

Text-generation

# Deep Learning Models

---

## Cellarius



# GPT

Generative  
Pre-trained  
Transformer

# OpenAI

GPT-1, GPT-2, GPT-3, Codex





# OpenAI



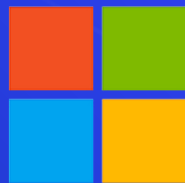
Mitgründer, CEO  
Sam Altman



Mitgründer, Spender  
Elon Musk



# OpenAI



Microsoft



**Spender**  
**Cloud Provider**

**Spender**

# OpenAI - Timeline

Gründung  
Non Profit  
\$1 Milliarde  
Investment

2015

2016

2017

2018

2019

2020

2021

GPT 2  
For Profit  
\$1 Milliarde  
Investment

GPT 3 Full  
Release  
Codex

GPT 1  
Musk verlässt  
Firma

GPT 3 Private  
Beta  
HQ mit Neuralink



# NLP Neuheit GPT

## Superverised Models (alt)

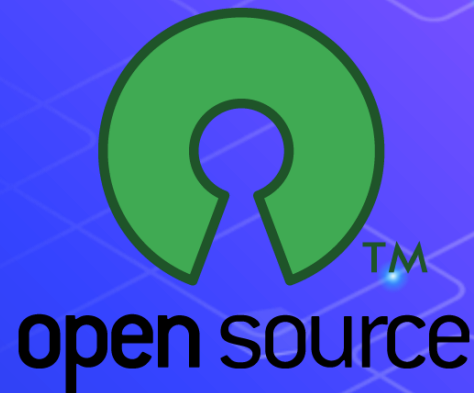
- ⬡ Auf spezifische Aufgaben trainiert
- ⬡ Brauchen sehr spezielle Datensätze
- ⬡ Nur für spezifische Aufgaben geeignet

## Semi-Superverised Models (neu)

- ⬡ Unsupervised Pre-Training
- ⬡ Supervised Fine-Tuning
- ⬡ Können mit allgemeinen Datensätzen trainiert werden
- ⬡ Für verschiedenste Aufgaben geeignet

# GPT-2

- ⬡ 1.5 Milliarden Parameter
- ⬡ Trainiert mit WebText Reddit Scrape (40 GB)
- ⬡ Fragen beantworten
- ⬡ Texte zusammenfassen
- ⬡ Texte übersetzen





# GPT-3

- ⬡ 175 Milliarden Parameter
- ⬡ Trainiert mit 5 verschiedenen Datensätzen (45 TB)
- ⬡ Humor
- ⬡ Programmerrätsel
- ⬡ Berühmte Schriftsteller imitieren



# The API

- ⬡ 18.11.2021 Full Release
- ⬡ Öffentlich zugänglich
- ⬡ Kostenpflichtig



# GPT-3 Base Models

Model	Parameteranzahl	Trainingskosten für 1000 Token	Laufzeitkosten für 1000 Token
Ada	2,7 Milliarden	\$ 0,0004	\$ 0,0016
Babbage	6,7 Milliarden	\$ 0,0006	\$ 0,0024
Curie	13 Milliarden	\$ 0,0030	\$ 0,0120
Davinci	175 Milliarden	\$ 0,0300	\$ 0,1200

# Codex

- ⬡ GPT-3 Abkömmling
- ⬡ Sprachbefehle in Code umwandeln
- ⬡ Kann alle Sprachen von PHP bis Shell
- ⬡ Python funktioniert am besten



# GPT-3

## Beispiele

<https://beta.openai.com/examples>

<https://gpt3demo.com/>



# Codex

## Beispiele

<https://copilot.github.com/>



# EleutherAI

GPT-Neo, GPT-J



# EleutherAI

- ⬡ Gegründet: 2020
- ⬡ Kollektiv von Freiwilligen, zur Entwicklung einer Open Source KI





# The Pile

- ⬡ 800 GB Dataset
- ⬡ 22 hoch qualitative Subsets (z.B.: wissenschaftliche Arbeiten, Bücher, GitHub-Repos)
- ⬡ Sehr große Vielfalt an Daten

# GPT-Neo

- ⬡ 1,3 / 2,7 Milliarden Parameter
- ⬡ Selbe Architektur wie GPT-2
- ⬡ Open Source



