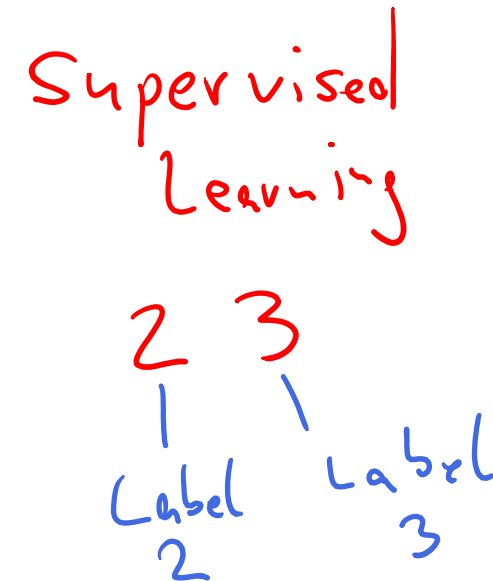
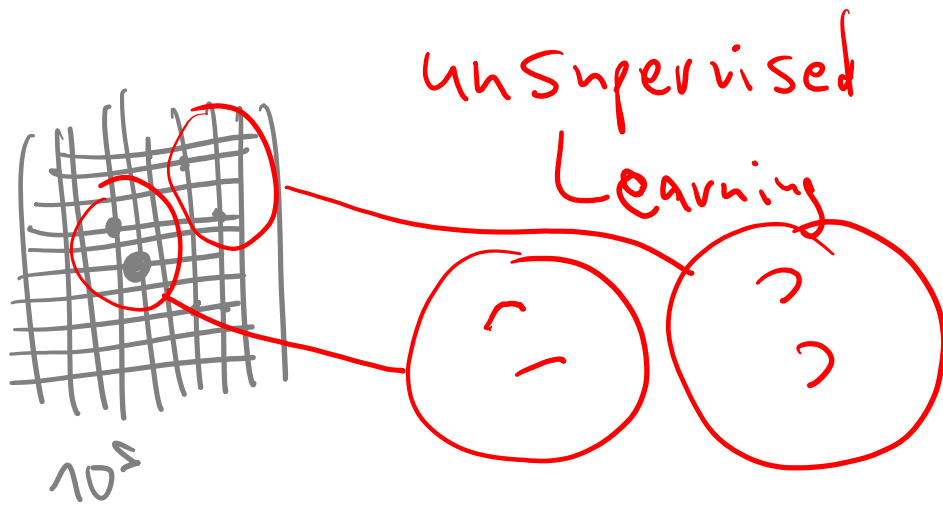
Zusammenfassung

- OpenAI hat eine Vielzahl von Modellen entwickelt, die auf Neuronalen Netzen basieren.
- Diese Modelle können Texte generieren, Bilder erstellen, die Verbindung zwischen Text und Bildern herstellen und sogar in Spielen eingesetzt werden.
- Die Forschung und Entwicklung von OpenAI auf dem Gebiet der Neuronalen Netze hat das Potenzial, viele Bereiche zu beeinflussen und die KI-Technologie weiter voranzutreiben.

Was ist ChatGPT?

Die Geschichte von ChatGPT

- ChatGPT ist ein Sprachmodell, das auf der GPT-3.5-Architektur von OpenAI basiert.
- Das Modell wurde entwickelt, um natürliche und interaktive Gespräche mit Benutzern zu führen.
- ChatGPT basiert auf umfangreichem Texttraining und kann verschiedene Aufgaben und Fragen bearbeiten.



GPT 4 1/2 Jahr

Die Entwicklung von ChatGPT

- Das erste GPT-Modell wurde 2018 von OpenAI veröffentlicht und beeindruckte die Fachwelt durch seine Fähigkeit, qualitativ hochwertigen Text zu generieren.
- Durch kontinuierliches Training und Verbesserungen wurde das Modell immer leistungsfähiger und vielseitiger.
- ChatGPT wurde entwickelt, um eine verbesserte Chatbot-Erfahrung zu bieten und in der Lage zu sein, komplexe Anfragen zu verstehen und zu beantworten.

Einsatzmöglichkeiten von ChatGPT

- ChatGPT findet Anwendung in verschiedenen Bereichen wie Kundensupport, Informationssuche und Unterhaltung.
- Es kann als persönlicher Assistent dienen, der Fragen beantwortet, Ratschläge gibt und Informationen bereitstellt.
- ChatGPT kann auch als kreative Schreibhilfe, zum Erstellen von Texten und zum Generieren von Ideen verwendet werden.

