

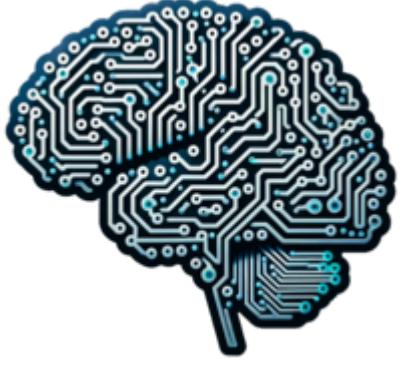
# Einführung in KI

Geschichte - Prinzipien -  
Anwendungen - Zukunft

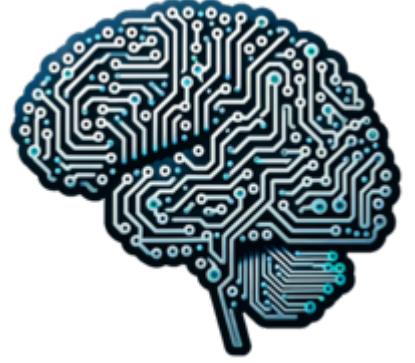
Dr. Frank Zimmer



# Fragen



- Wer hat schon Erfahrung mit künstlicher Intelligenz gesammelt?
- Wer ist im Bereich Informationstechnologie tätig?



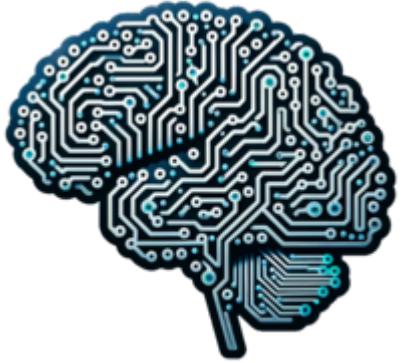
# Einführung in Künstliche Intelligenz

- Geschichte der KI
- Prinzipien der KI
- Trainieren eines Neuronen Netzes
- KI-Anwendungen
  - Kommerzieller Ansatz
  - Open Source Ansatz
- Zukunft



# Geschichte der KI

**Dr. Frank Zimmer**

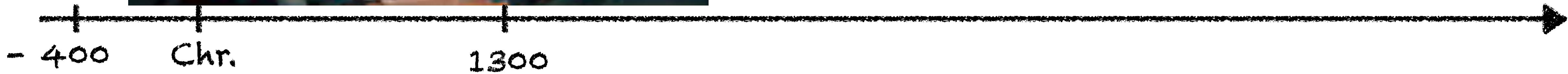
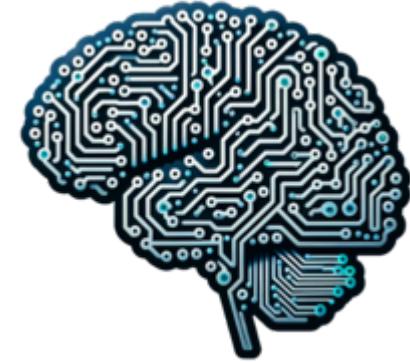


## Talos von Kreta

- ca. 400 v. Chr.
- riesige automatisierte Bronze Figur
- Schutz Kretas vor Piraten und Eindringlingen



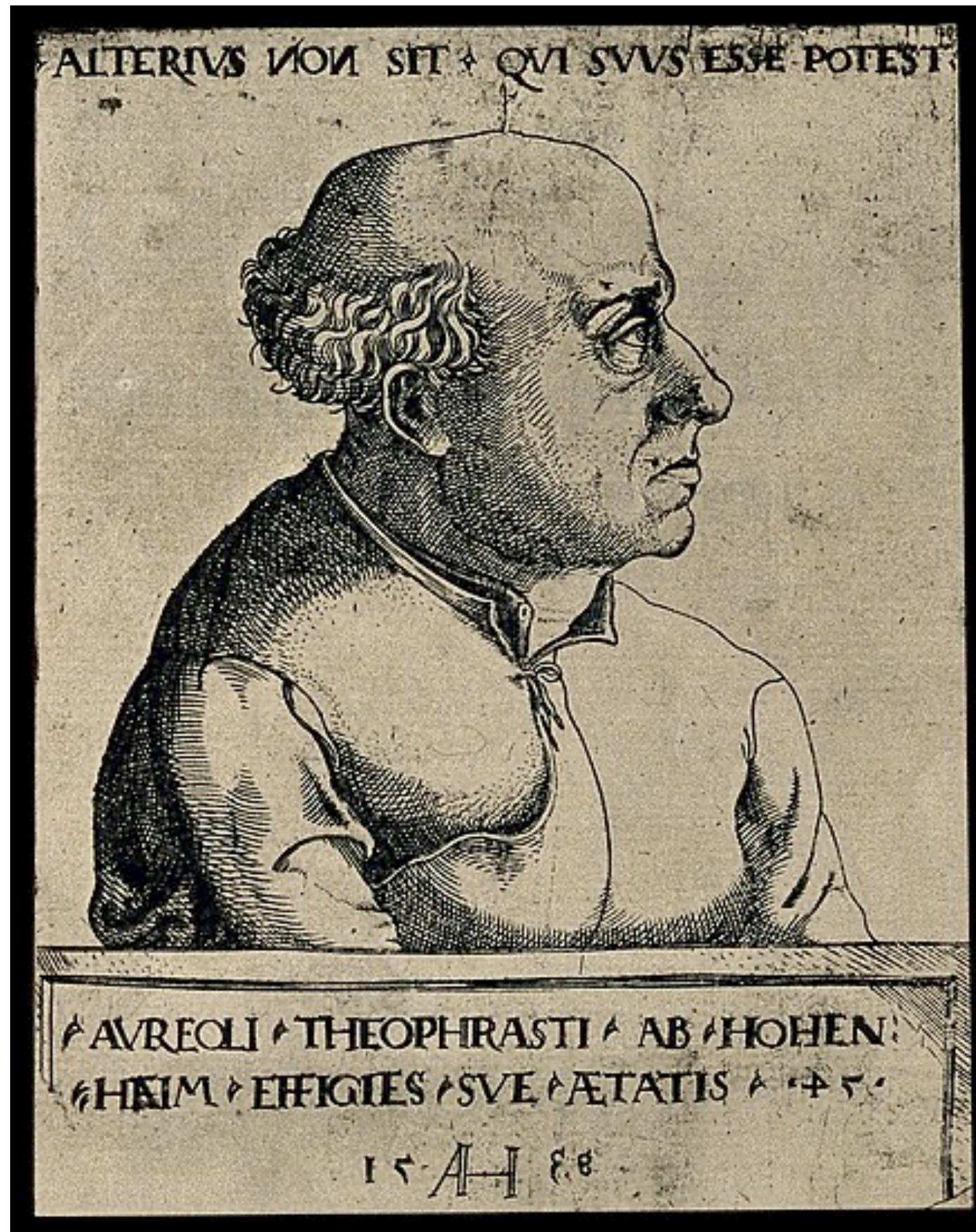
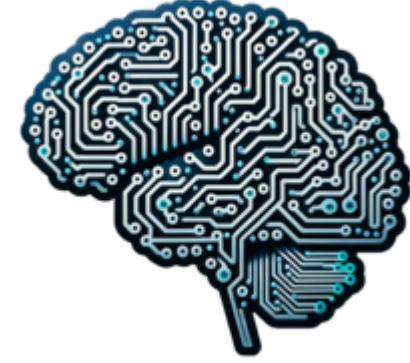
# (Vor-) Geschichte der KI



## Golem

- früheste Erwähnungen ca. 1300 n. Chr. durch Eleazar ben Judah von Worms
- Menschenähnliche Figur aus Ton

# (Vor-) Geschichte der KI



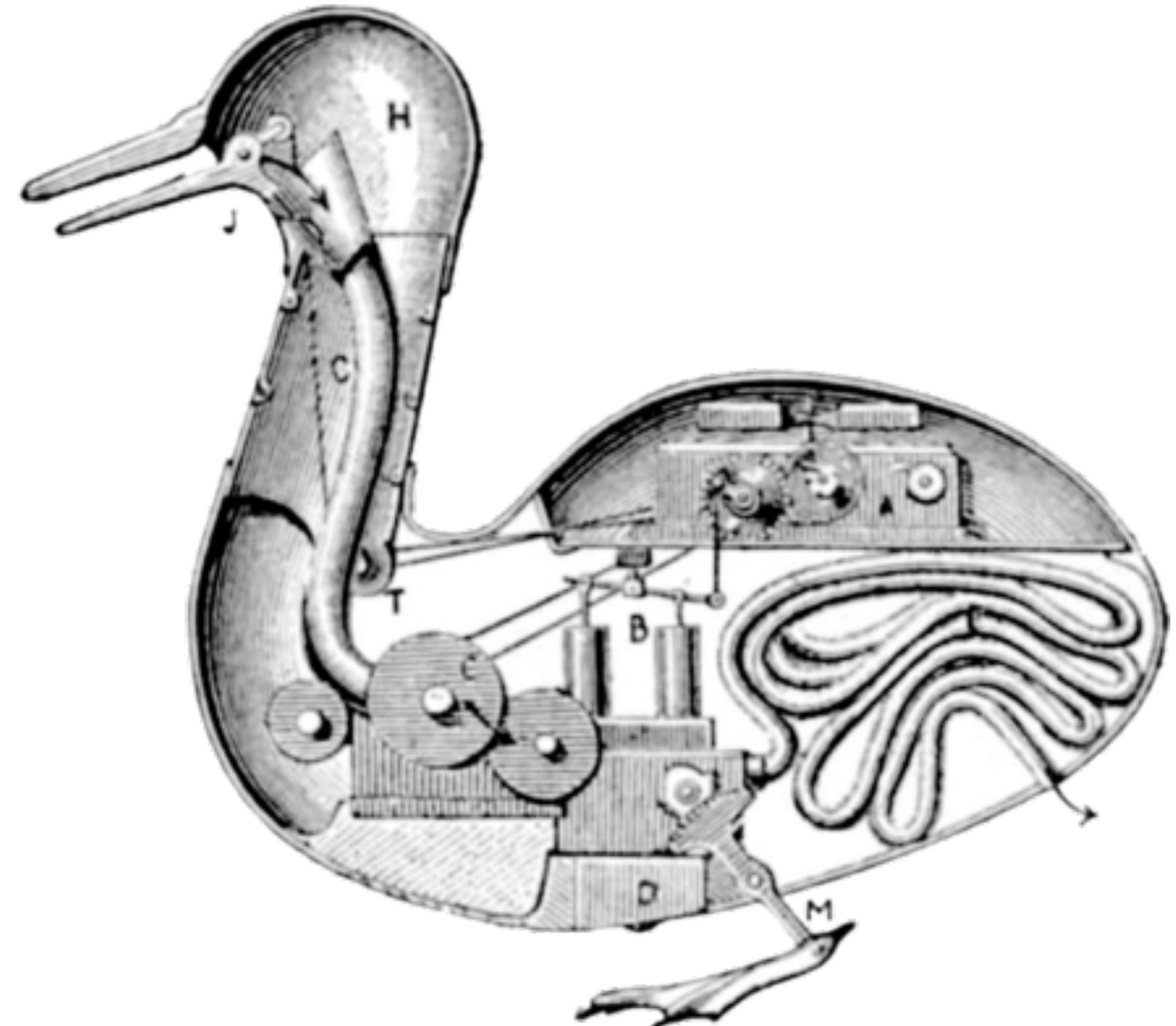
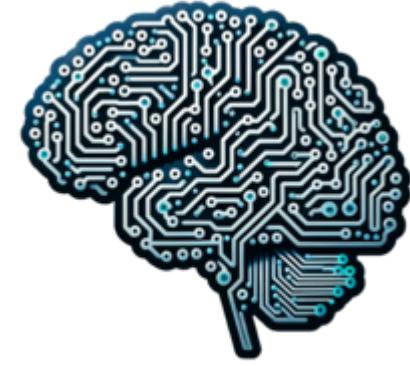
- 400      Chr.      1300      1500



## Paracelsus

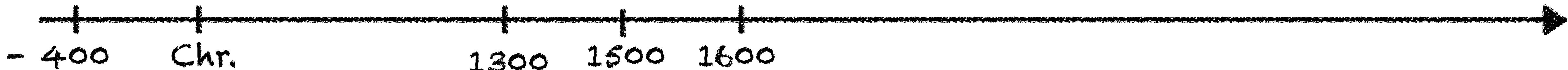
- ca. 1500 n. Chr.
- erzeugen eines künstlichen Menschen durch Spermien im Pferdemist
- Geschöpf entsteht nach ca. 40 Tagen

# (Vor-) Geschichte der KI

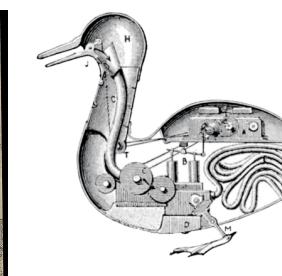
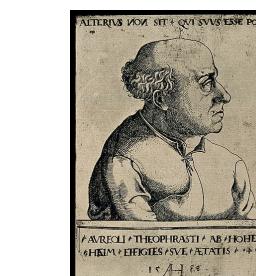
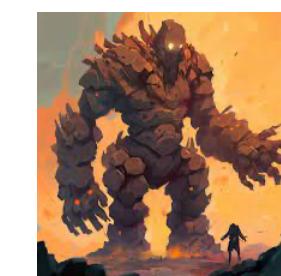
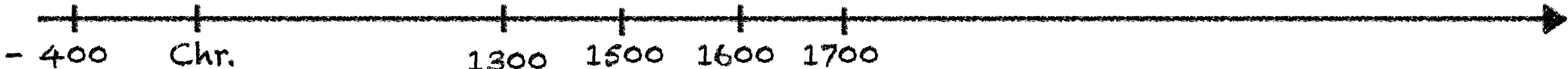
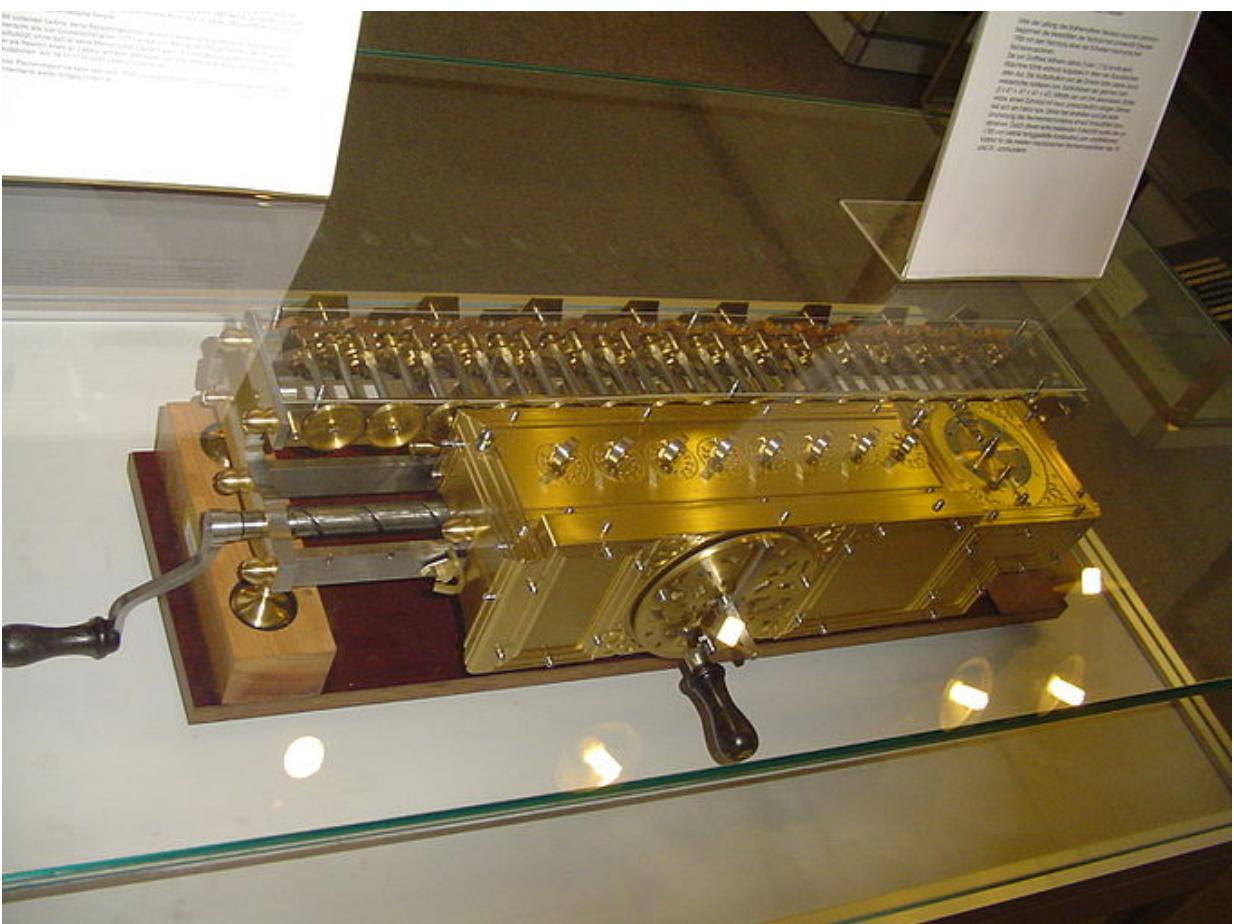
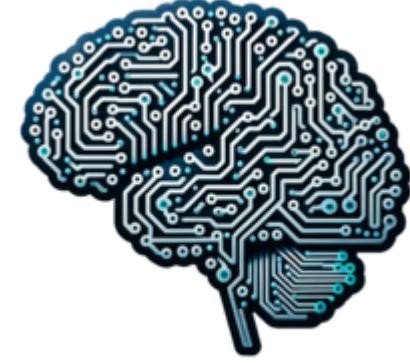


## Automtons

- ca. 1700 n. Chr., aber auch schon vorher
- Nachbildungen von Menschen und Tieren, die sich durch Mechanik bewegen
- für den Betrachter sieht es zunächst wie intelligent aus