# GPT-J-6B

- ⬡ 6 Milliarden Parameter
- ⬡ xmap: Aufteilen des Modells auf Recheneinheiten
- ⬡ Basiert auf JAX
  - ⬡ Google
  - ⬡ Deeplearning (einfach)
  - ⬡ Just in time-compiler (schnell)
- ⬡ Beste Performance bei down-streaming tasks
- ⬡ Open Source

# Token

<https://beta.openai.com/tokenizer>



# KI Playground

<https://beta.openai.com/playground>



# KI Training



# Wie trainiert man eine KI?

- ⬡ Neuronales Netz aufsetzen
- ⬡ Trainingsdaten einspeisen, auf die man den Algorithmus trainieren möchte

# Wie trainiert man eine KI?

- ⬡ Trainiertes Modell mit neuen Testdaten testen
- ⬡ Trainiert und getestetes Modell kann verwendet werden



# Kosten

Selber Hosten oder Cloud



# Cloud

- ⬡ Einfaches Trainieren/Ausführen
- ⬡ Einfaches Skalieren
- ⬡ Auf längere Zeit teurer

GPT-3 Ada	\$0.0008 / 1K tokens
GPT-3 Babbage	\$0.0012 / 1K tokens
GPT-3 Curie	\$0.0060 / 1K tokens
GPT-3 Davinci	\$0.0060 / 1K tokens

# Selber Hosten

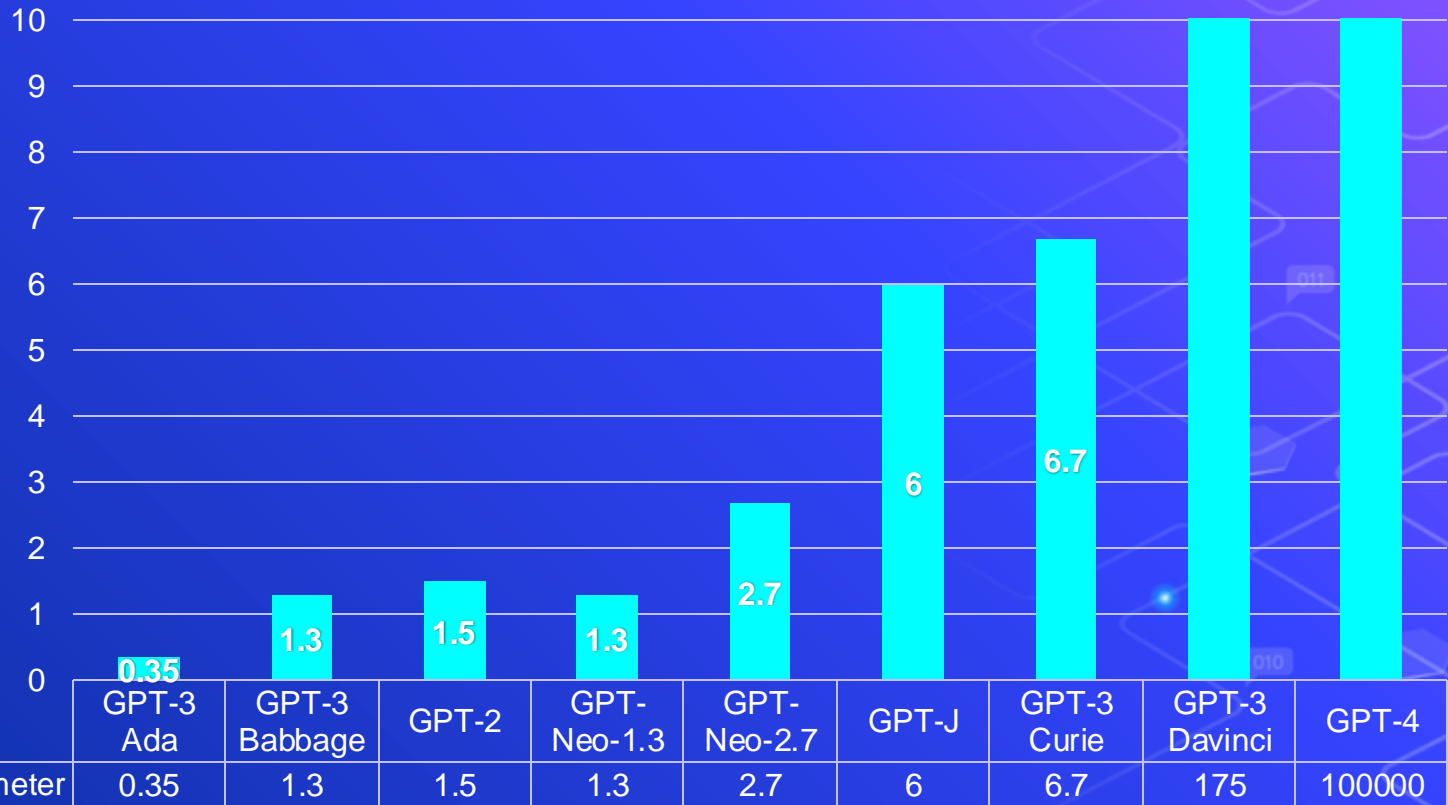
- ⬡ Man muss sich mit Deep Learning Hardware beschäftigen
- ⬡ Nach 1-2 Jahre deutlich Günstiger als Cloud
- ⬡ Sehr hohe Anschaffungskosten
  - Ab 5000€ bis 1000000€+