CodeX



Weiterentwicklungen und Herausforderungen

- OpenAI arbeitet kontinuierlich an Verbesserungen von ChatGPT und der KI-Technologie im Allgemeinen.
- Herausforderungen bestehen darin, das Modell genauer, verständlicher und weniger anfällig für Fehlinformationen zu machen.
- OpenAI bemüht sich auch um den verantwortungsvollen Einsatz von ChatGPT und berücksichtigt ethische Fragen und mögliche Missbrauchsrisiken.



Die Zukunft von ChatGPT

- OpenAI plant, ChatGPT weiterhin zu verbessern und neue Versionen mit noch fortschrittlicheren Fähigkeiten zu veröffentlichen.
- Die Zukunft von ChatGPT könnte mehr Interaktivität, besseres Verständnis von Kontext und spezifischeren Einsatz in verschiedenen Branchen umfassen.
- OpenAI strebt an, die KI-Technologie zugänglicher zu machen und die Vorteile von ChatGPT breit zu verteilen, während gleichzeitig mögliche Herausforderungen angegangen werden.

Wie funktioniert ChatGPT?



Grundlagen von ChatGPT

- ChatGPT basiert auf der GPT-3.5-Architektur von OpenAI.
- Es handelt sich um ein neuronales Netzwerk, das durch maschinelles Lernen trainiert wurde.
- Das Modell verwendet einen Transformer, um Texte zu verstehen und darauf zu antworten.



Training von ChatGPT

- ChatGPT wurde mit großen Mengen an Textdaten trainiert.
- Es wurden Texte aus dem Internet, Büchern, Artikeln und anderen Quellen verwendet.
- Durch das Training erlernt das Modell Sprachmuster, Kontextverständnis und Antwortgenerierung.

Eingabe und Verarbeitung

- Benutzer geben ihre Eingabe in natürlicher Sprache ein.
- Die Eingabe wird in Textform an ChatGPT übermittelt.
- Das Modell verarbeitet den Text, analysiert den Kontext und versucht, eine passende Antwort zu generieren.



Grundlagen der Textanalyse

- Tokenisierung: Aufteilung von Text in sinnvolle Einheiten
- Embedding: Transformation von Tokens in numerische Vektoren



Die Transformer-Architektur

- Struktur und Funktionsweise von Transformer-Blöcken
- Multi-Head-Aufmerksamkeitsmechanismus
- Feedforward-Netzwerk



Schritte der Textanalyse

- Tokenisierung des eingegebenen Textes
- Embedding der Tokens für semantische Repräsentation
- Durchlaufen mehrerer Transformer-Blöcke zur Erfassung von Beziehungen

Tokenisierung

- Die Eingabe wird in einzelne Tokens aufgeteilt, um sie für die Verarbeitung durch das Modell zu strukturieren.
- Jedes Token repräsentiert einen Teil der Eingabe, z. B. ein Wort oder ein Satzzeichen.

Embedding

- Jedes Token wird in einen numerischen Vektor umgewandelt, der als Embedding bezeichnet wird.
- Das Embedding erfasst die Bedeutung und den Kontext des Tokens im Bezug zur gesamten Eingabe.



Sequenzielle Verarbeitung

- Die Embeddings der einzelnen Tokens werden in einer bestimmten Reihenfolge an das Modell übergeben.
- Das Modell verarbeitet die Eingabe sequenziell und erfasst dabei die Beziehungen zwischen den einzelnen Tokens.

Anwendungen der Textanalyse

- Beantwortung von Fragen und Anweisungen
- Generierung von Texten basierend auf dem Kontext



Attention-Mechanismus

- Während der Verarbeitung der Eingabe verwendet das Modell einen Attention-Mechanismus.
- Dieser Mechanismus erlaubt dem Modell, auf bestimmte Teile der Eingabe zu fokussieren und relevante Informationen zu extrahieren.



Trainingsprozess und maschinelles Lernen

- Verarbeitung großer Mengen an Textdaten
- Lernen von Zusammenhängen und Mustern im Text



Kontextverständnis

- ChatGPT berücksichtigt den vorherigen Dialogverlauf, um den Kontext besser zu verstehen.
- Das Modell erfasst Informationen aus vorherigen Fragen, Antworten und Konversationen.
- Dadurch kann ChatGPT relevantere und kohärentere Antworten generieren.