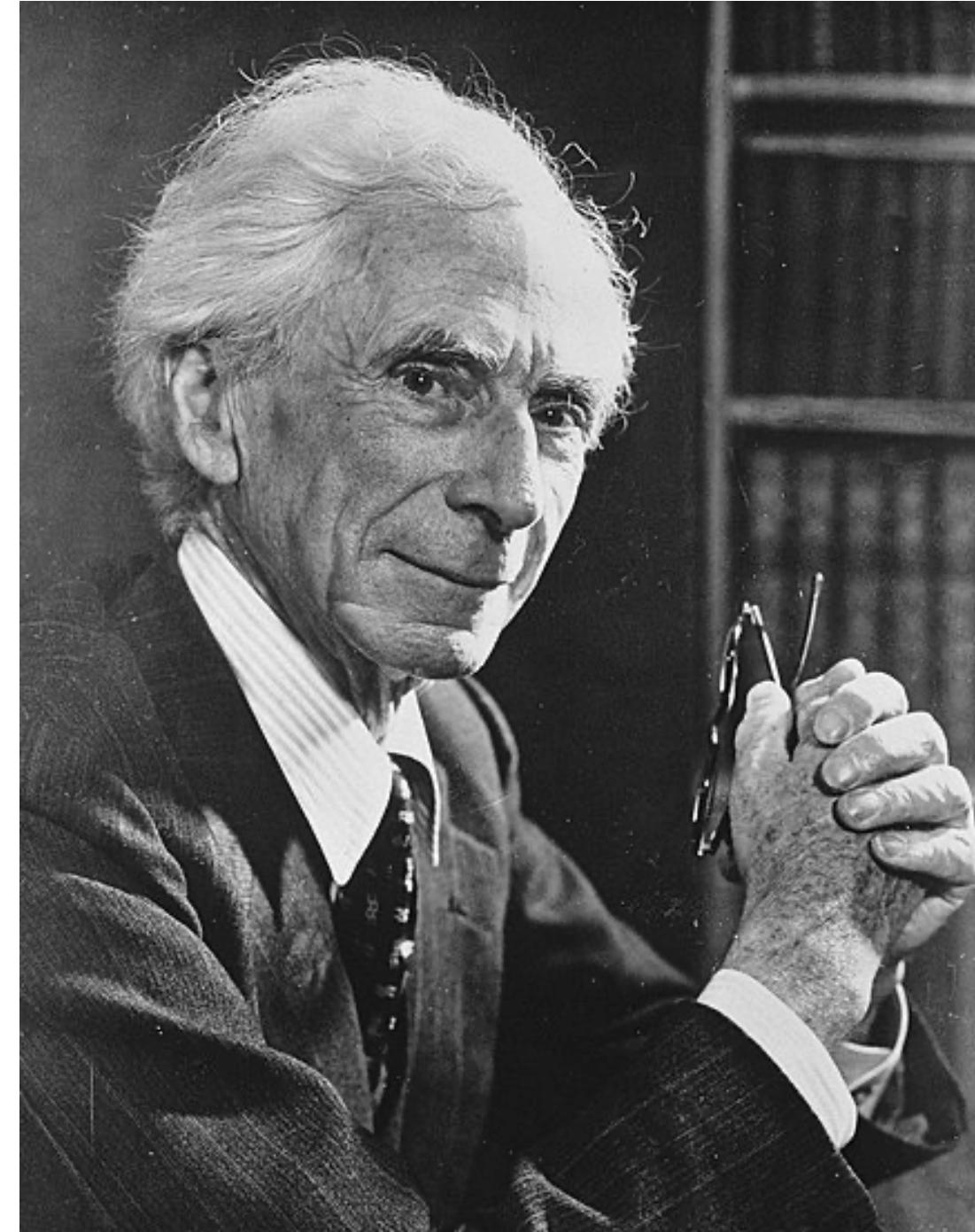
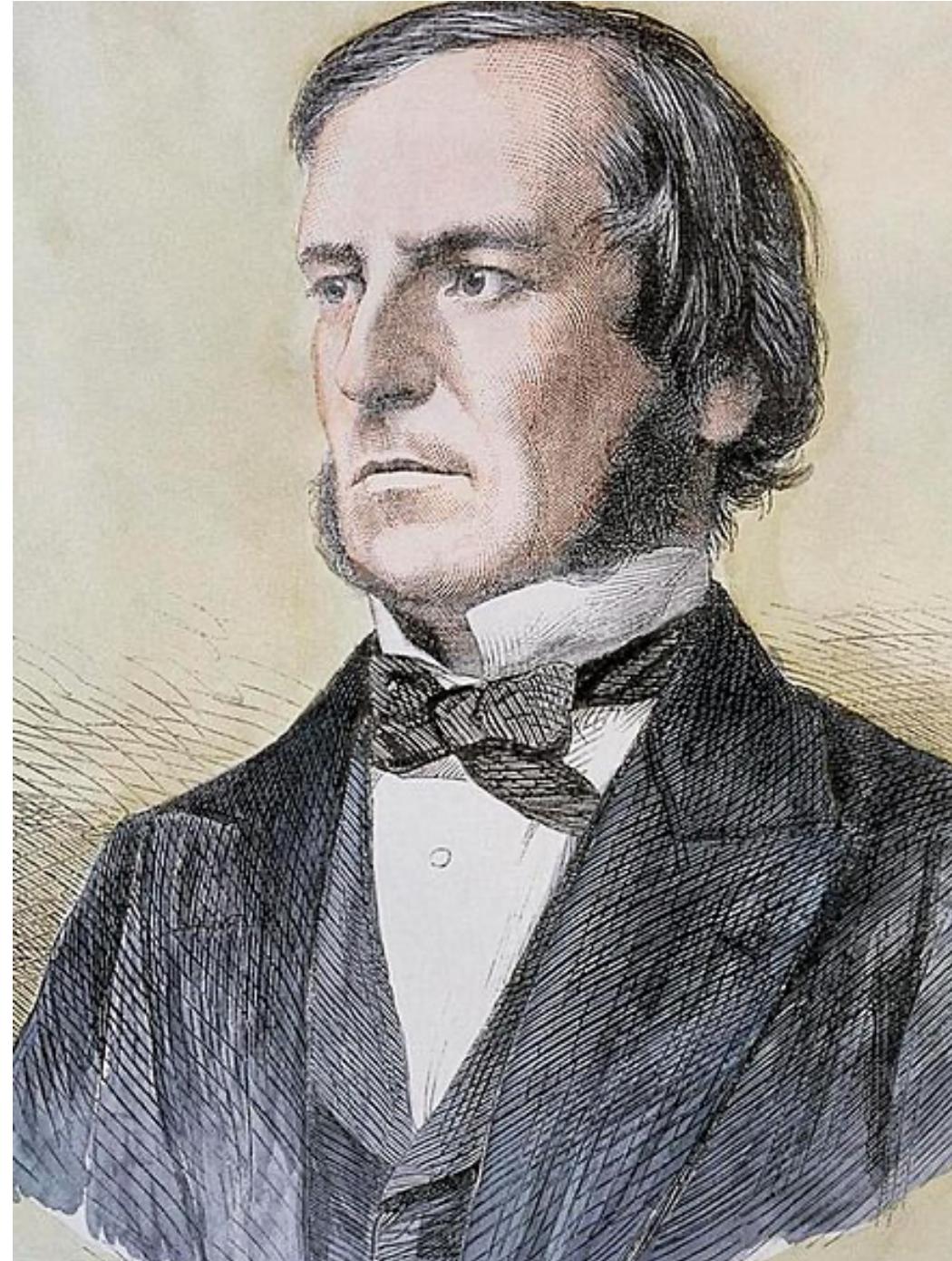
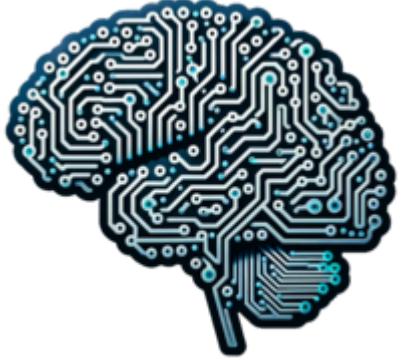


# (Vor-) Geschichte der KI



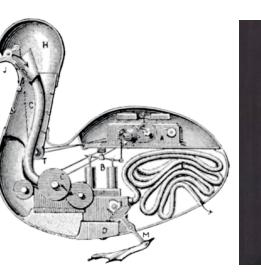
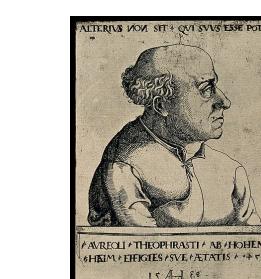
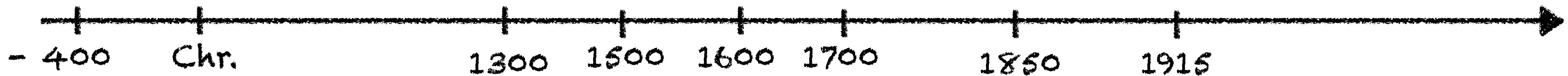
## Formale Logik 1

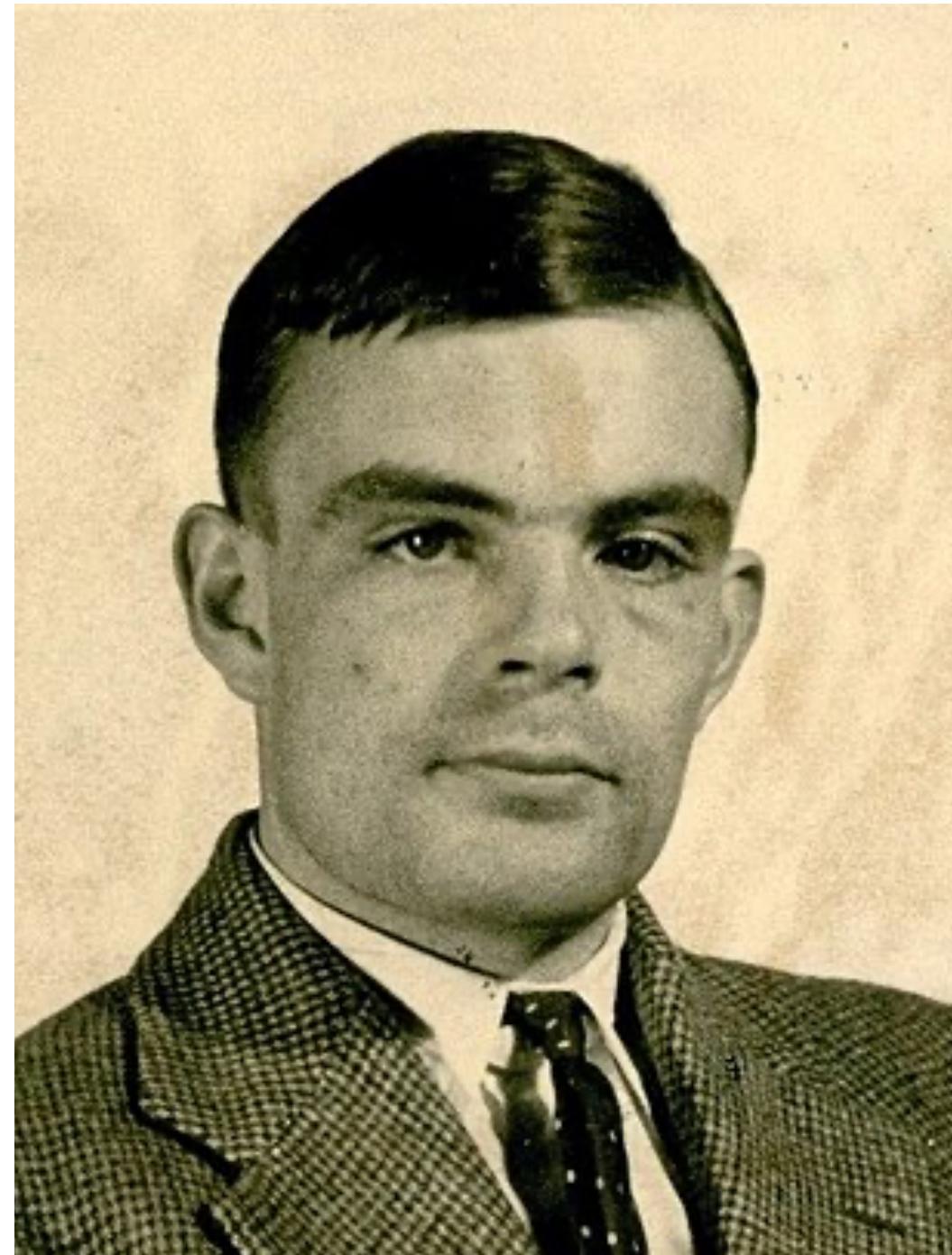
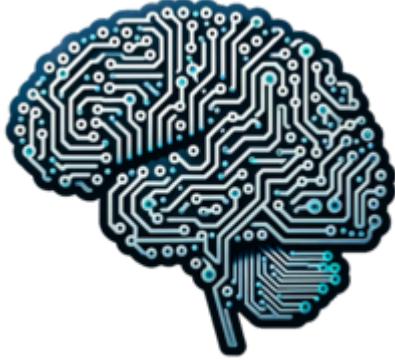
- Leibniz, Hobbes und Descartes arbeiten an rationalem, mechanisiertem Denken
- Argumentieren soll wie Rechnen werden
- Formalisierung der mathematischen Sprache



## Formelle Logik 2

- Boole erfindet Boolesche Algebra in *The laws of thought*; wichtig Entwicklung von Computern
- Frege formalisiert mathematische Sprache weiter im Sinne von Leibniz
- Russel und Whitehead veröffentlichen Principia Mathematica



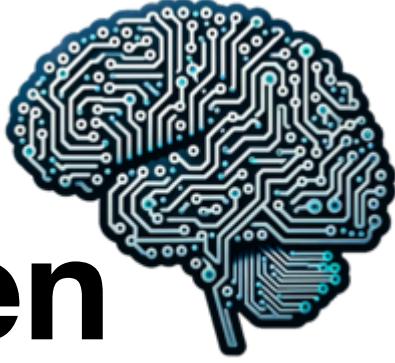


## Erfindung Computer

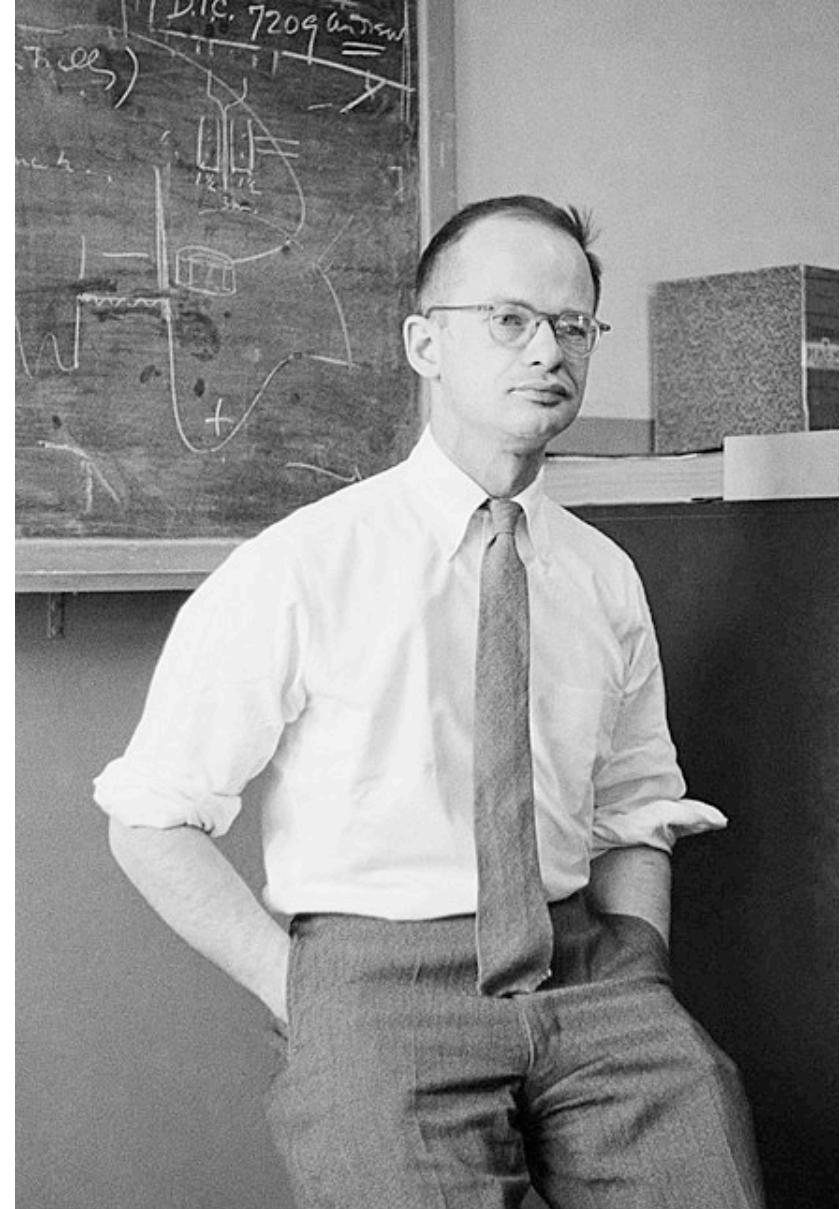
- Turing entwickelt die Turing machine als Konzept für einen Computer
- John von Entwickelt bis heute gültige Architektur von Computer
- 1945 wird ENIC von von Neumann und anderen entwickelt



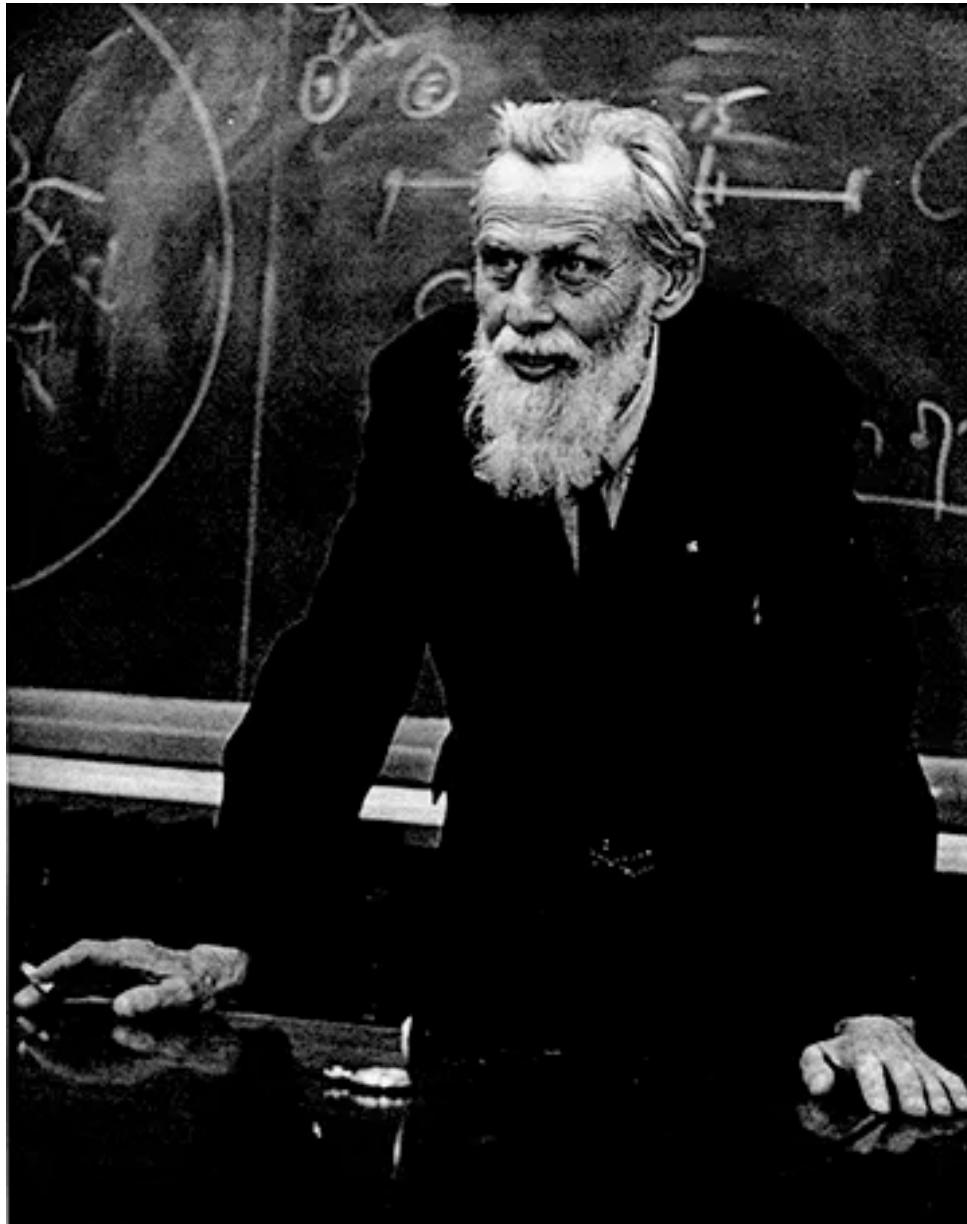
# (Moderne) Geschichte der KI



## Konzept künstlicher Neuronen



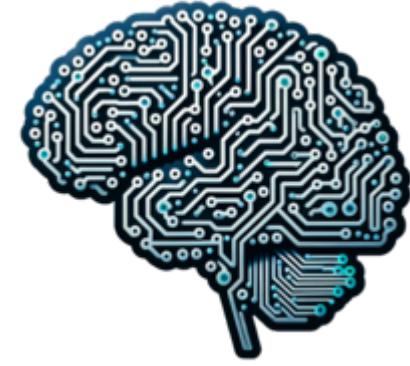
Pitts



McCulloch

- 1943: Pitts und McCulloch veröffentlichen Arbeit zu künstlichen Neuronen
- Neuron ist die logische Basiseinheit eines Gehirns laut Ihrer Arbeit
- Ein künstliches Neuron ist eine Funktion, welche über eine gewichtete Summe von Eingangswerten ein Ausgangswert bestimmt.
- Einführung des Begriffs Neuronales Netz