# Zukunft

Wie geht es weiter mit OpenAI

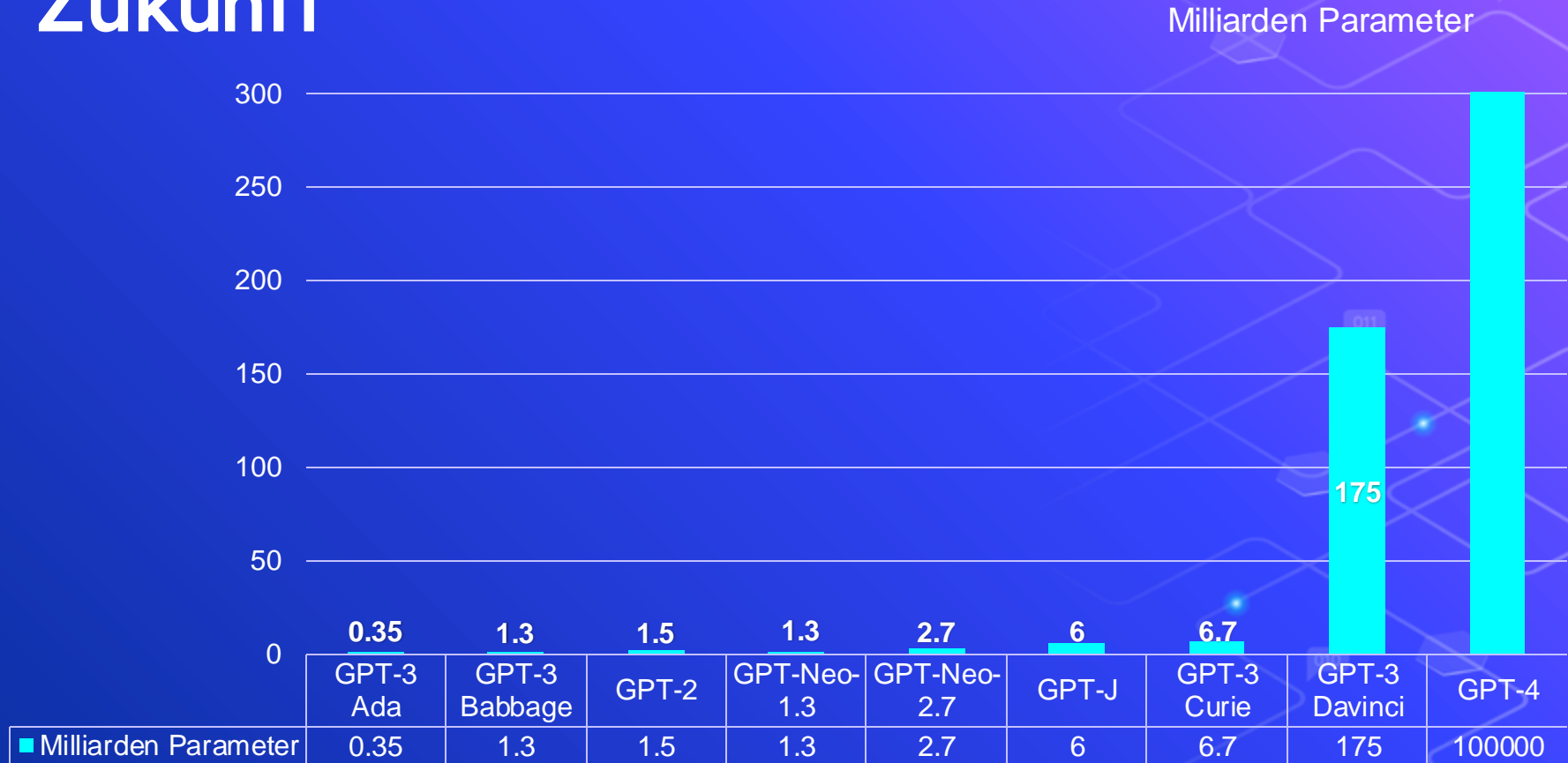


# Zukunft

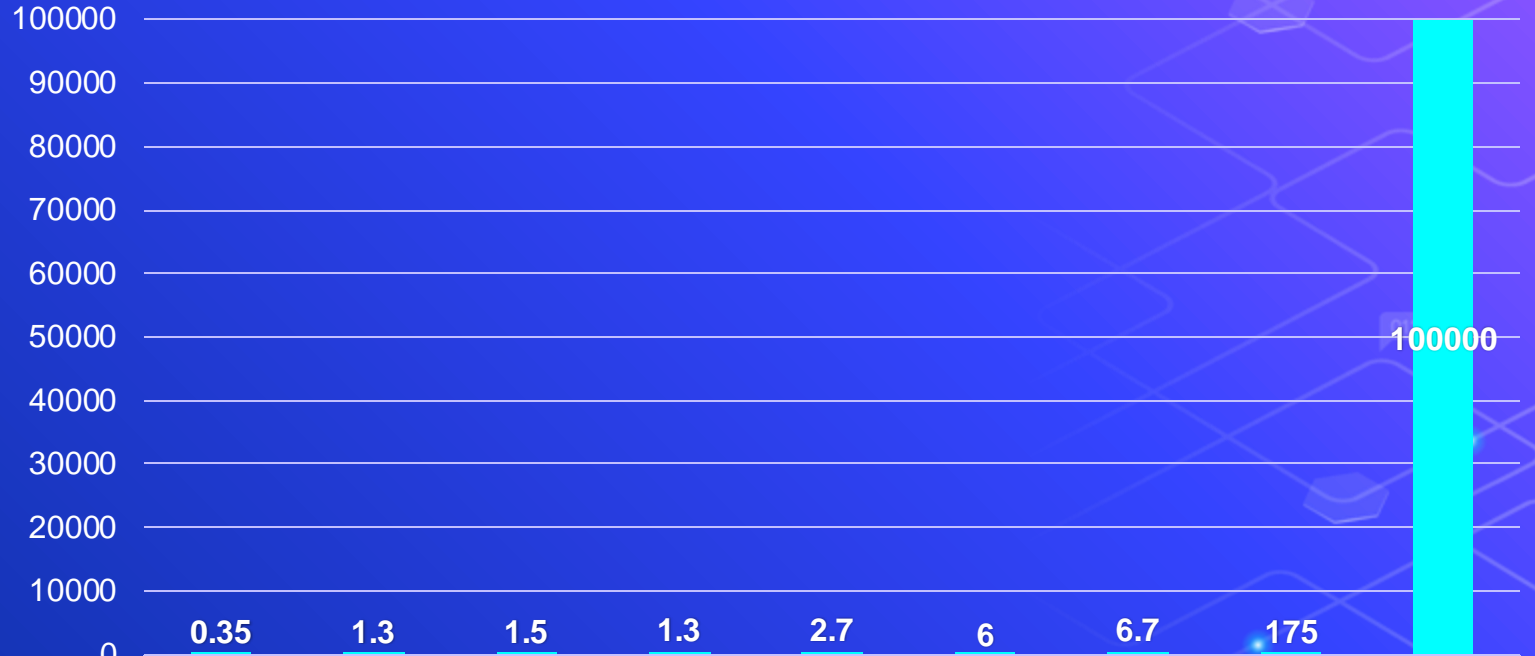


■ Milliarden Parameter

# Zukunft



# Zukunft



■ Milliarden Parameter

GPT-3 Ada	GPT-3 Babbage	GPT-2	GPT-Neo-1.3	GPT-Neo-2.7	GPT-J	GPT-3 Curie	GPT-3 Davinci	GPT-4
0.35	1.3	1.5	1.3	2.7	6	6.7	175	100000



# Cellarius