Antwortgenerierung

- Basierend auf dem verarbeiteten Text und dem gelernten Wissen generiert ChatGPT eine Antwort.
- Das Modell wählt aus einer Vielzahl von möglichen Antwortoptionen die wahrscheinlichste aus.
- Die generierte Antwort wird dem Benutzer präsentiert.

Feedbackschleifen

- ChatGPT kann von Benutzerfeedback lernen und sich verbessern.
- Positive und negative Rückmeldungen helfen dem Modell, seine Antworten anzupassen und zukünftige Interaktionen zu optimieren.
- Dieser Lernprozess trägt zur Weiterentwicklung und Verfeinerung von ChatGPT bei.



Herausforderungen und Verbesserungen

- ChatGPT hat einige Herausforderungen wie das Risiko von Fehlinformationen oder unangemessenen Inhalten.
- OpenAI arbeitet kontinuierlich daran, das Modell zu verbessern und solche Probleme zu minimieren.
- Durch Feedback und Iterationen werden neue Versionen entwickelt, um ChatGPT leistungsfähiger und verlässlicher zu machen.



ChatGPT



Interact with our flagship language models in a conversational interface

DALL·E



Create realistic images and art from a description in natural language

API

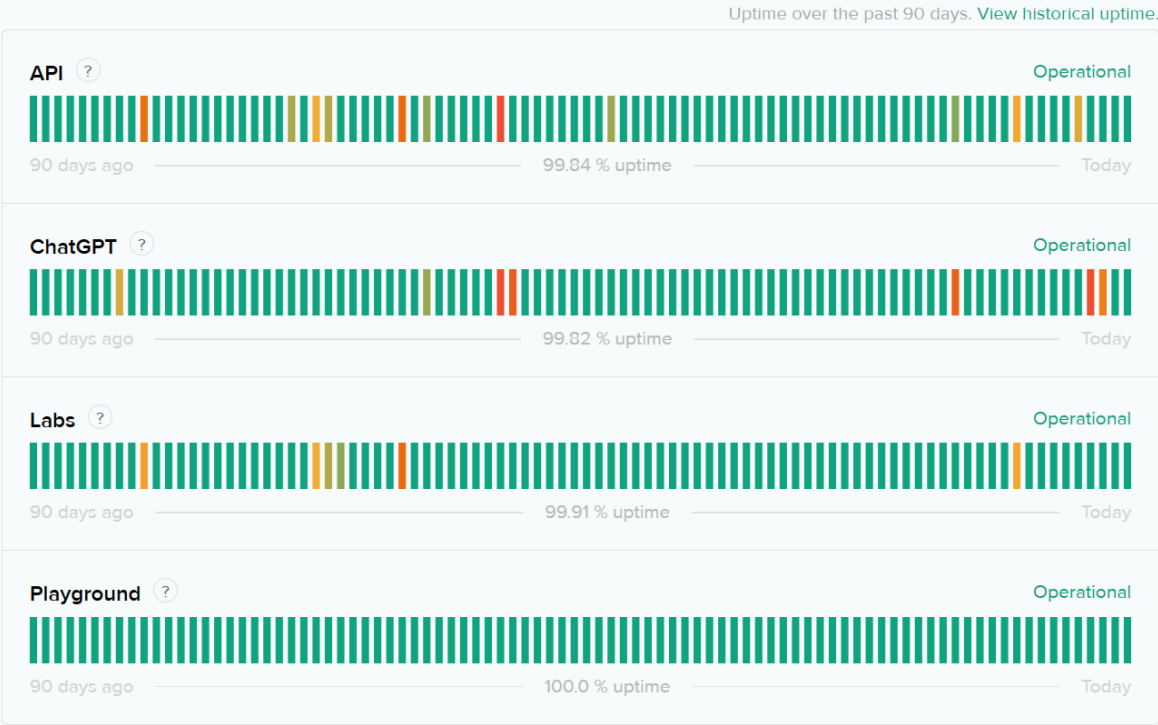


Integrate OpenAI models into your application or business



SUBSCRIBE TO UPDATES

All Systems Operational



Elevated error rate for ChatGPT

Incident Report for OpenAI

- Resolved

This incident has been resolved.
Posted 2 days ago. Jul 12, 2023 - 09:20 PDT
- Monitoring