1943



Newell

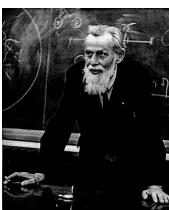


Simon

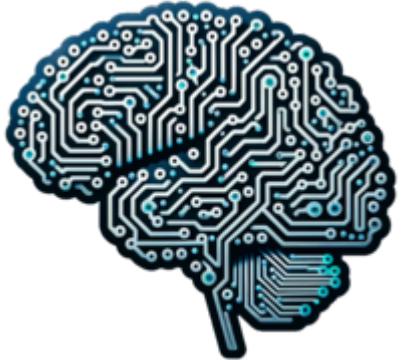
- 1955: Newell und Simon schreiben **erstes KI Programm**
- Programm wird zum automatisierte Schlussfolgern genutzt
- Programm bewies 38 Sätze aus Principia Mathematica und fand auch neue Beweise
- Ausgangspunkt für Lisp und Prolog



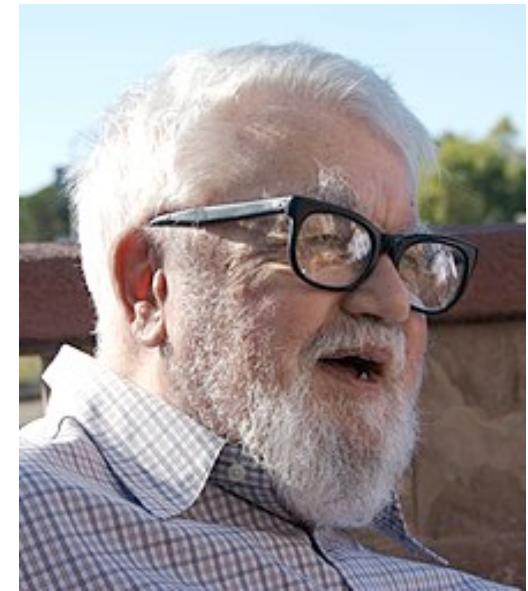
1943 1955



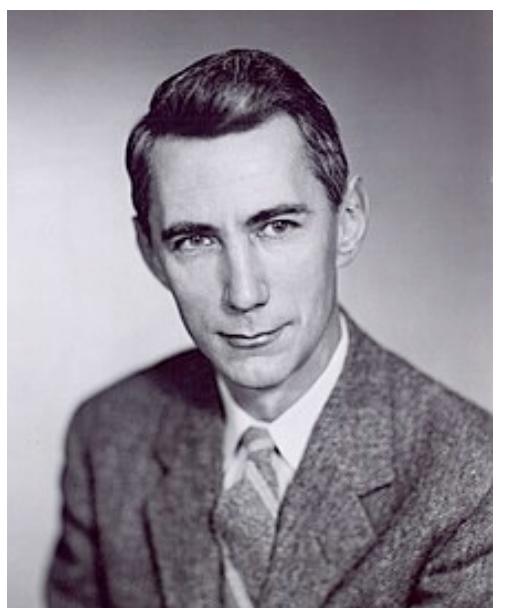
# (Moderne) Geschichte der KI



Minsky



McCarthy



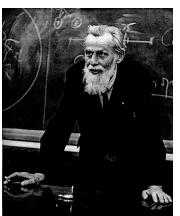
Shannon



Rochester



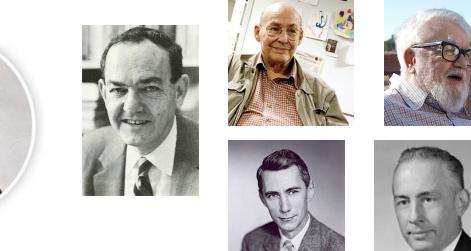
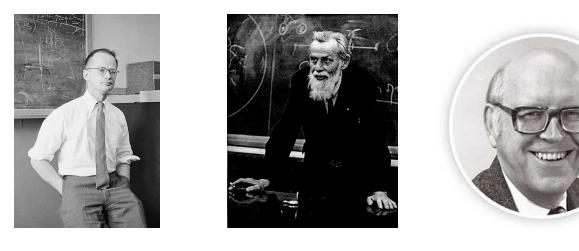
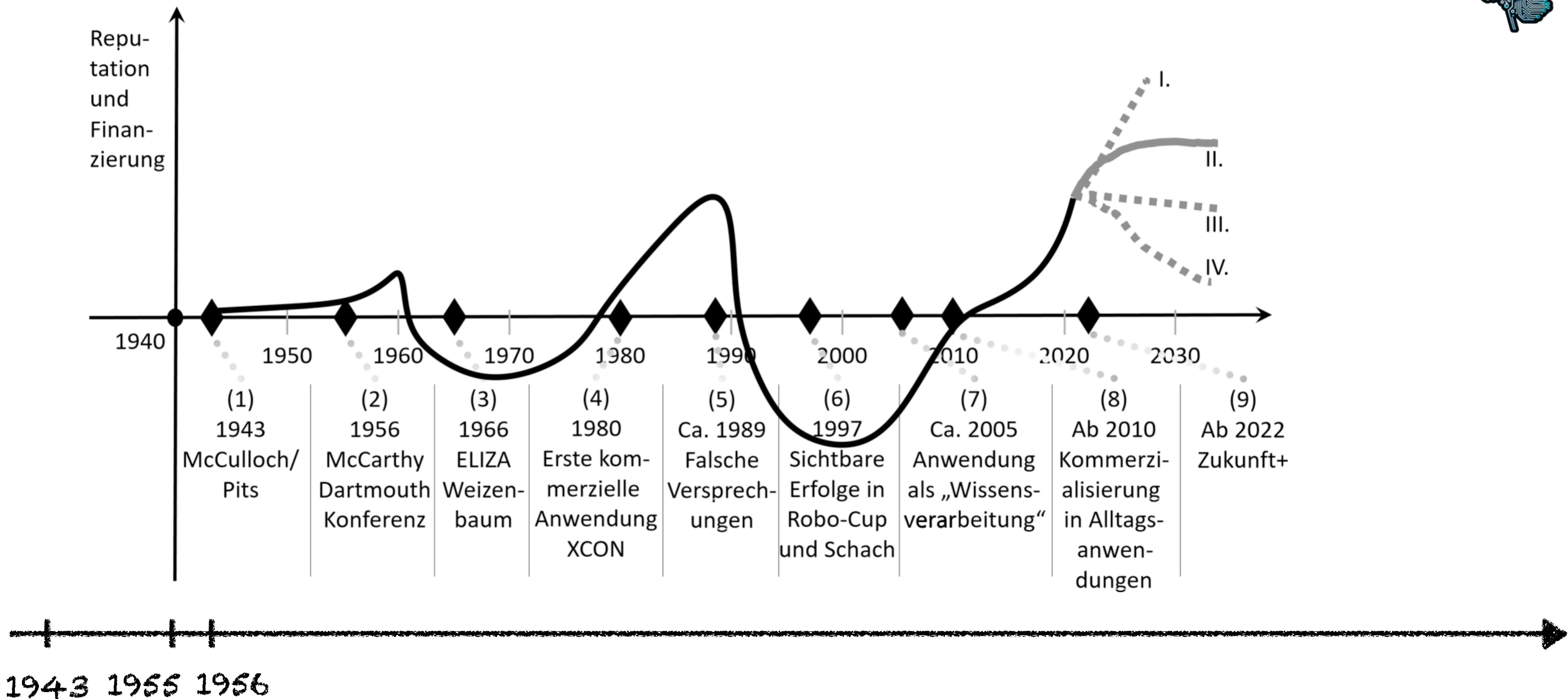
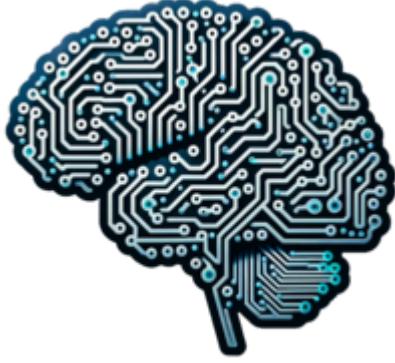
1943 1955 1956

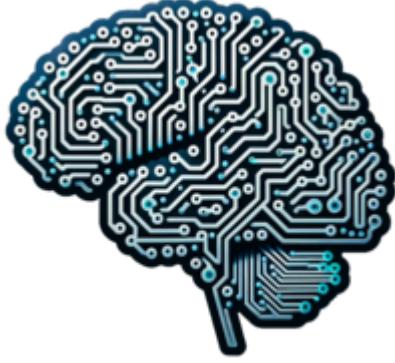


## Dartmouth Workshop

- 1956: Minsky (AI MIT Lab), McCarthy (Lisp), Shannon (Entropie), Rochester (Assembler) organisieren Treffen zum Thema: Möglichkeiten einer Maschine zum Simulieren eines menschlichen Gehirns
- Der **Begriff KI/AI** wurde im Rahmen der Konferenz eingeführt.
- Stellt die **Geburtsstunde** der modernen Forschung zur KI da.

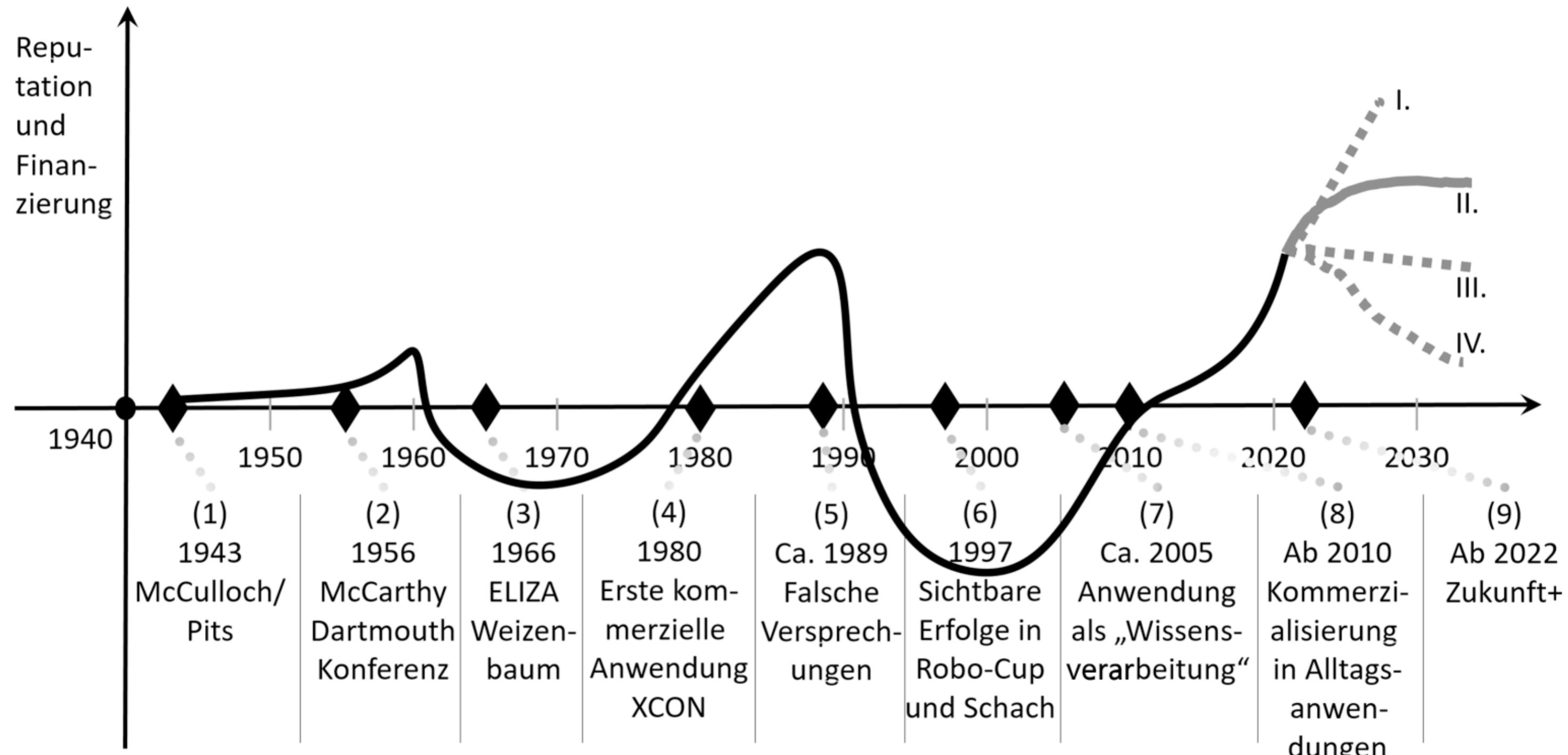
# (Moderne) Geschichte der KI





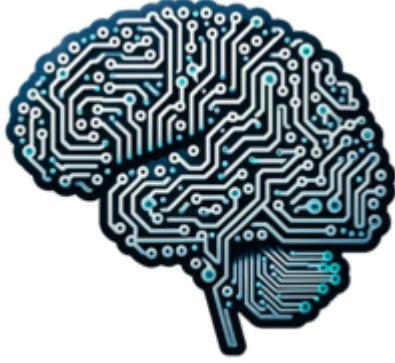
## Überoptimistisch

- 1958: Simon, Newell: „within ten years digital computer will be the world's chess champion and will discover and prove new mathematical theorems“
- 1965: Simon: „machines will be capable, within twenty years, of doing work a man can do“
- 1970: Minsky: „In from three to eight years we will have a machine with the general intelligence of an average human being.“



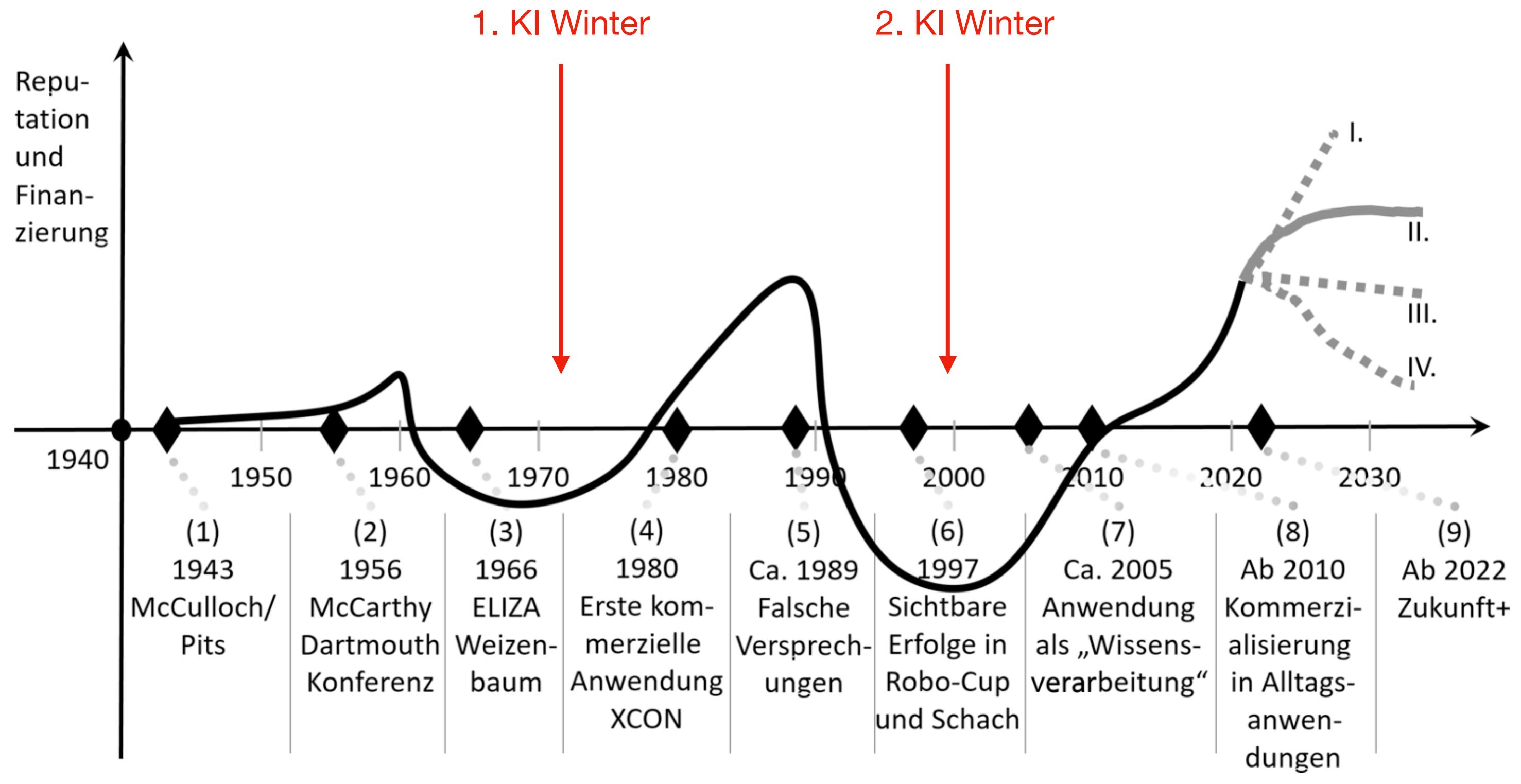
1943 1955 1956





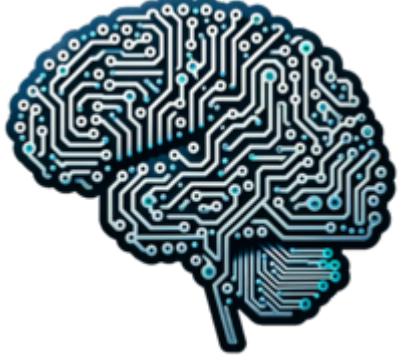
## Überoptimistisch

- 1958: Simon, Newell: „within ten years digital computer will be the world's chess champion and will discover and prove new mathematical theorems“
- 1965: Simon: „machines will be capable, within twenty years, of doing work a man can do“
- 1970: Minsky: „In from three to eight years we will have a machine with the general intelligence of an average human being.“