A fix has been implemented and we are monitoring the results.
Posted 2 days ago. Jul 12, 2023 - 09:10 PDT
- Update

Some users are seeing a "Failed to get service status" error. We are investigating
Posted 2 days ago. Jul 12, 2023 - 08:37 PDT
- Update

We are continuing to work on the problem and things are improving. Logins are starting to work.
Posted 2 days ago. Jul 12, 2023 - 08:14 PDT
- Investigating

ChatGPT is currently unavailable for most users, and may be slow to load. We've disabled login while we recover the service.
Posted 2 days ago. Jul 12, 2023 - 07:43 PDT

This incident affected: ChatGPT.

Power shell 5
7

```
$Key = 'sk-xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx'
```

```
$Headers = @{
    "Authorization" = "Bearer $Key"
    "Content-Type" = "application/json"
}
```

```
$Body = ConvertTo-Json @{
    "model" = 'text-davinci-003'
    "prompt" = "This is a test"
}
```

```
Invoke-WebRequest `
-Uri "https://api.openai.com/v1/completions"
-Method POST `
-Headers $Headers `
-Body $Body
```

← Back Tic

Wie kann ChatGPT die moderne IT unterstützen?



Einführung

- ChatGPT, entwickelt von OpenAI, ist ein fortschrittliches Sprachmodell, das in der Lage ist, natürliche und interaktive Gespräche zu führen.
- Durch seine Fähigkeit, menschenähnliche Texte zu generieren, kann ChatGPT die moderne IT in verschiedenen Bereichen unterstützen.



Kundensupport und Helpdesk

- ChatGPT kann als virtueller Assistent eingesetzt werden, um Kunden bei technischen Fragen und Problemen zu unterstützen.
- Durch die Bereitstellung von genauen und relevanten Informationen kann ChatGPT die Effizienz des Kundensupports verbessern.

KDC

en / de

Hauptverteilungsmittelpunkt
key distribution center



Fehlerbehebung und Diagnose

- Bei der Fehlerbehebung und Diagnose von IT-Problemen kann ChatGPT als Wissensbasis dienen.
- Es kann Fragen zu gängigen Problemen beantworten, Lösungen vorschlagen und den Benutzern bei der Behebung von Schwierigkeiten helfen.



Windows Server
→ Web Server install 3 Möglichkeiten

Code-Snippets und Programmierung

- Entwickler können ChatGPT verwenden, um nach Code-Snippets, Beispielen oder Programmierhilfe zu fragen.
- Es kann dabei helfen, schnelle Lösungen und Syntaxvorschläge zu erhalten, um die Programmierarbeit effizienter zu gestalten.

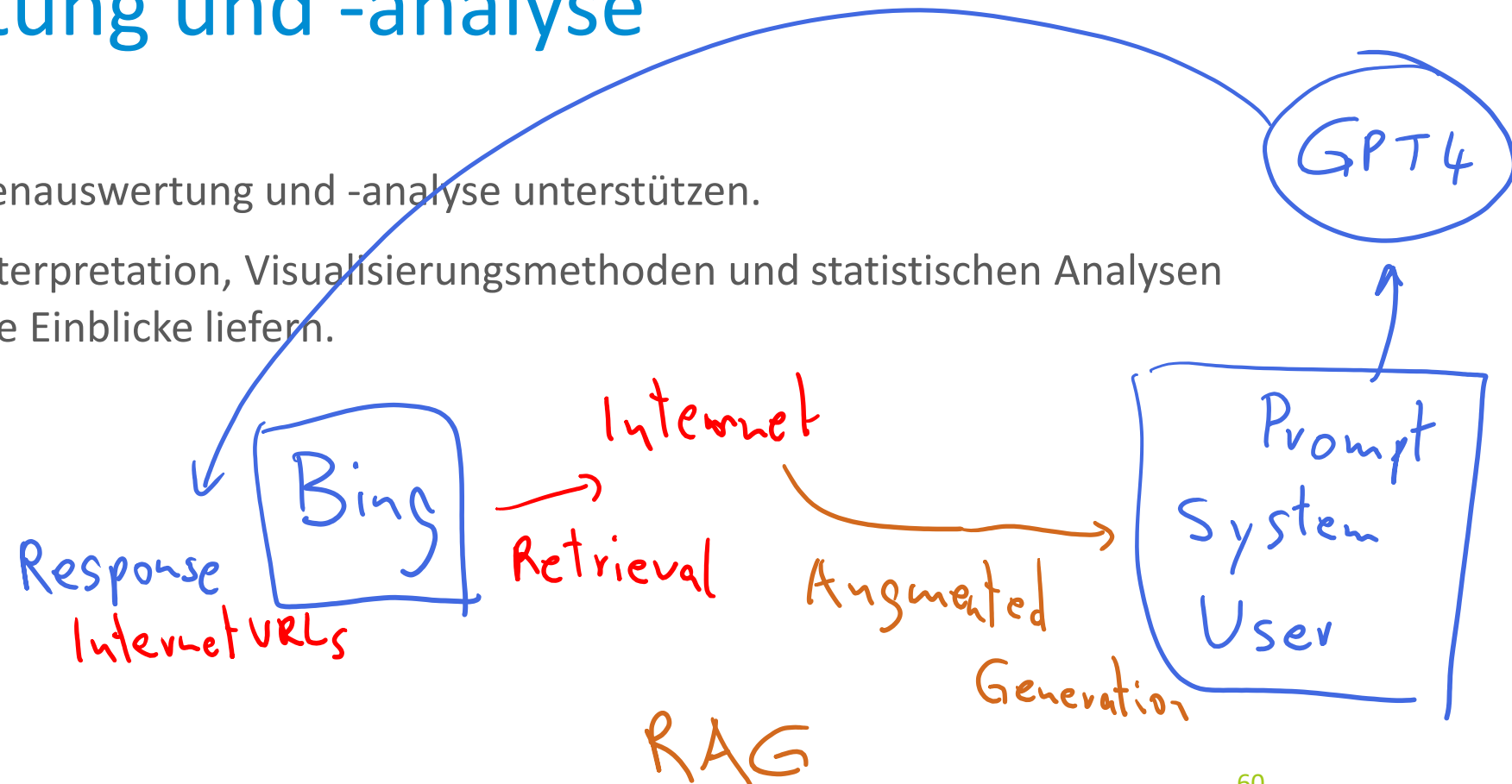


Wissensmanagement

- ChatGPT kann als leistungsfähiges Werkzeug für das Wissensmanagement in IT-Unternehmen dienen.
- Es kann verwendet werden, um firmeninternes Wissen zu konsolidieren, FAQs zu erstellen und den Zugriff auf wichtige Informationen zu erleichtern.

Datenauswertung und -analyse

- ChatGPT kann bei der Datenauswertung und -analyse unterstützen.
- Es kann Fragen zu Dateninterpretation, Visualisierungsmethoden und statistischen Analysen beantworten und wertvolle Einblicke liefern.





IT-Sicherheit und Datenschutz

- ChatGPT kann Unternehmen dabei unterstützen, Sicherheitslücken zu erkennen und bewährte Verfahren im Bereich IT-Sicherheit und Datenschutz zu implementieren.
- Es kann Schulungen und Richtlinien für Mitarbeiter bereitstellen und bei der Sensibilisierung für Sicherheitsrisiken helfen.



Automatisierung und Prozessoptimierung

- Durch den Einsatz von ChatGPT können Unternehmen Prozesse automatisieren und optimieren.
- Es kann wiederkehrende Aufgaben übernehmen, wie z. B. das Generieren von Berichten, das Verfolgen von Projekten oder das Erstellen von Dokumentationen.