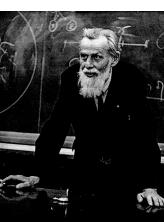
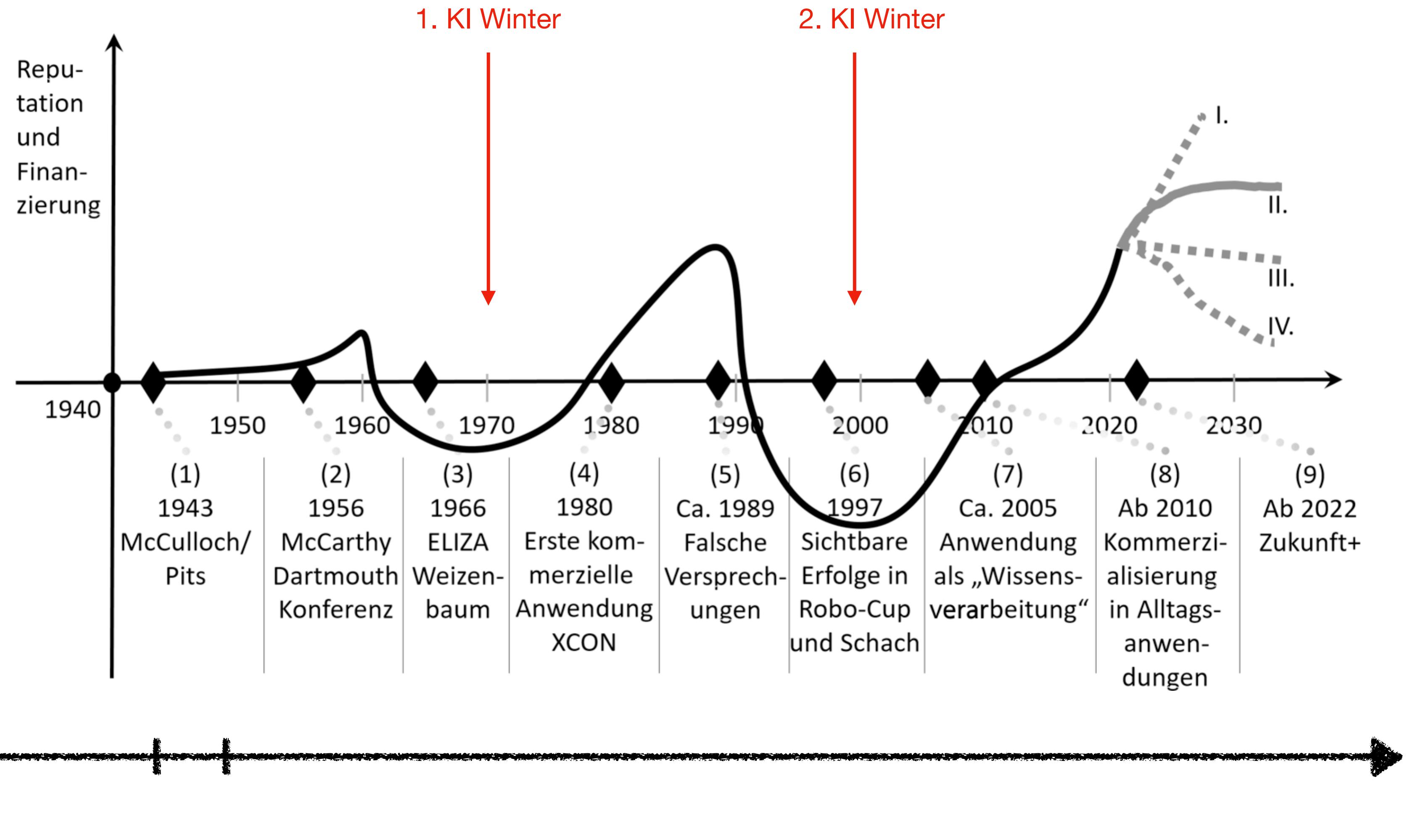
1943 1955 1956

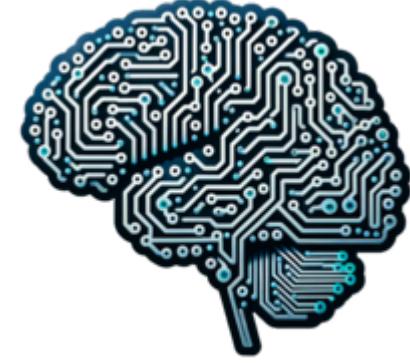




## Probleme

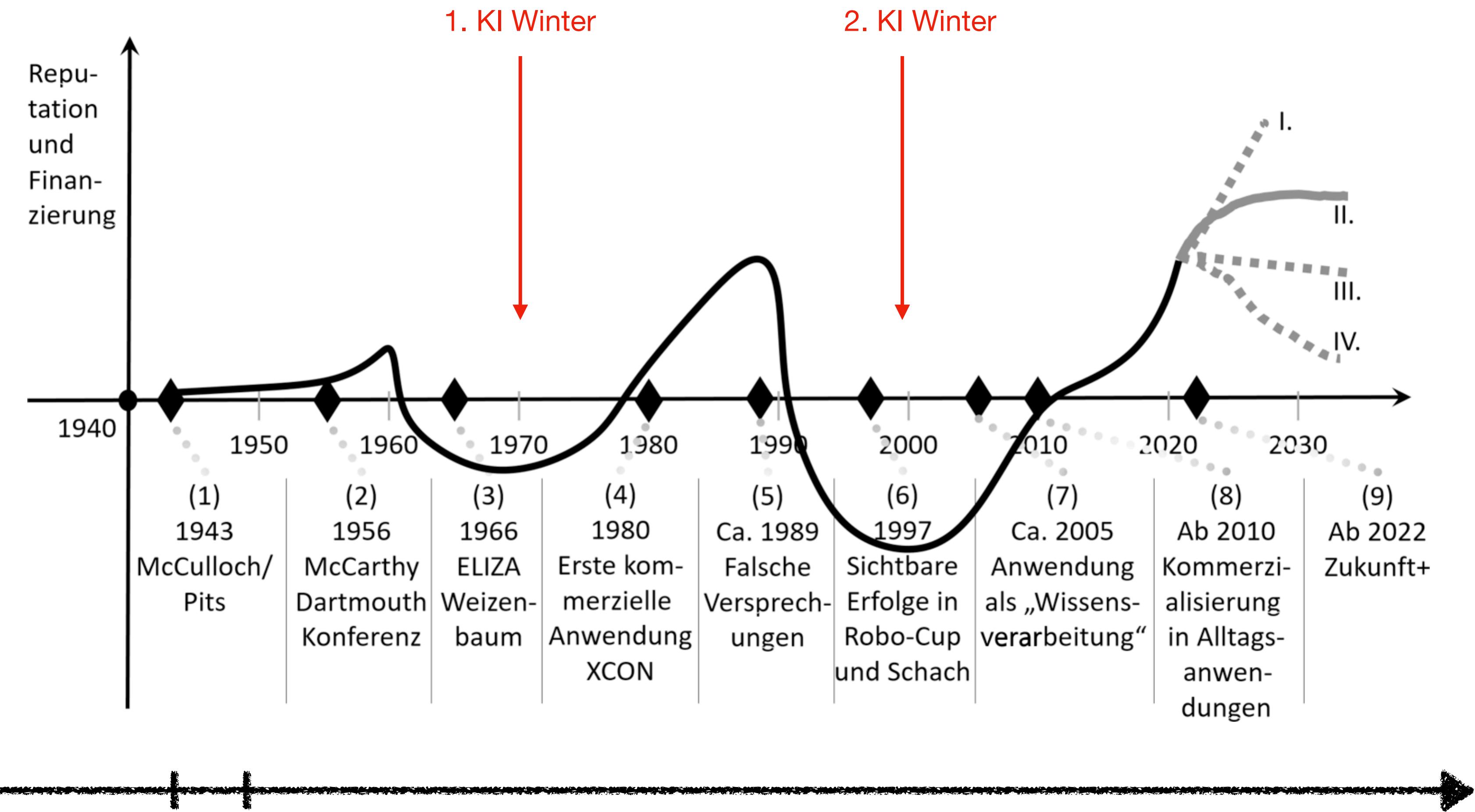
- **Begrenzte Compute Power:** nicht genügend Speicher und Prozessgeschwindigkeit
- Unlösbarkeit und **kombinatorische Explosion:** für einige Ansätze wurde erkannt, dass sie nicht mit digitalen Computern lösbar sind
- Gesunder Menschenverstand und vernünftige Argumentation brauchen große Datenmenge
- **Moravecches Paradox:** für Menschen einfache Sachen (Gehen, Sprechen, etc) sind für Computer schwieriger zu „erlernen“ als Lösung mathematischer Probleme



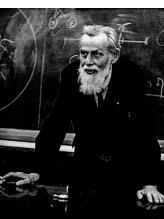


## Probleme

- **Begrenzte Compute Power:** nicht genügend Speicher und Prozessgeschwindigkeit
- Unlösbarkeit und **kombinatorische Explosion:** für einige Ansätze wurde erkannt, dass sie nicht mit digitalen Computern lösbar sind
- Gesunder Menschenverstand und vernünftige Argumentation brauchen große Datenmenge
- **Moravecches Paradox:** für Menschen einfache Sachen (Gehen, Sprechen, etc) sind für Computer schwieriger zu „erlernen“ als Lösung mathematischer Probleme

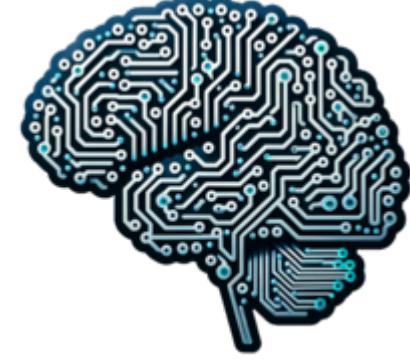


1943 1955 1956



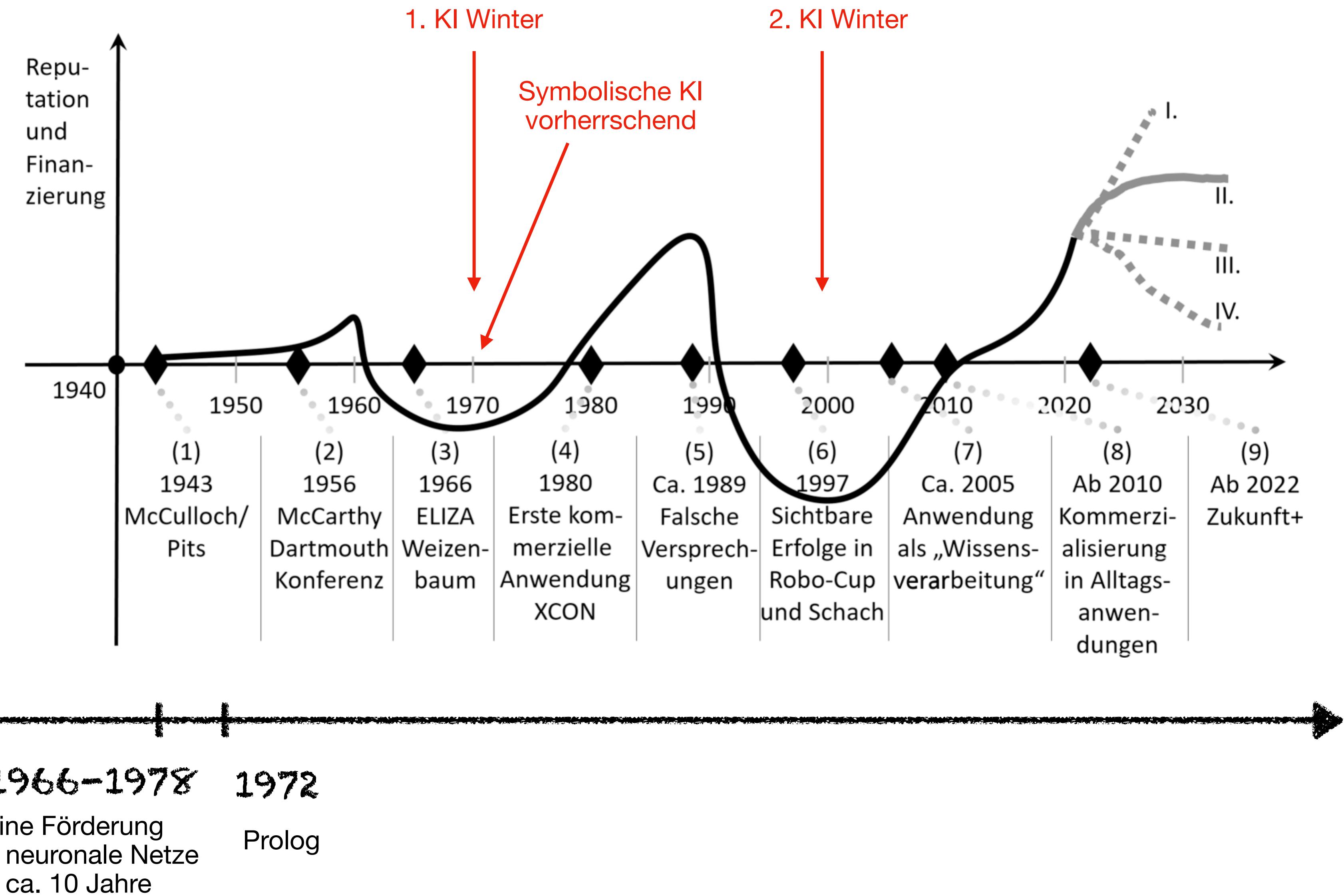
1966-1978

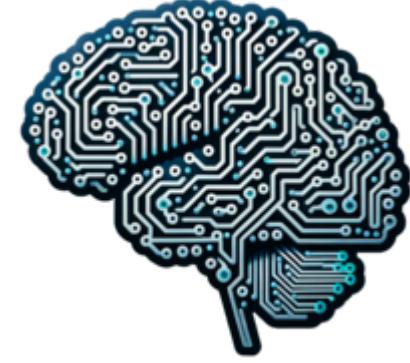
Keine Förderung  
für neuronale Netze  
für ca. 10 Jahre



## Probleme

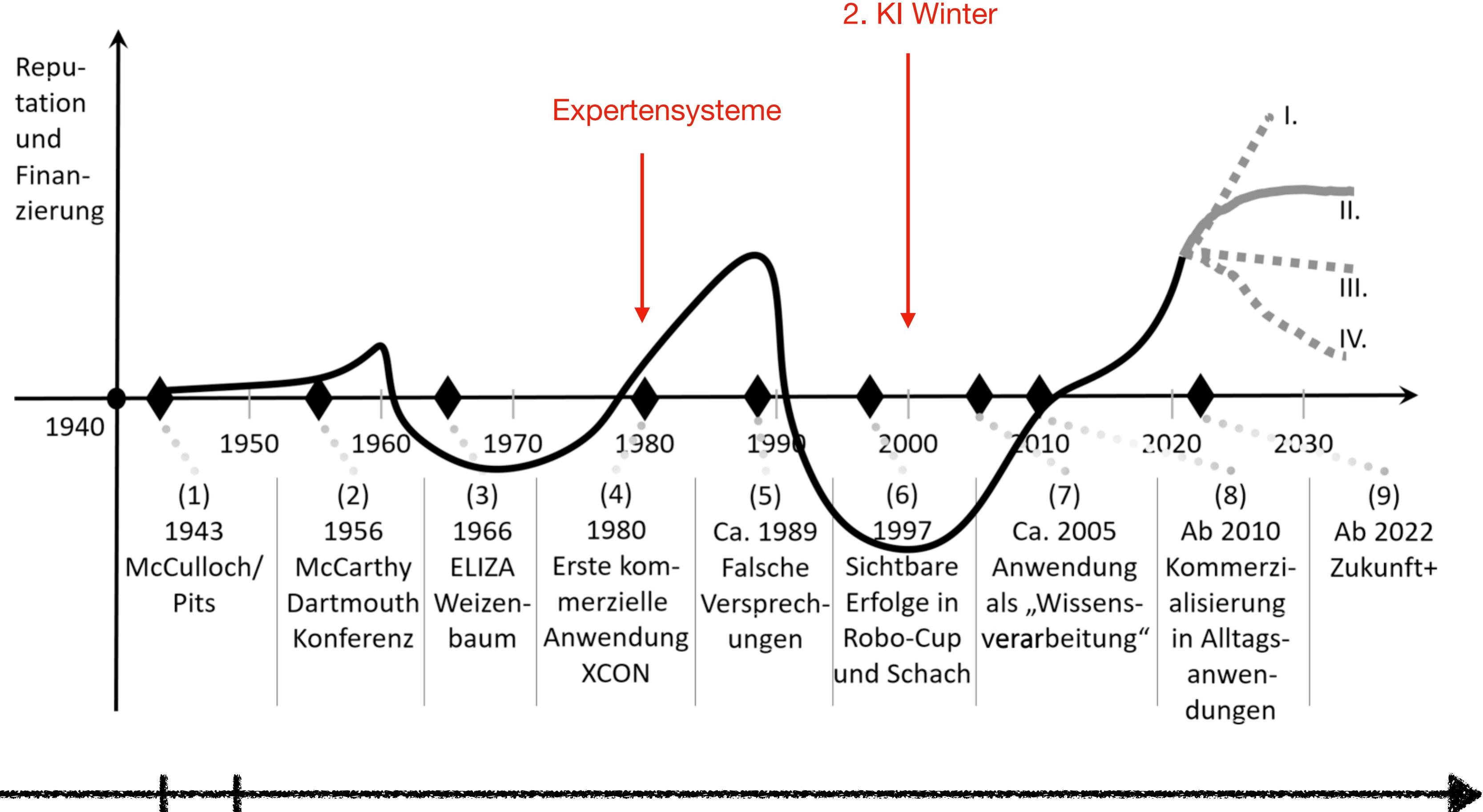
- **Begrenzte Compute Power:** nicht genügend Speicher und Prozessgeschwindigkeit
- Unlösbarkeit und **kombinatorische Explosion:** für einige Ansätze wurde erkannt, dass sie nicht mit digitalen Computern lösbar sind
- Gesunder Menschenverstand und vernünftige Argumentation brauchen große Datenmenge
- **Moravecches Paradox:** für Menschen einfache Sachen (Gehen, Sprechen, etc) sind für Computer schwieriger zu „erlernen“ als Lösung mathematischer Probleme



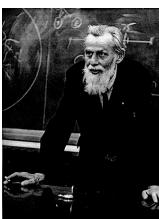


## Verschiedene Ansätze

- **Experten Systeme:** Domänen Experten bauen Domänen-Kenntnisse und Regeln auf; System nutzt dies um Problem zu lösen und Fragen zu beantworten
- Experten Systeme sind in vielen Fällen zu teuer, um diese zu pflegen
- **Spezialisierte Computer-Hardware** für KI Projekte wären ein Ansatz, aber durch Einführung des PCs zu teuer
- Erwartung werden **wieder nicht** erfüllt