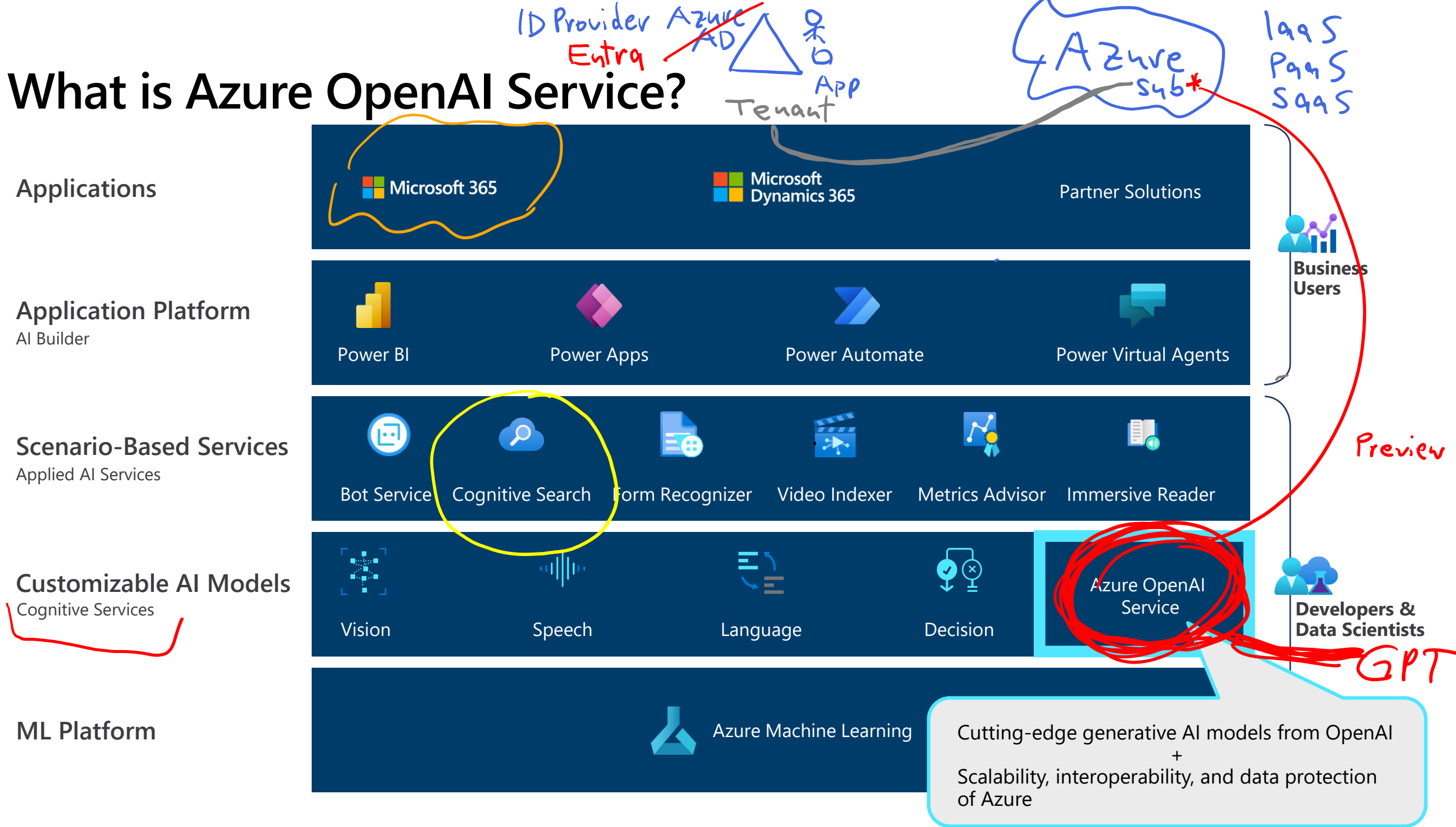
Forschung und Entwicklung

- In der IT-Forschung und -Entwicklung kann ChatGPT als kreativer Assistent dienen.
- Es kann Ideen für neue Technologien, Algorithmen oder Lösungsansätze liefern und bei der Entwicklung innovativer IT-Konzepte unterstützen.

Zusammenfassung

- ChatGPT bietet vielfältige Möglichkeiten, die moderne IT zu unterstützen.
- Von Kundensupport bis hin zur Prozessautomatisierung kann es Unternehmen helfen, effizienter zu arbeiten und bessere IT-Dienstleistungen anzubieten.

What is Azure OpenAI Service?



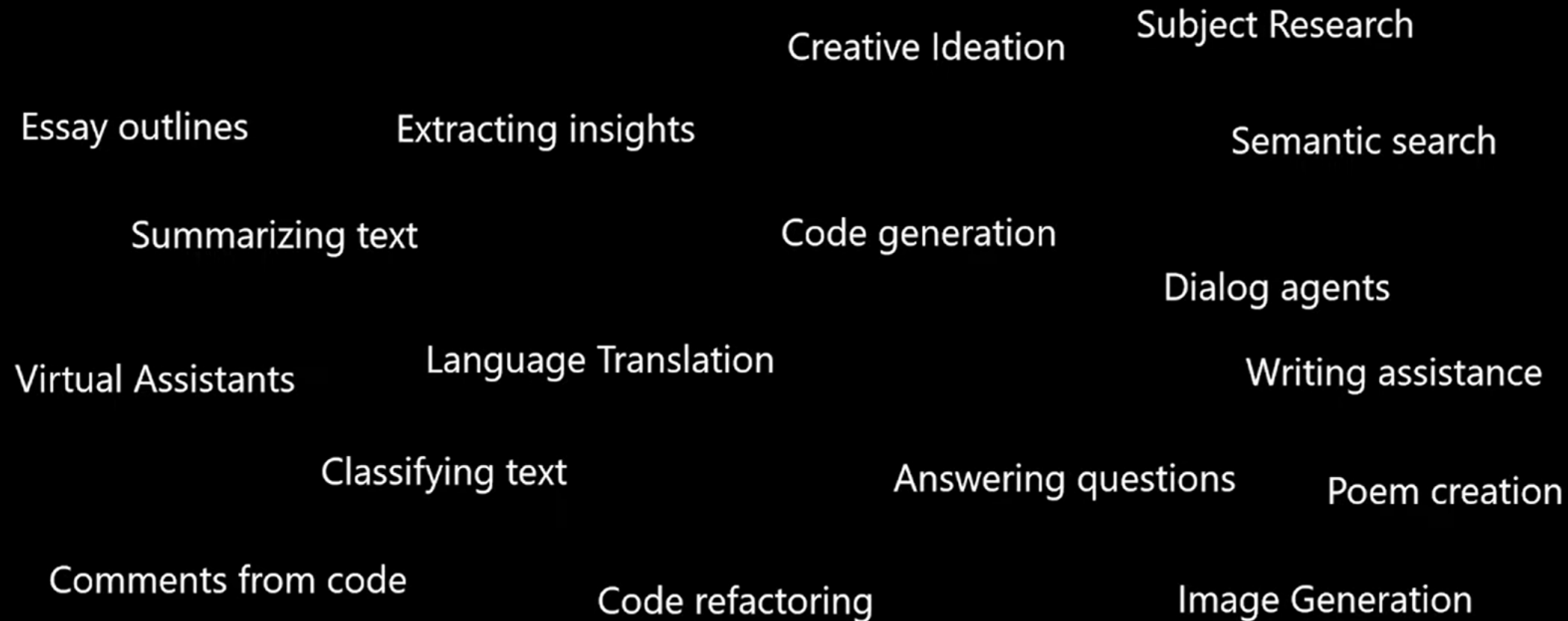
Using prompts to get completions from models

Task	Prompt	Completion
Classifying content	Tweet: I enjoyed the training course. Sentiment:	Positive
Generating new content	Write a poem about databases	Databases, oh databases, You keep our information safe, From the small to the large, You store our data in a place.
Transformation/Translation	English: Hello French:	Bonjour
Summarization	Scotland is [<i>long description of Scotland...</i>] Summarize the previous text	Scotland is [<i>summarized description...</i>]
Continuation	One way to grow tomatoes is to	start with seeds...
Question answering	How many moons does Earth have?	Earth has one moon.
Chat	<i>Setup, followed by messages...</i>	<i>A sequence of relevant responses</i>