1943 1955 1956

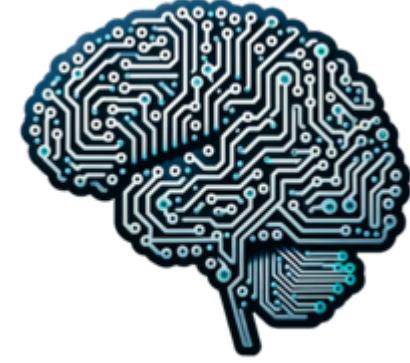


1966-1978 1972

Keine Förderung  
für neuronale Netze  
für ca. 10 Jahre

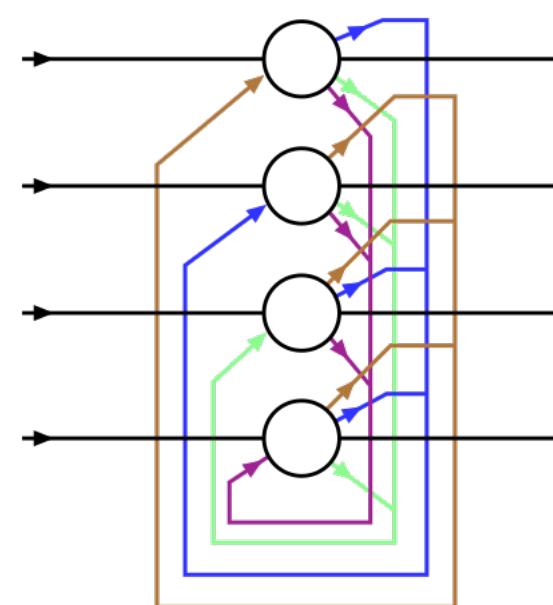
Prolog

# (Moderne) Geschichte der KI



## Hopfield

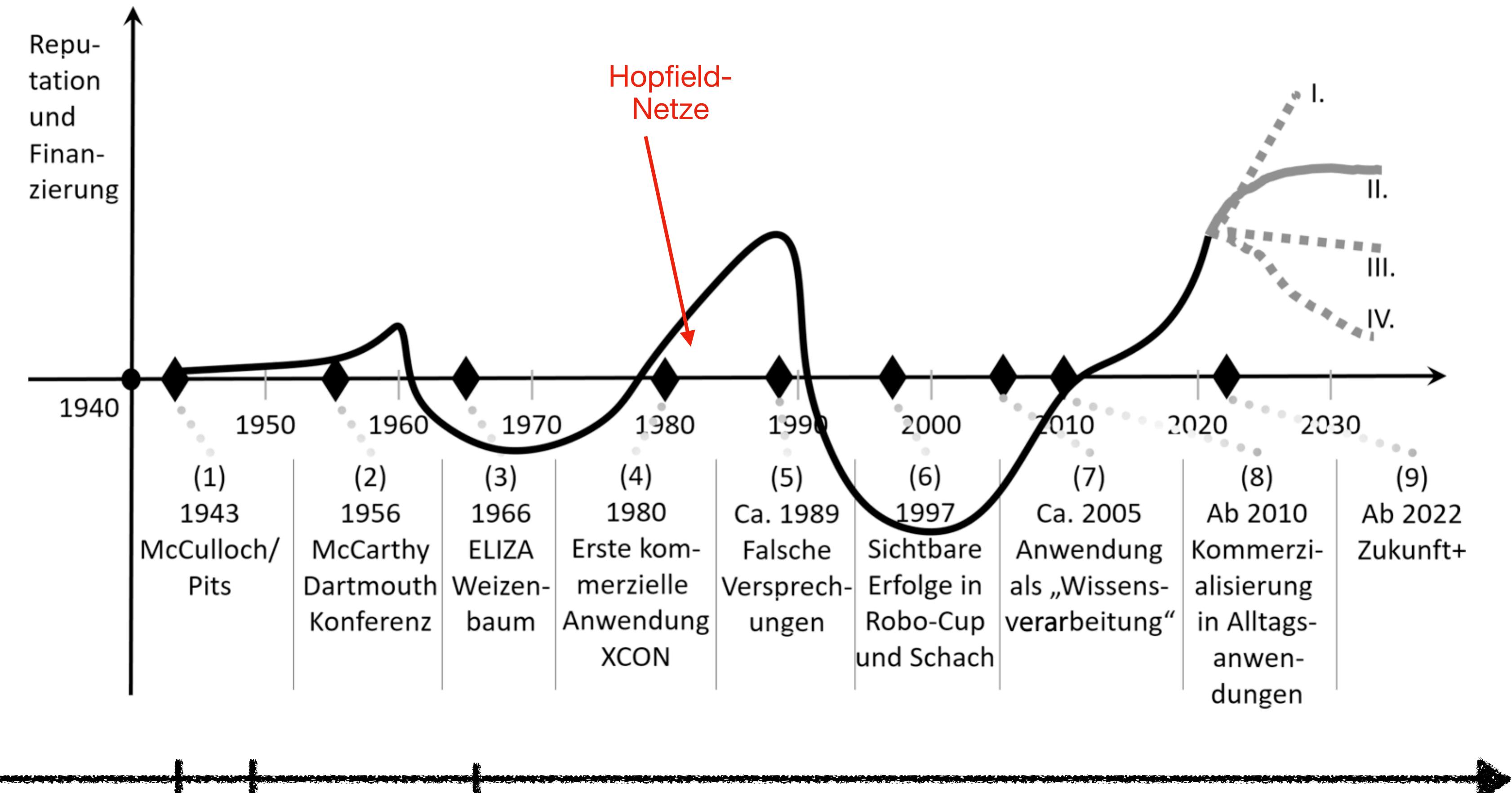
### Hopfield-Netze



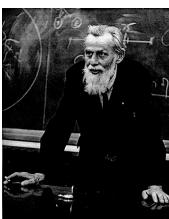
Netze mit Rückkopplung

1. Hopfield konnte zeigen, dass Netze Information lernen und prozessieren konnten.

2. Lern Prozess konvergiert nach endlich vielen Schritten und führt **nicht** zu chaotischer Dynamik



1943 1955 1956



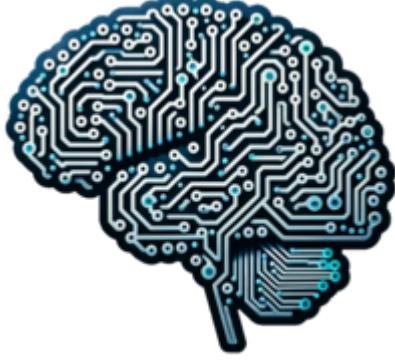
1966-1978 1972 1982

Keine Förderung für neuronale Netze für ca. 10 Jahre

Prolog

Hopfield-Netze

# (Moderne) Geschichte der KI



Hinton

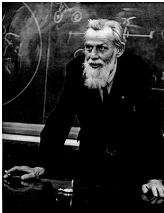
Godfather moderner KI

## Backpropagation

Demonstriert, dass das vorher, in einem anderen Zusammenhang entwickelte Verfahren, bei neuronalen Netzen erfolgreich angewandt werden kann.



1943 1955 1956



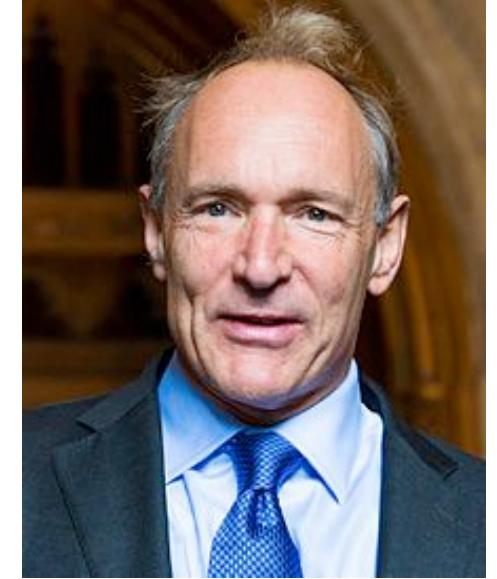
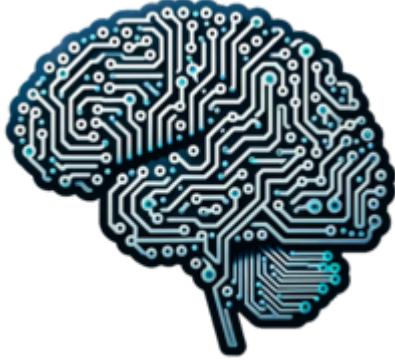
1966-1978 1972 1982

Keine Förderung  
für neuronale Netze  
für ca. 10 Jahre

Prolog



# (Moderne) Geschichte der KI

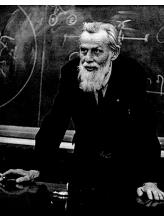


Tim Berners-Lee

Internet

Zugang zu vielen Daten

1943 1955 1956



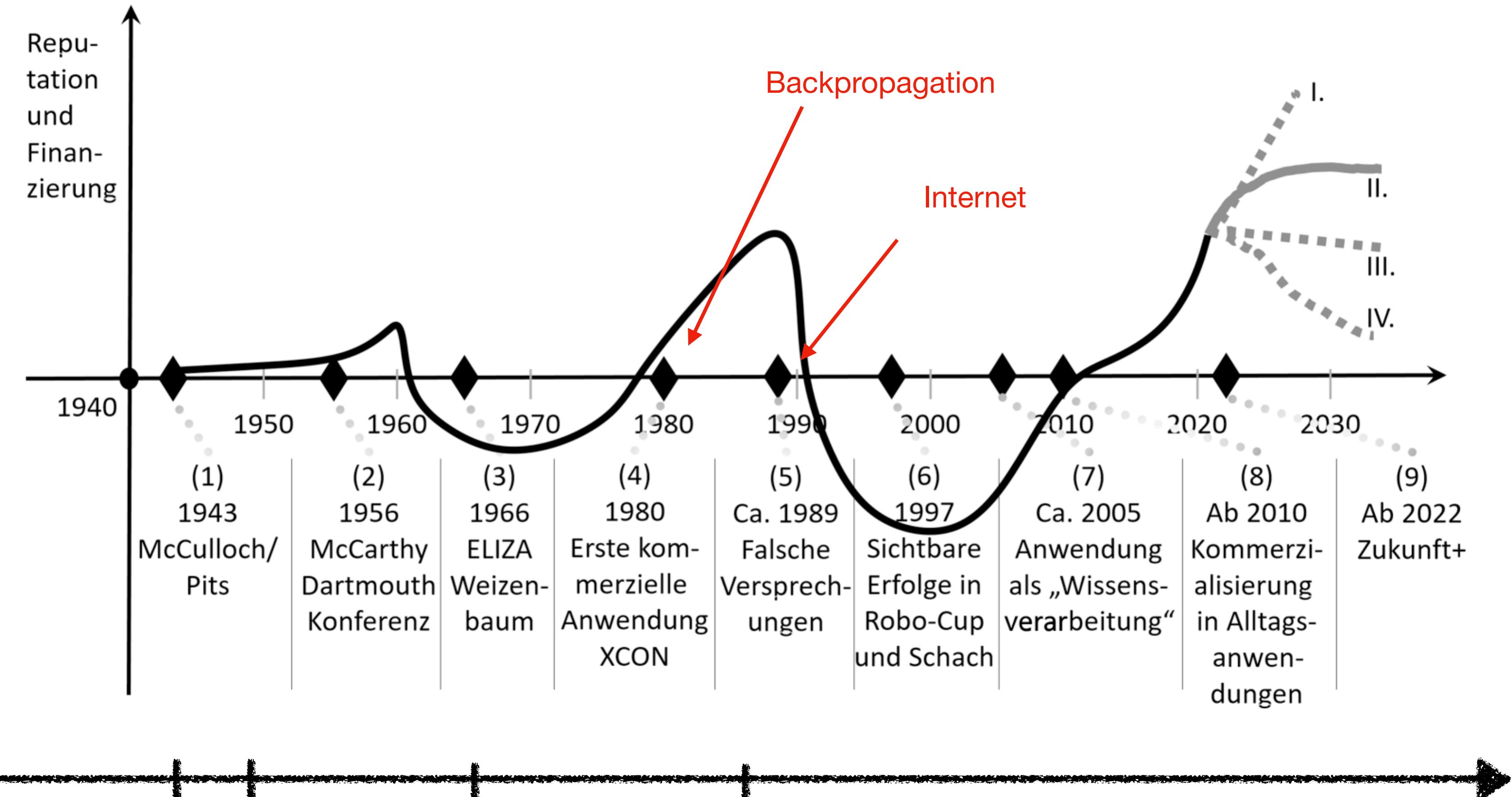
Keine Förderung  
für neuronale Netze  
für ca. 10 Jahre

Prolog

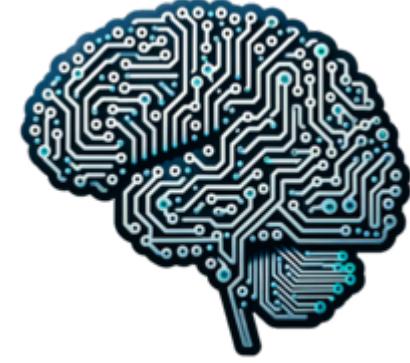


1990

Internet

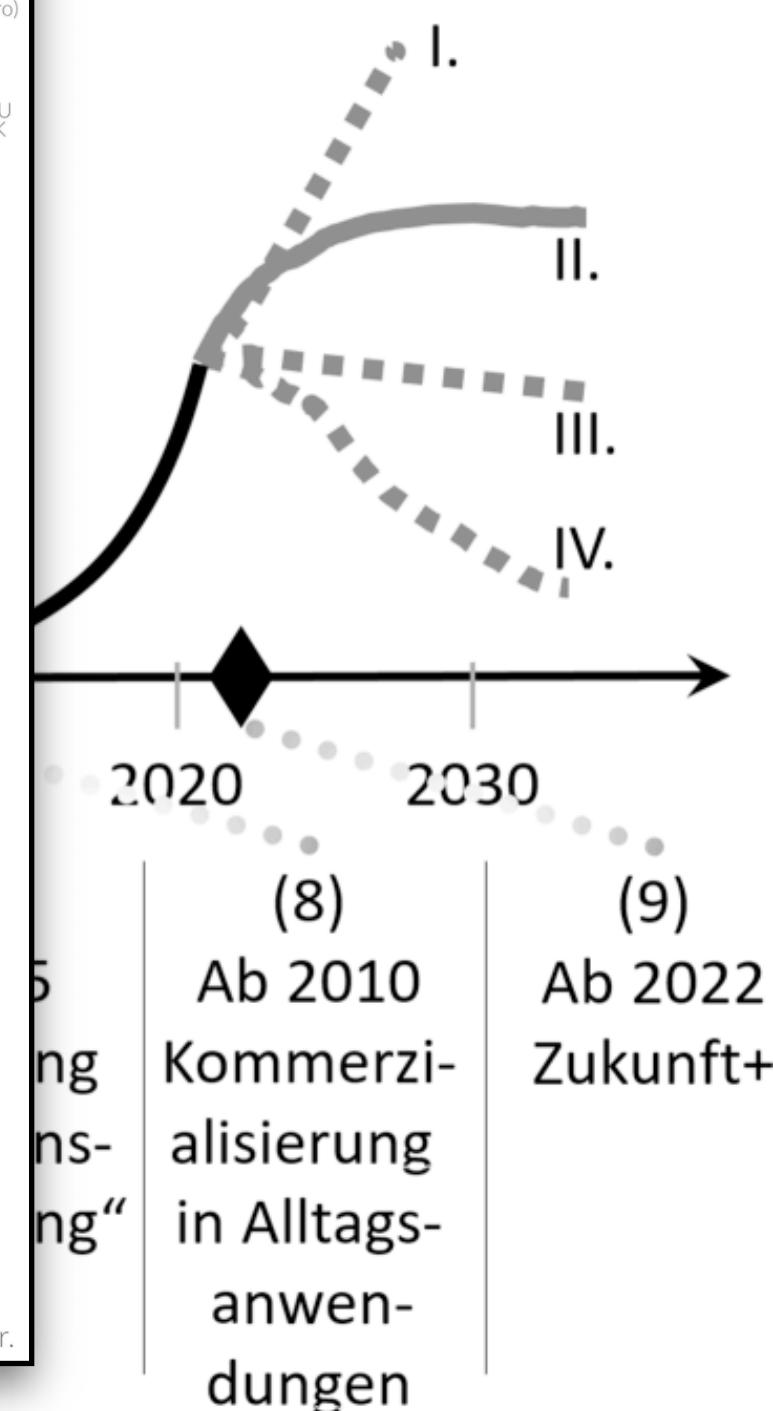
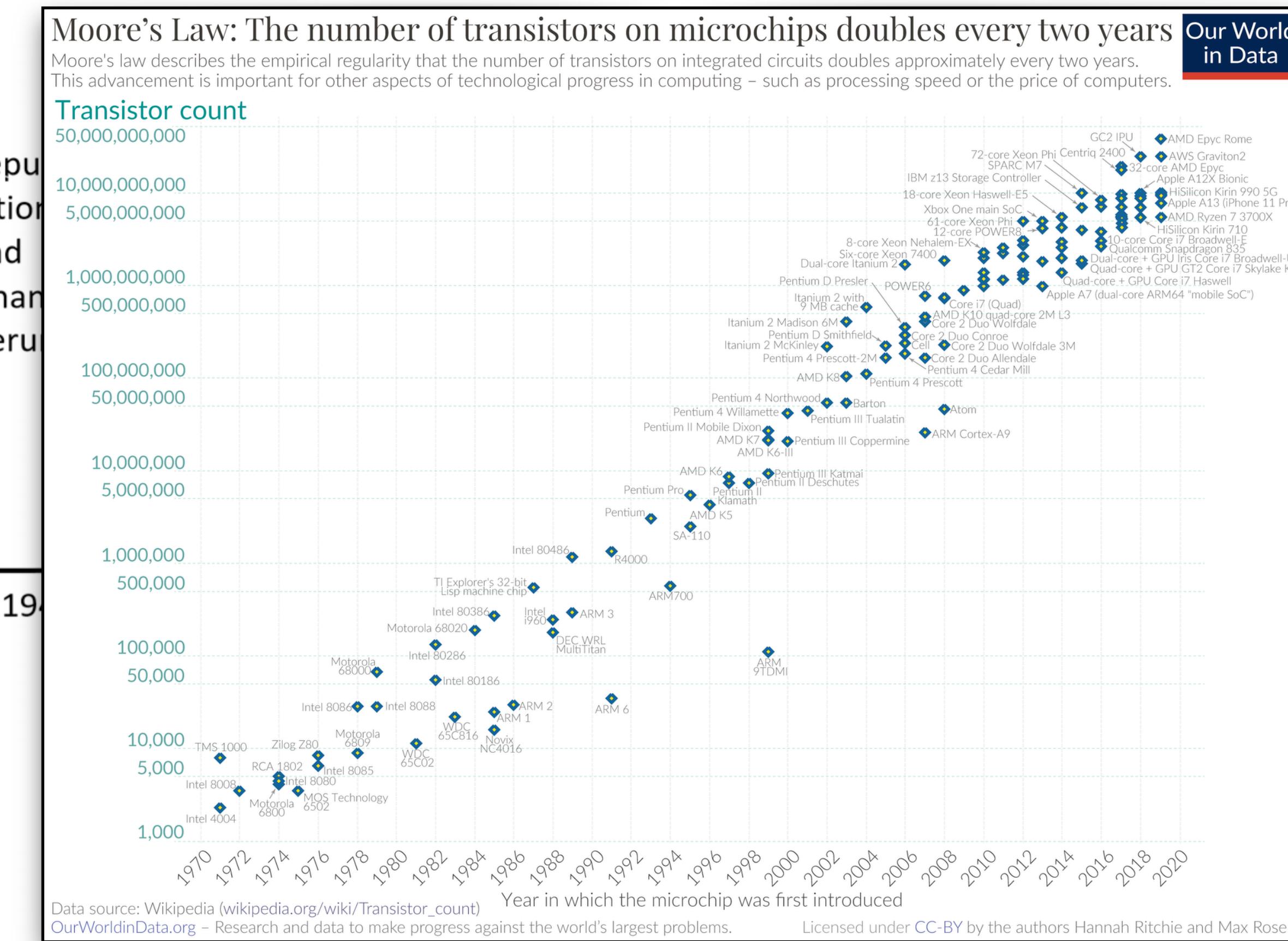


# (Moderne) Geschichte der KI

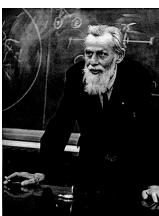


## Zunahme Rechenleistung

- Die Prozessorgeschwindigkeit hat zugenommen über die Jahre.
- Insbesondere die Zahl der Prozessoren in einem Computer haben zugenommen.
- Grafikkartenprozessoren (GPU) haben sich exponential verbessert.



1943 1955 1956



1966-1978

Keine Förderung  
für neuronale Netze  
für ca. 10 Jahre

1972

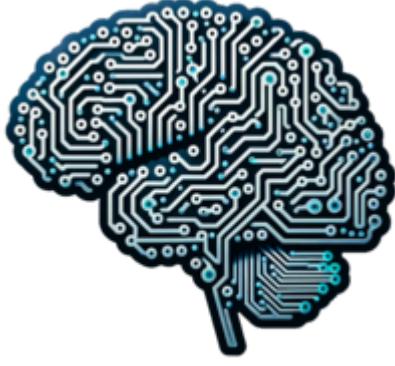
Prolog



1990



90-00 Jahre

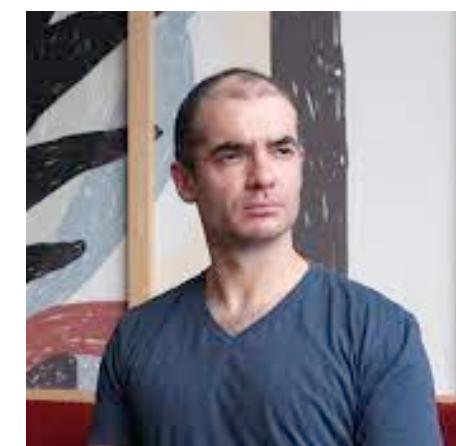


## Erfolgreiche NN

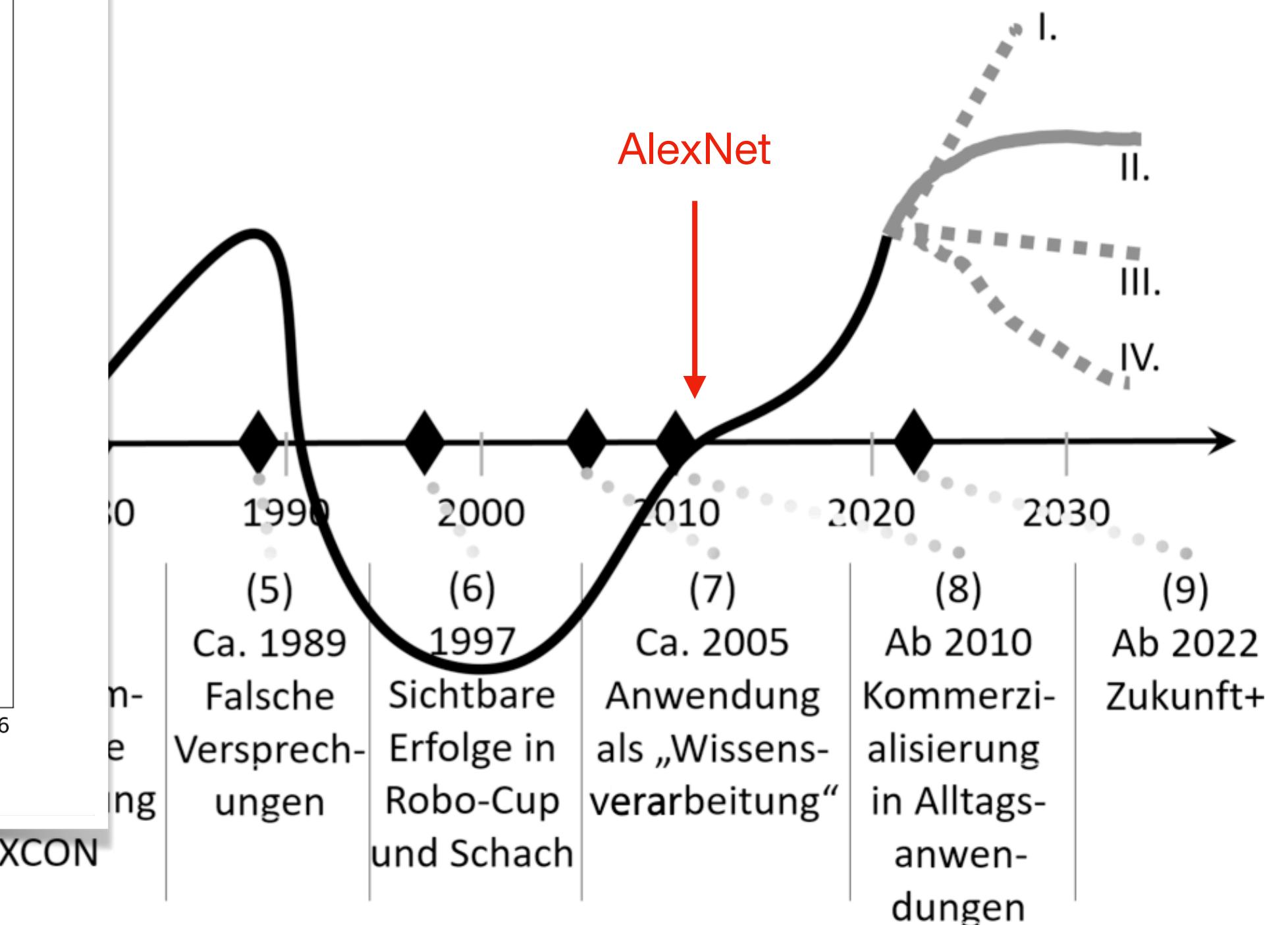
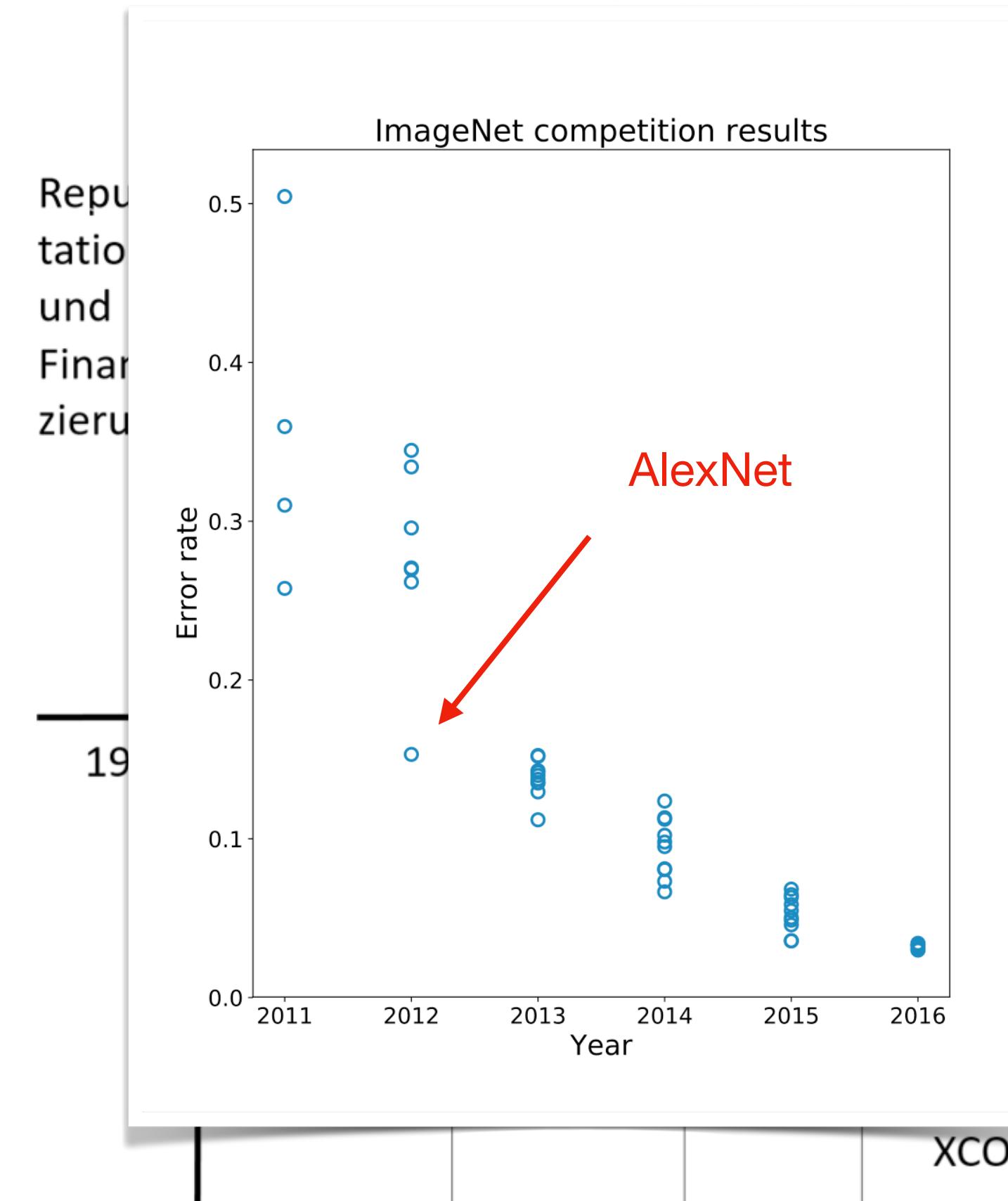
- Fei-Fei Li hat 14 Millionen Bilder mit ca. 20.000 Kategorien für Bilderkennung zusammengestellt (ImageNet)
- Convolutional Neural Network AlexNet war mit Abstand besser als Konkurrent



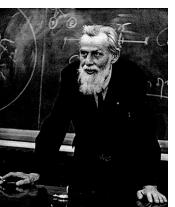
Krizhevsky



Sutskever



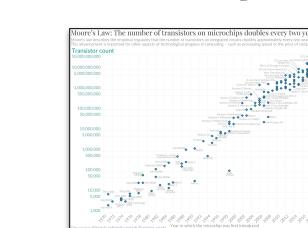
1943 1955 1956



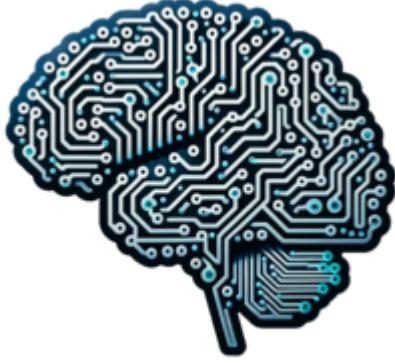
1966-1978 1972 1982 1990 90-00 Jahre 2012

Keine Förderung  
für neuronale Netze  
für ca. 10 Jahre

Prolog

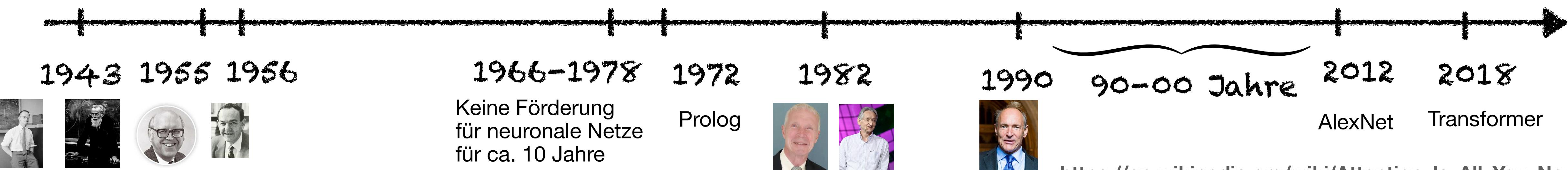
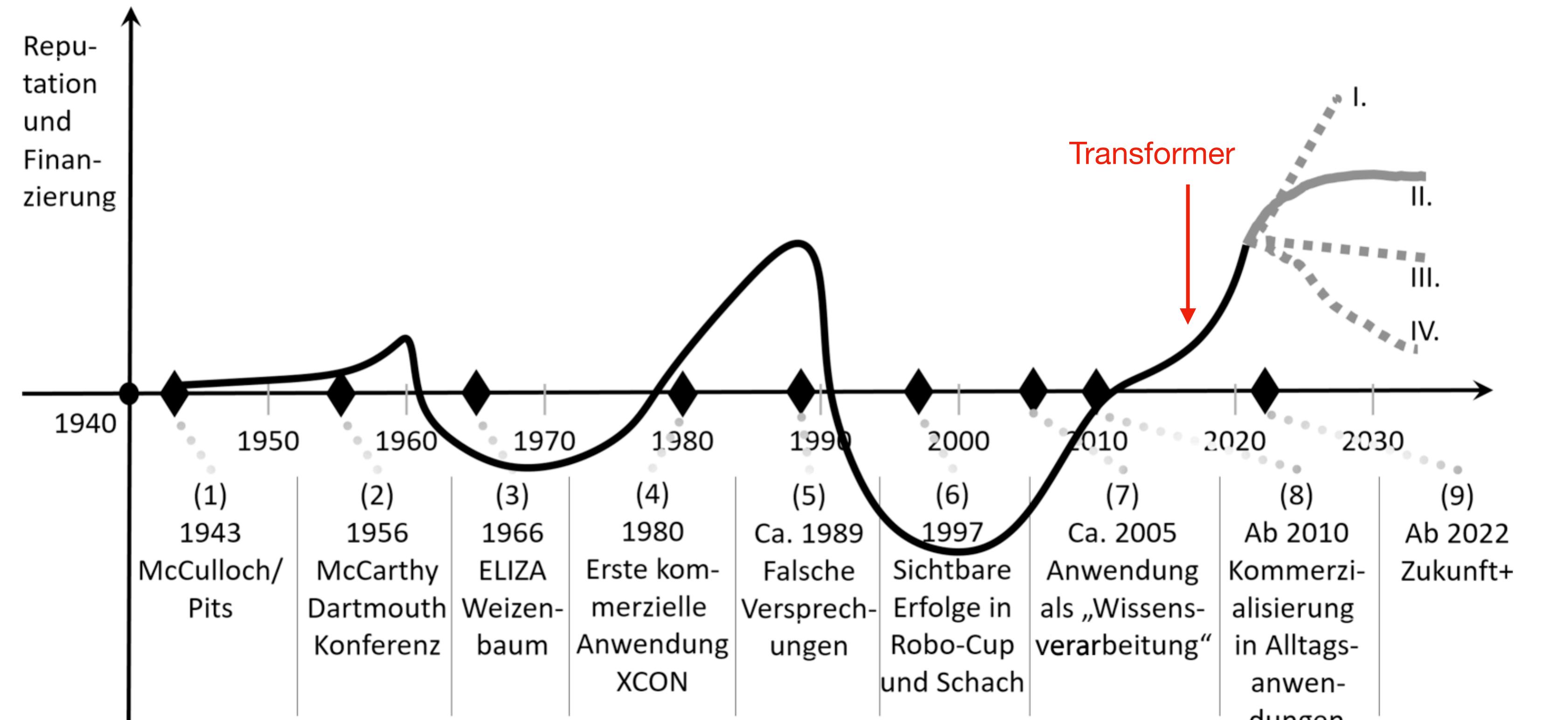


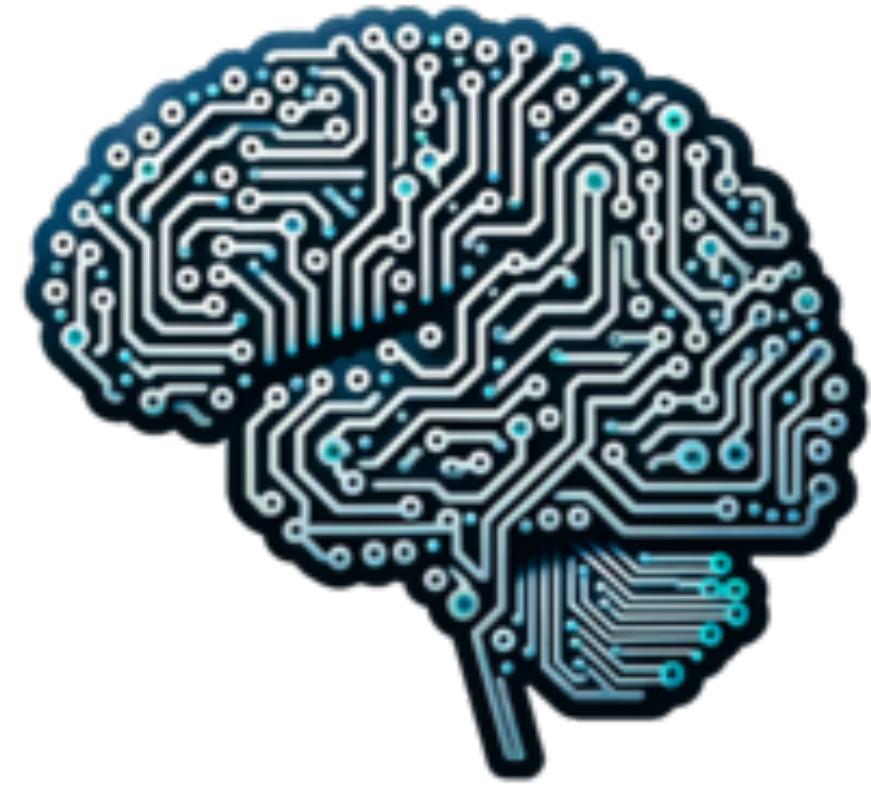
AlexNet



## Transformer

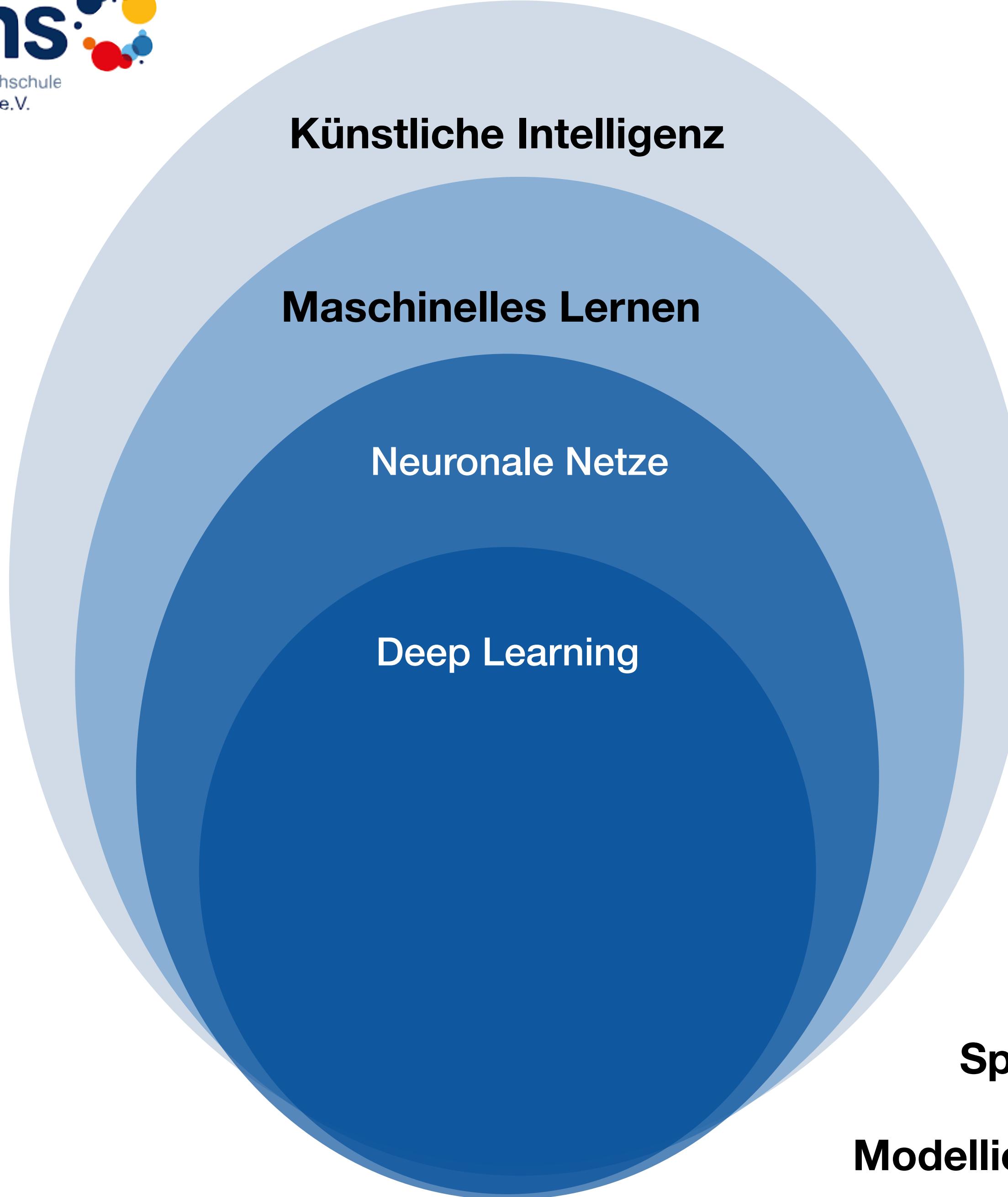
- Google veröffentlicht Publikation „Attention is all you need“
- Führt Transformer-Architektur bei Convolutional Neural Networks ein
- Wird als „Founding Document“ der moderner KI angesehen
- Vorgeschlagene Architektur ist Hauptarchitektur für Large Language Models