| Azure OpenAI | GPT-3 Family of Models

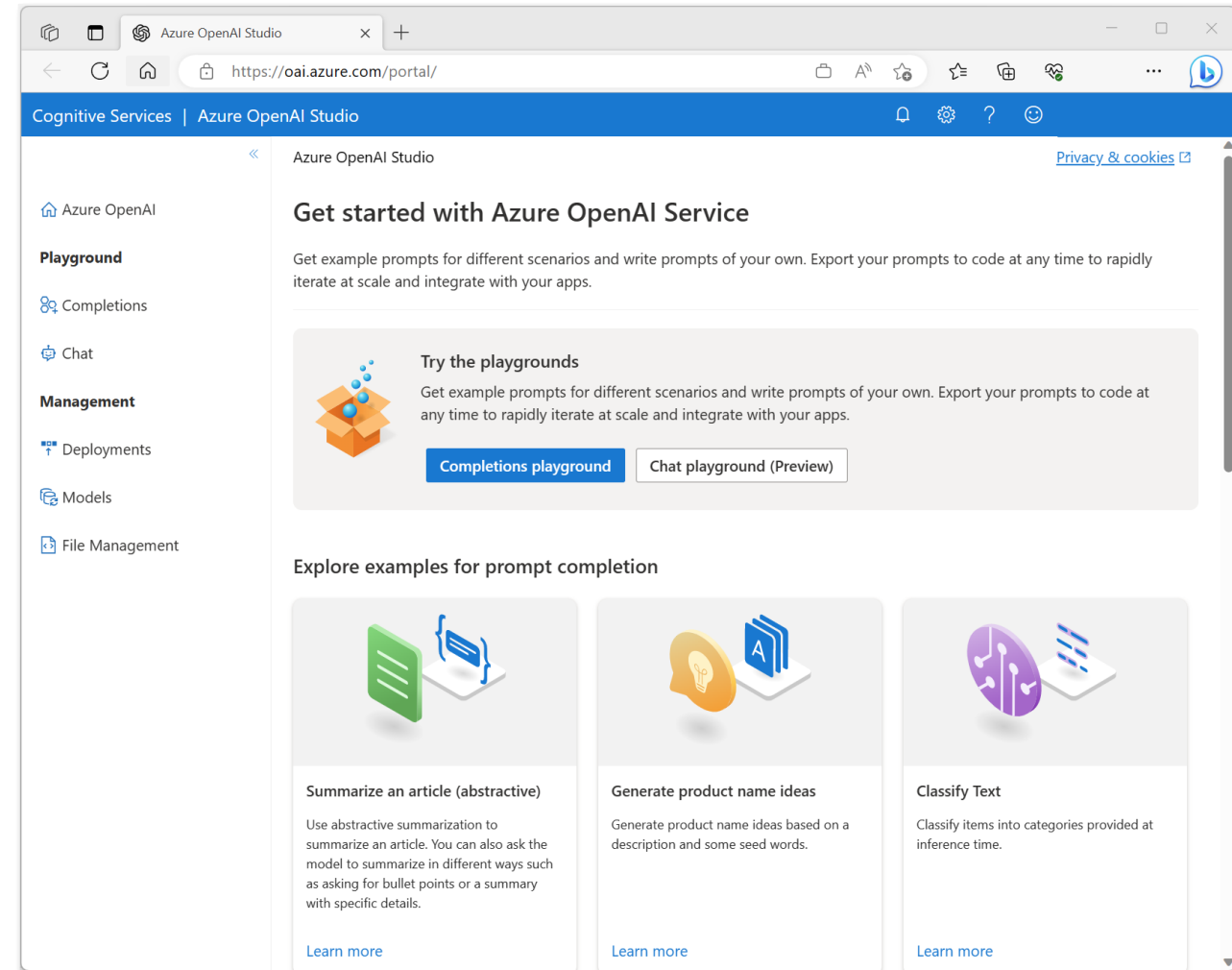
Model	Request	Description, performance, cost	Use cases
Davinci	4,000 tokens	Most capable GPT-3 model. Can do any task the other models can do, often with <i>higher quality, longer output</i> and <i>better instruction-following</i> .	Complex intent, cause and effect, summarization for audience
Curie	2048 tokens	Very capable , but <i>faster</i> and <i>lower cost</i> than Davinci.	Language translation, complex classification, text sentiment, summarization
Babbage	2048 tokens	Capable of straightforward tasks, <i>very fast</i> , and <i>lower cost</i> .	Moderate classification, semantic search classification
Ada	2048 tokens	Capable of very simple tasks, usually the <i>fastest</i> model in the GPT-3 series, and <u>lowest cost</u> .	Parsing text, simple classification, address correction, keywords

| Azure OpenAI Service Capabilities



Azure OpenAI Studio

- Web portal for working with Azure OpenAI models:
<https://oai.azure.com/>
- View and deploy base models
- Manage fine tuning and data files for custom models
- Test models in visual playgrounds:
 - **Completions** (GPT-3 and earlier models)
 - **Chat** (GPT-3.5-Turbo and later models)



More efficient methods of tuning

Parameter-Efficient Tuning Methods (PETM)

Methods for tuning an LLM on your own custom data without duplicating the model. The base model itself is not altered. Instead, a small number of add-on layers are tuned, which can be swapped in and out at inference time.

Prompt Tuning

One of the easiest Parameter Efficient Tuning Methods.

ChatGPT

Eine Bereicherung für die täglichen Aufgaben