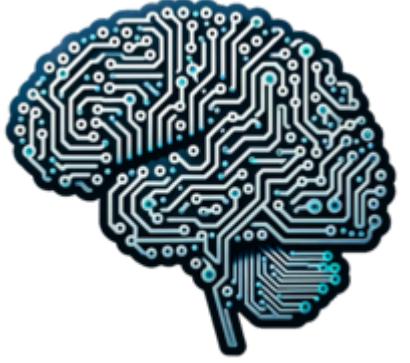
# Prinzipien der KI

**Dr. Frank Zimmer**



## Künstliche Intelligenz

Erzeugen einer Maschine, die den Menschen und seine rationales Entscheidungen simuliert



## Maschinelles Lernen

Computer erlernen mittels vieler Daten (Big Data)

## Neuronale Netze

Nachempfindung des menschlichen Gehirns im Computer

## Deep Learning

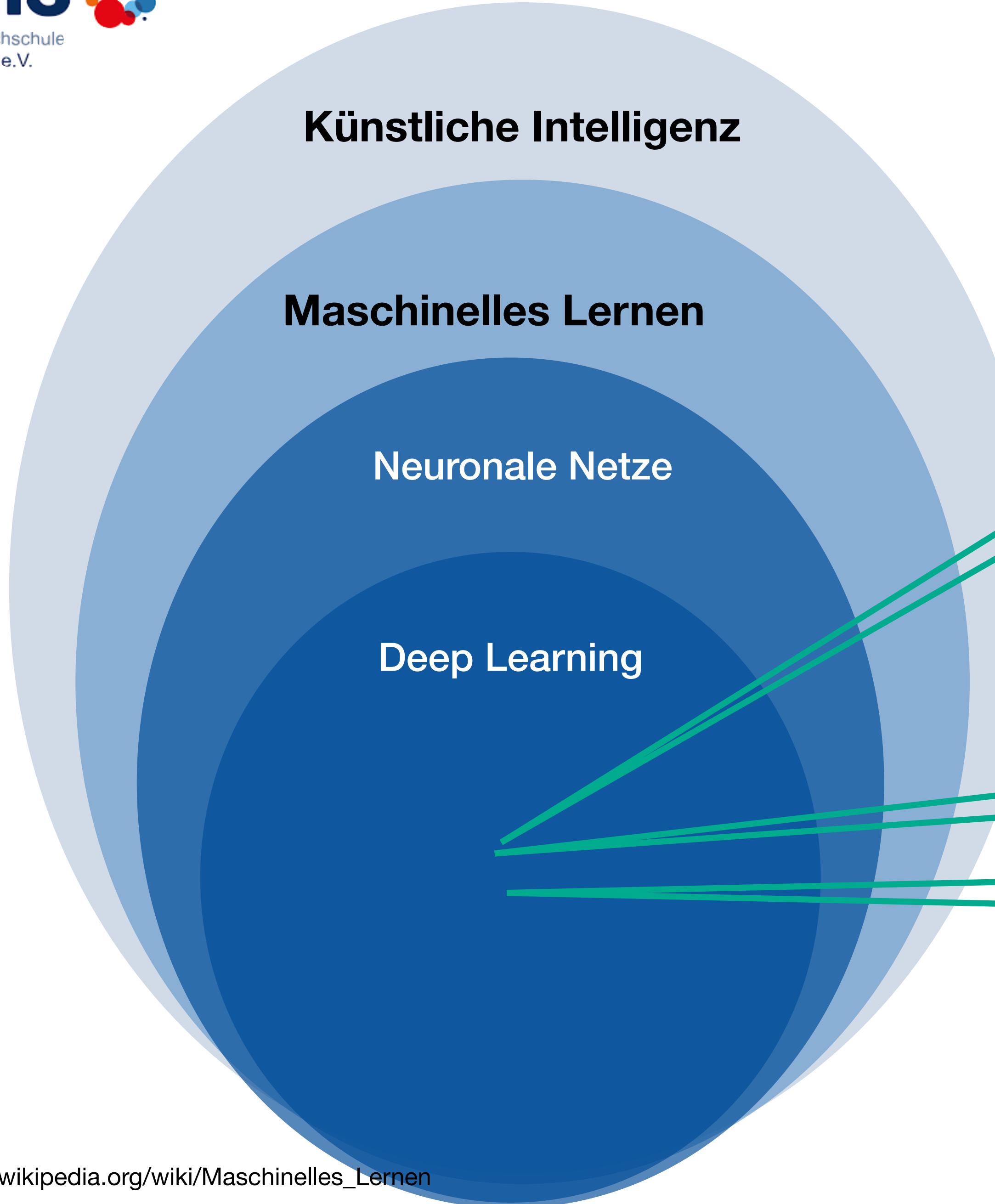
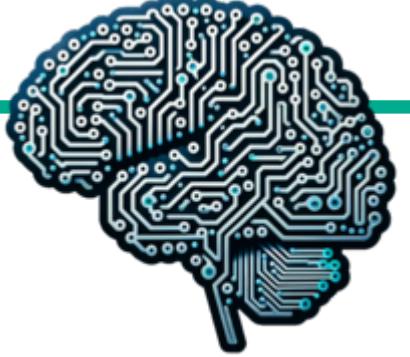
Große neuronale Netze verbunden mit vielen Daten

### Anwendungsgebiete

**Computer Vision**  
**Spracherkennung**

**Modellierung**

**Bildanalyse**  
**Sprachverarbeitung**  
**Übersetzungen**  
**Taktische Spiele**



## **Supervised Learning - Überwachtes Lernen**

Dies entspricht dem Lernen in der Schule. Das zu Lernende wird dem Computer in Form von Trainingsdaten bereitgestellt. Durch ein Fehlerfunktion (Note), wird dem Computer gesagt wie gut er die Konzepte gelernt hat. Das Lernen wird so lange wiederholt, bis die gemachten Fehler möglichst klein sind. Die Trainingsdaten wurde von Menschen vorher kategorisiert.

## **Unsupervised Learning - Unüberwachtes Lernen**

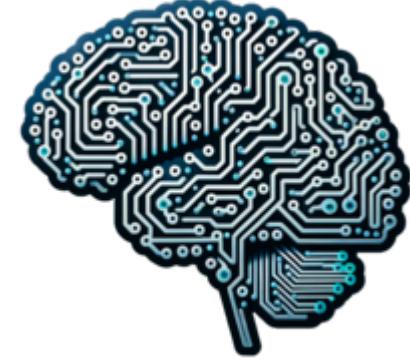
Dies entspricht dem Lernen einer Sprache durch ein Baby. Der Computer erkennt selbst die Muster/ Kategorien in der zur Verfügung gestellten Daten.

## **Reinforcement Learning - Bestärktes Lernen**

Dies entspricht dem Lernen beim Spielen. Die SpielerIn erhält für seine Handlungen Belohnung oder Bestrafungen. Der Lernende fest dabei seinen Gewinn/ Wohlsein zu maximieren.

# Neuronale Netze

## Unterschiede zwischen Mensch und Computer



Mensch kann viele komplexe Dinge:

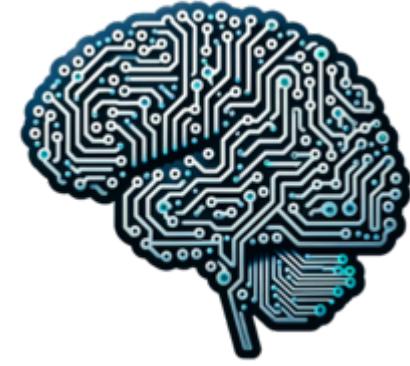
- Hören, Sprechen, Sehen
- Krabbeln, Gehen, Schwimmen
- Erkennen, Erlernen, etc

Obwohl:

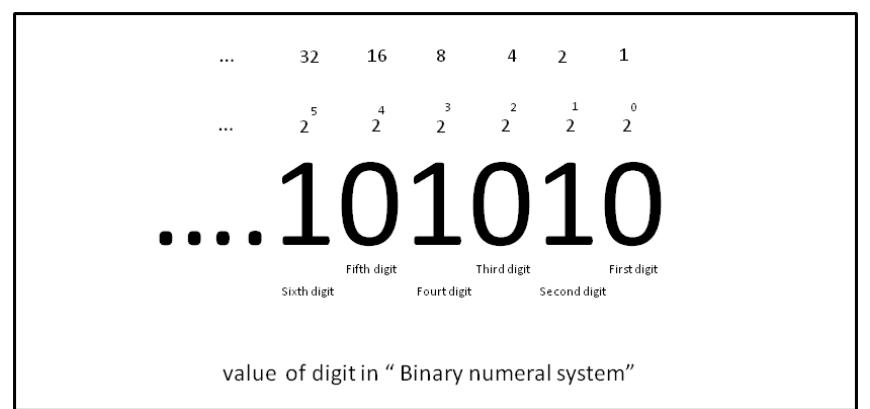
- „Denkgeschwindigkeit“ klein ist
- Er/sie keine schnellen Berechnung dafür durchführen kann / muss

# Neuronale Netze

## Unterschiede zwischen Mensch und Computer



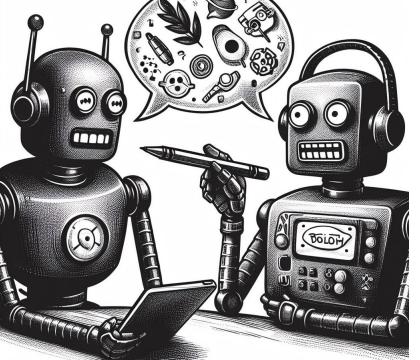
**Einfach**



Rechnen

Große Datenmengen

**Langsam  
Schwer**



ab ca 2020



ab 2012

Der Computer

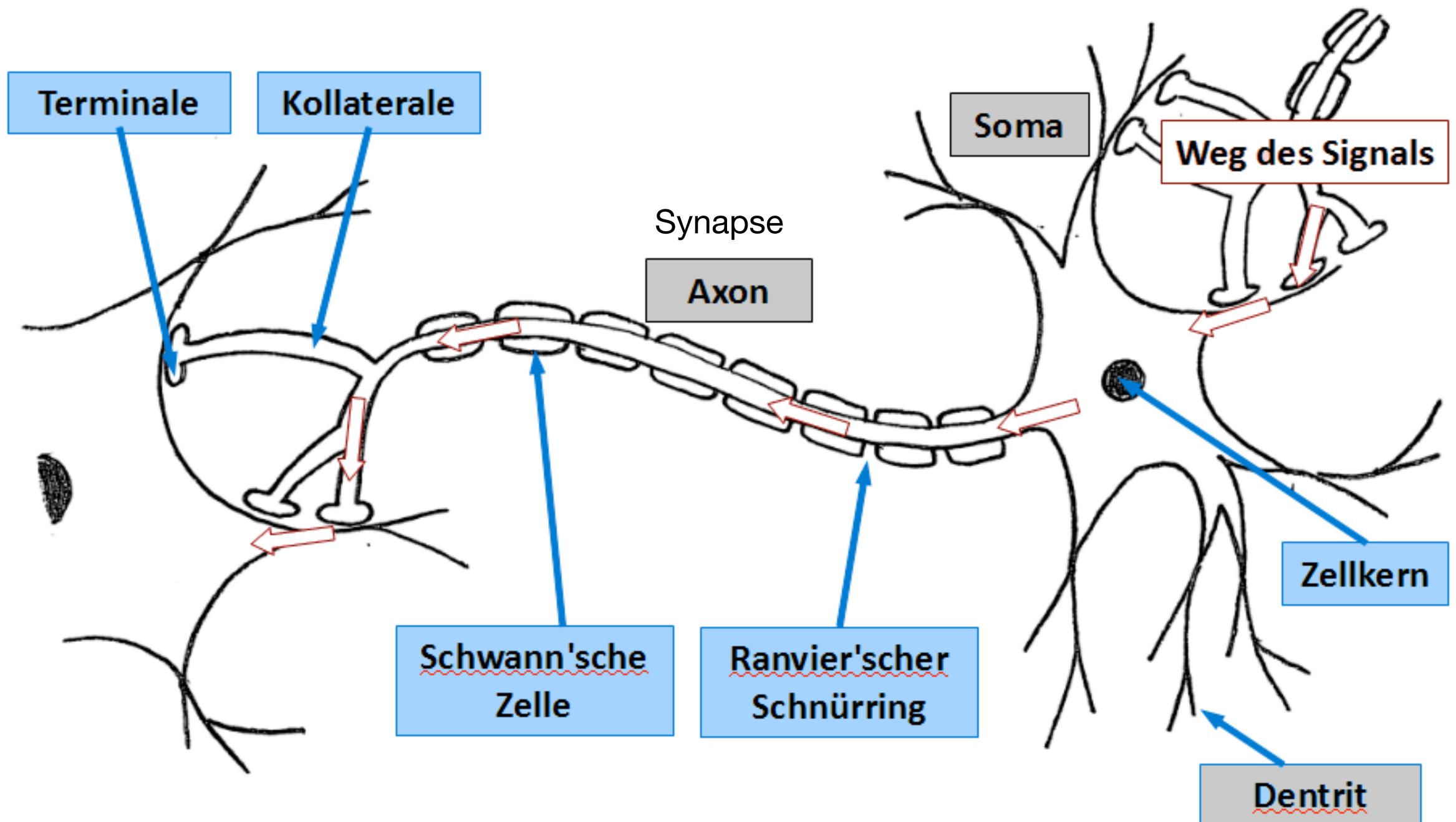
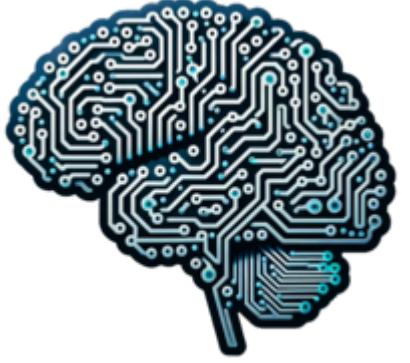
- arbeitet mit Hoher Taktfrequenz im GHz-Bereich
- kann große Zahlenmengen ohne Fehler manipulieren

Aber

- Hat erst in den letzten Jahren gelernt:
  - Gesprochene Sprache und Texte zu „Verstehen“
  - Bilder und ihren Inhalt zu erkennen
  - Bewegungen ähnlich einem Menschen durchzuführen

# Neuronale Netze

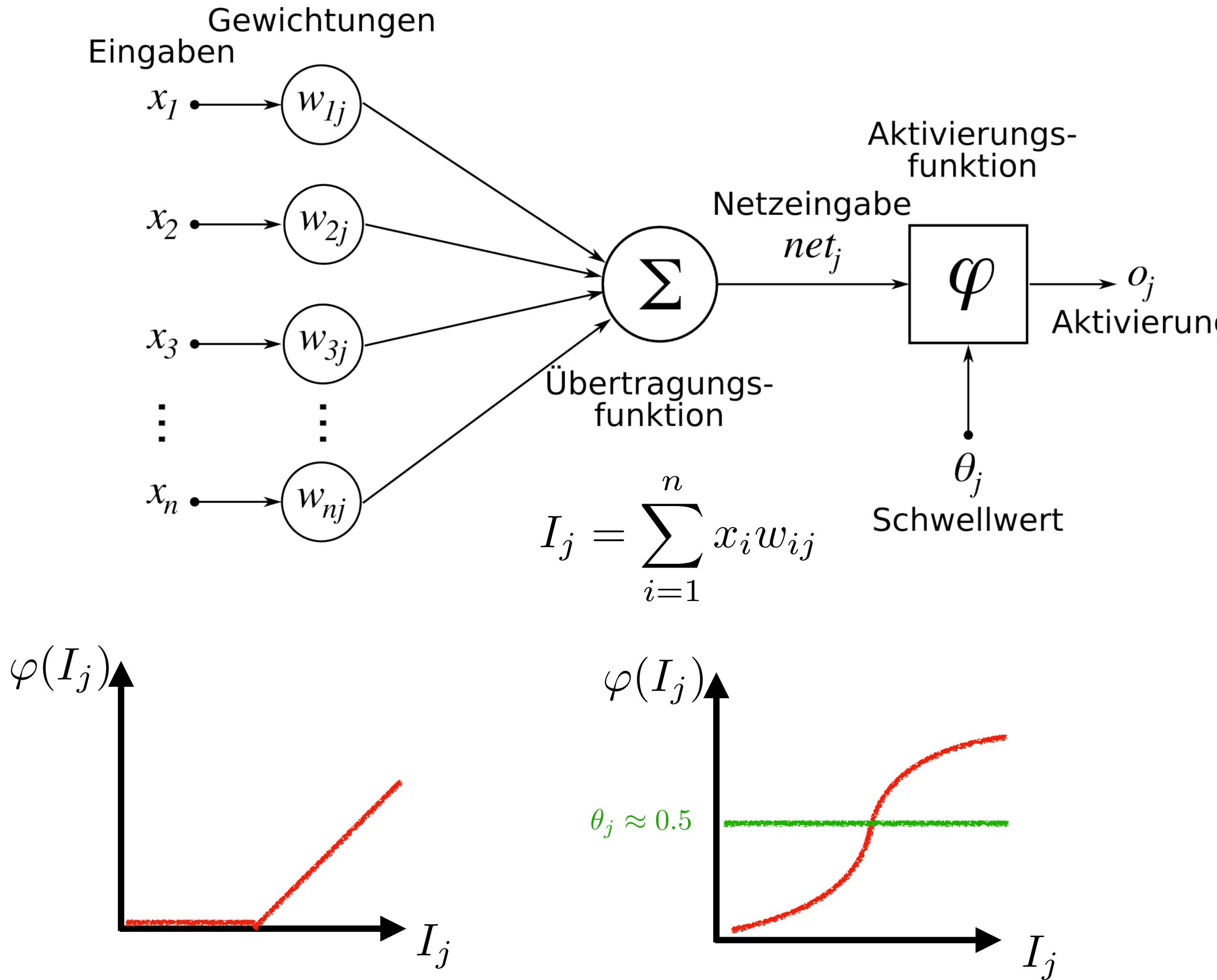
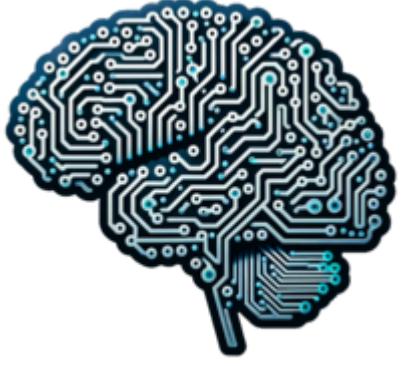
## Wie funktioniert das Lernen beim Menschen ?



- Gehirn besteht aus Milliarden von **Neuronen**, die miteinander verbunden sind
- Gedanken/Gefühle/Informationen wandern via **elektrischer** oder **chemischer Impulse**
- Signale werden nur weitergeleitet, wenn ein **Schwellspannung** überschritten wird
- Menschliches Lernen führt zur **Verstärkung** oder Entstehung neuer **neuronaler Verbindungen**
- Vergessen entspricht dem Schwächerwerden oder Verschwinden von neuronalen Verbindungen

# Neuronale Netze

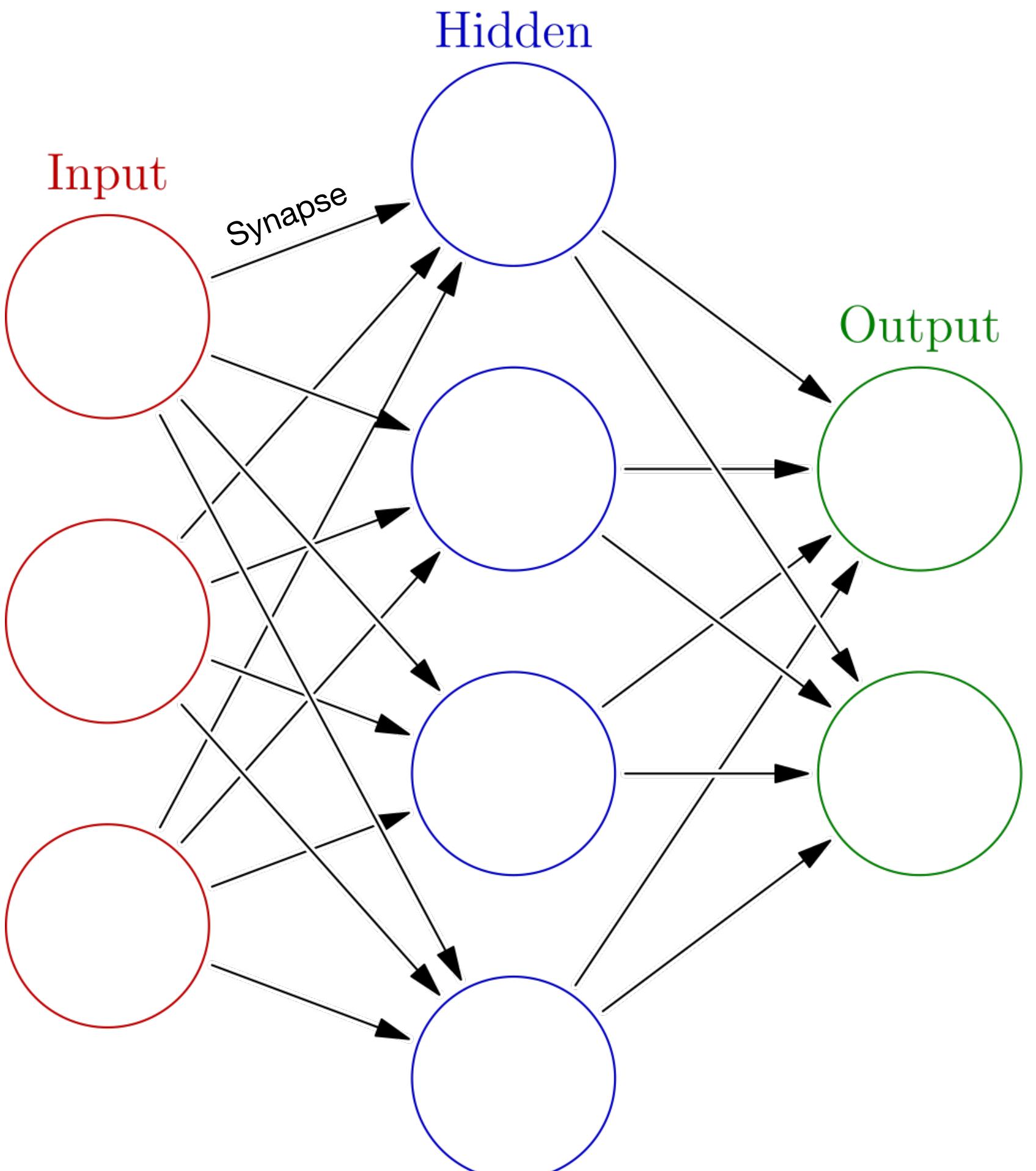
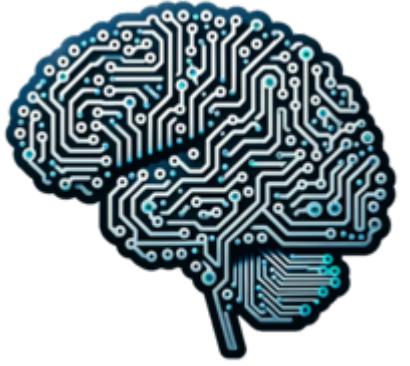
## Wie funktionieren künstliche Neuronen ?



- Grundbaustein eines neuronalen Netzes
- Führen mehrere Eingangssignale in ein Ausgangssignal über
- Hängen von einer Übertragungs- und Aktivierungsfunktion ab
- Unter einem bestimmten Schwellwert gibt es kein Ausgangssignal
- Die Signale sind hier reelle Zahlen

# Neuronale Netze

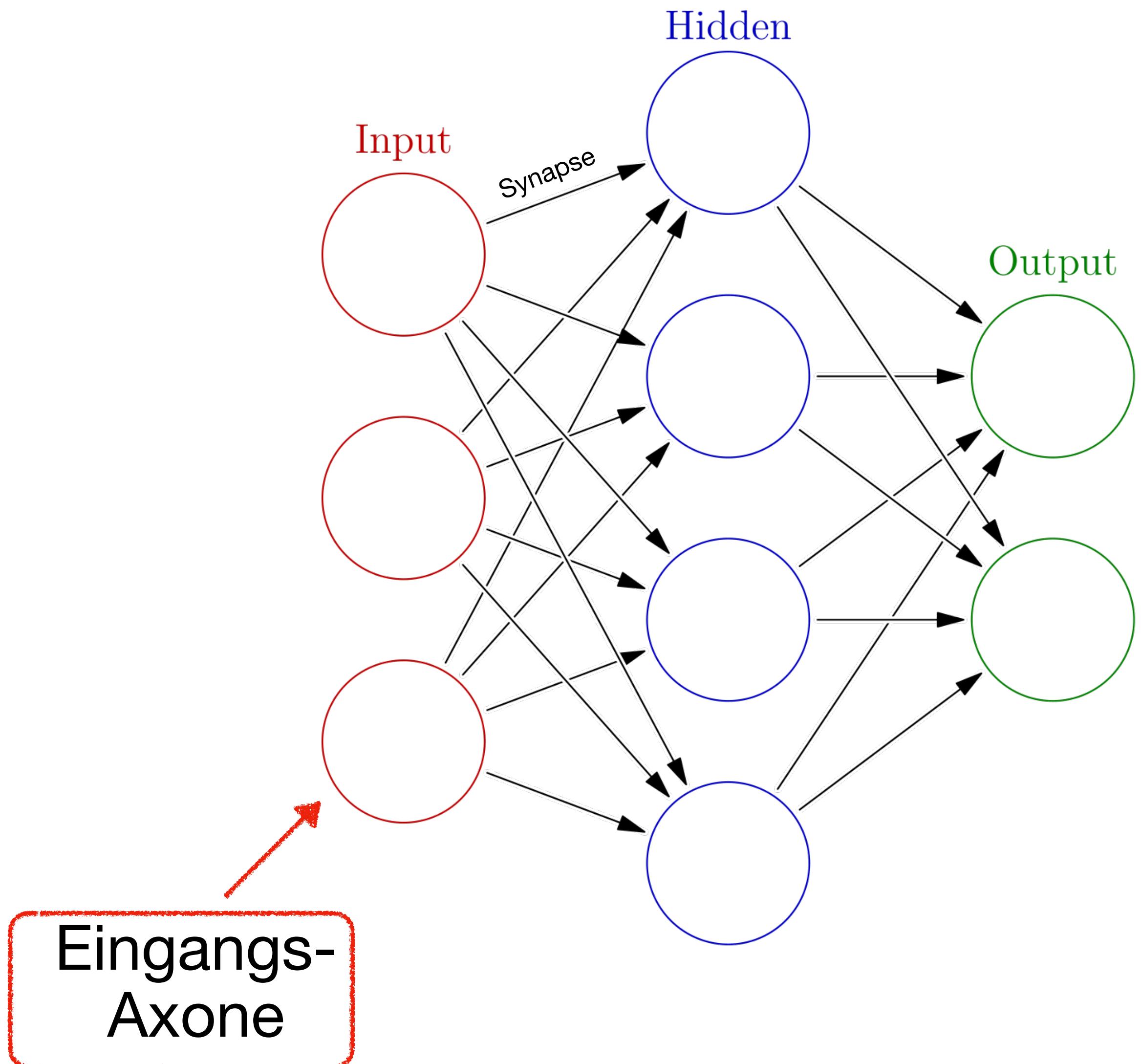
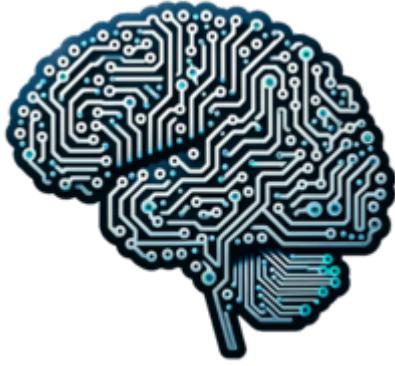
## Wie sind **Neuronale Netze** im **Computer** aufgebaut?



- Neuronale Netze bestehen aus Eingangs-, Hidden- und Ausgangsschicht
- Eingangsebene nimmt Eingangsdaten, z.B. Pixel in einem Bild, in Form eines Vektors auf
- Daten wandern über die gewichteten Verbindung von einer Neuronenschicht zur Nächsten
- Je nach Lernmethode werden Ausgangsdaten behandelt

# Neuronale Netze

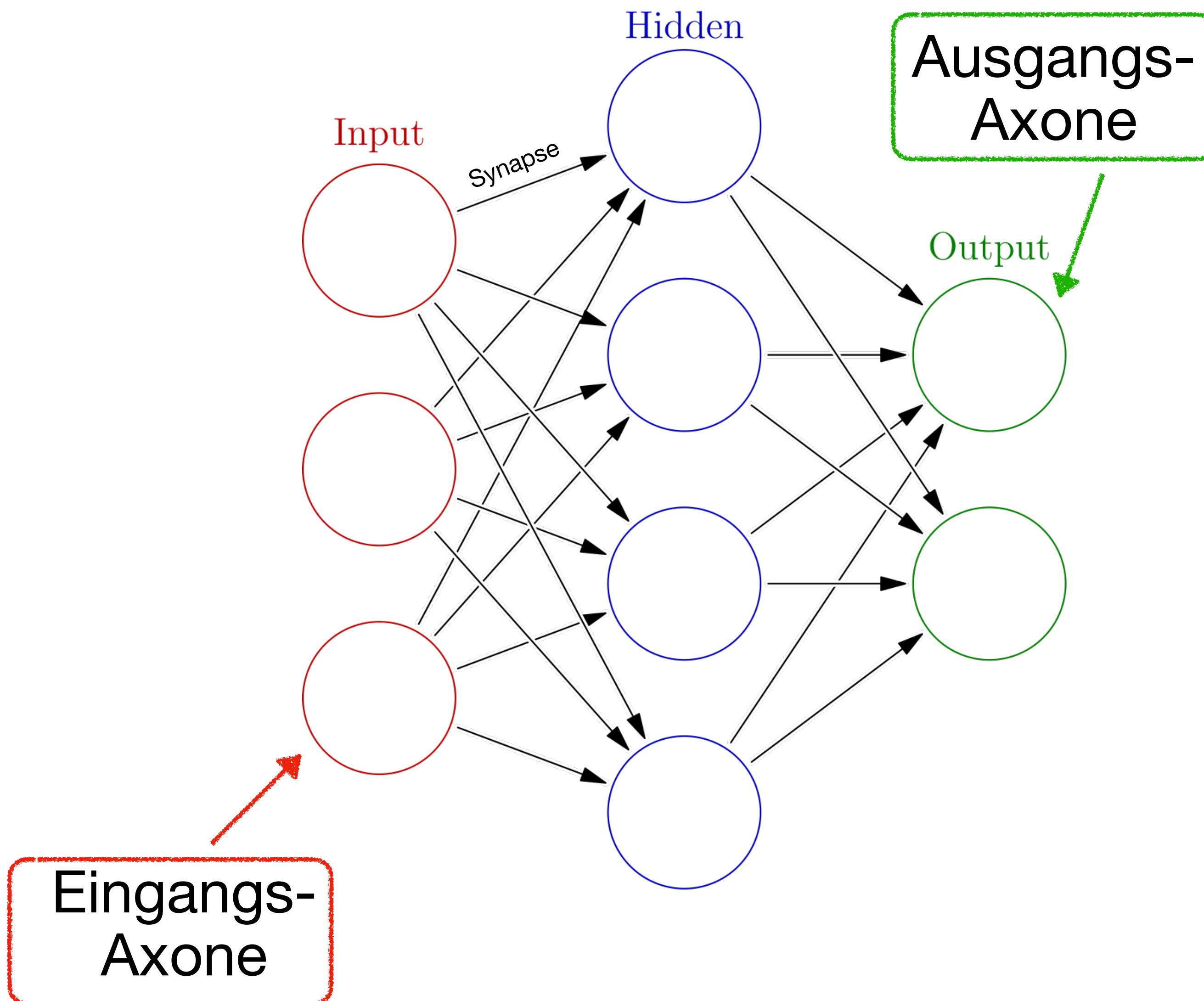
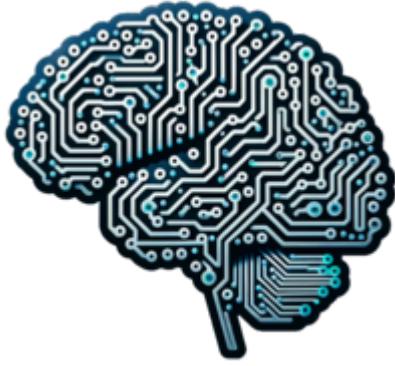
## Wie sind **Neuronale Netze** im **Computer** aufgebaut?



- Neuronale Netze bestehen aus Eingangs-, Hidden- und Ausgangsschicht
- Eingangsebene nimmt Eingangsdaten, z.B. Pixel in einem Bild, in Form eines Vektors auf
- Daten wandern über die gewichteten Verbindung von einer Neuronenschicht zur Nächsten
- Je nach Lernmethode werden Ausgangsdaten behandelt

# Neuronale Netze

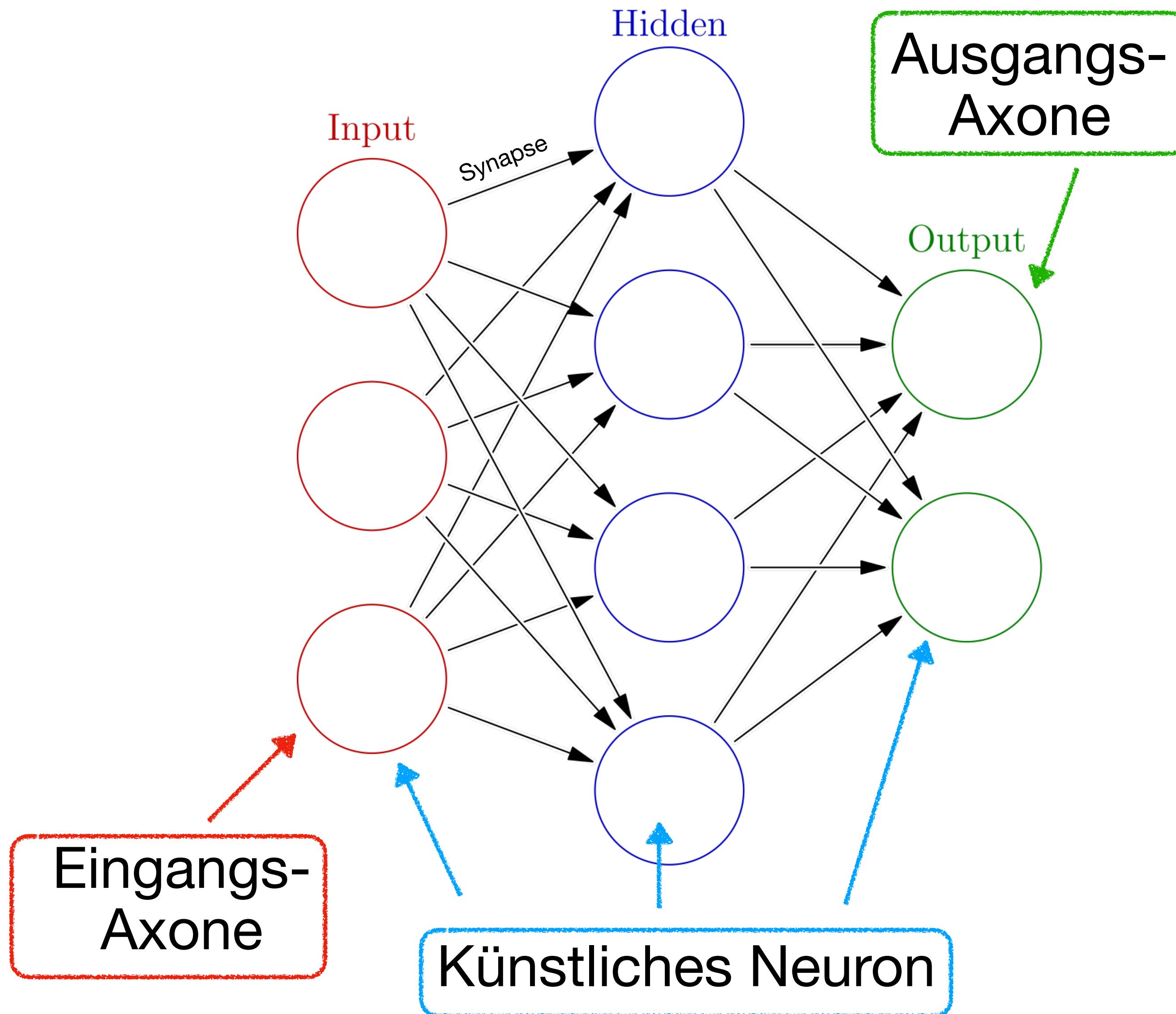
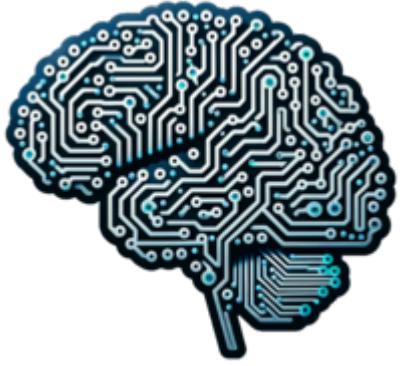
Wie sind **Neuronale Netze** im **Computer** aufgebaut?



- Neuronale Netze bestehen aus Eingangs-, Hidden- und Ausgangsschicht
- Eingangsebene nimmt Eingangsdaten, z.B. Pixel in einem Bild, in Form eines Vektors auf
- Daten wandern über die gewichteten Verbindung von einer Neuronenschicht zur Nächsten
- Je nach Lernmethode werden Ausgangsdaten behandelt

# Neuronale Netze

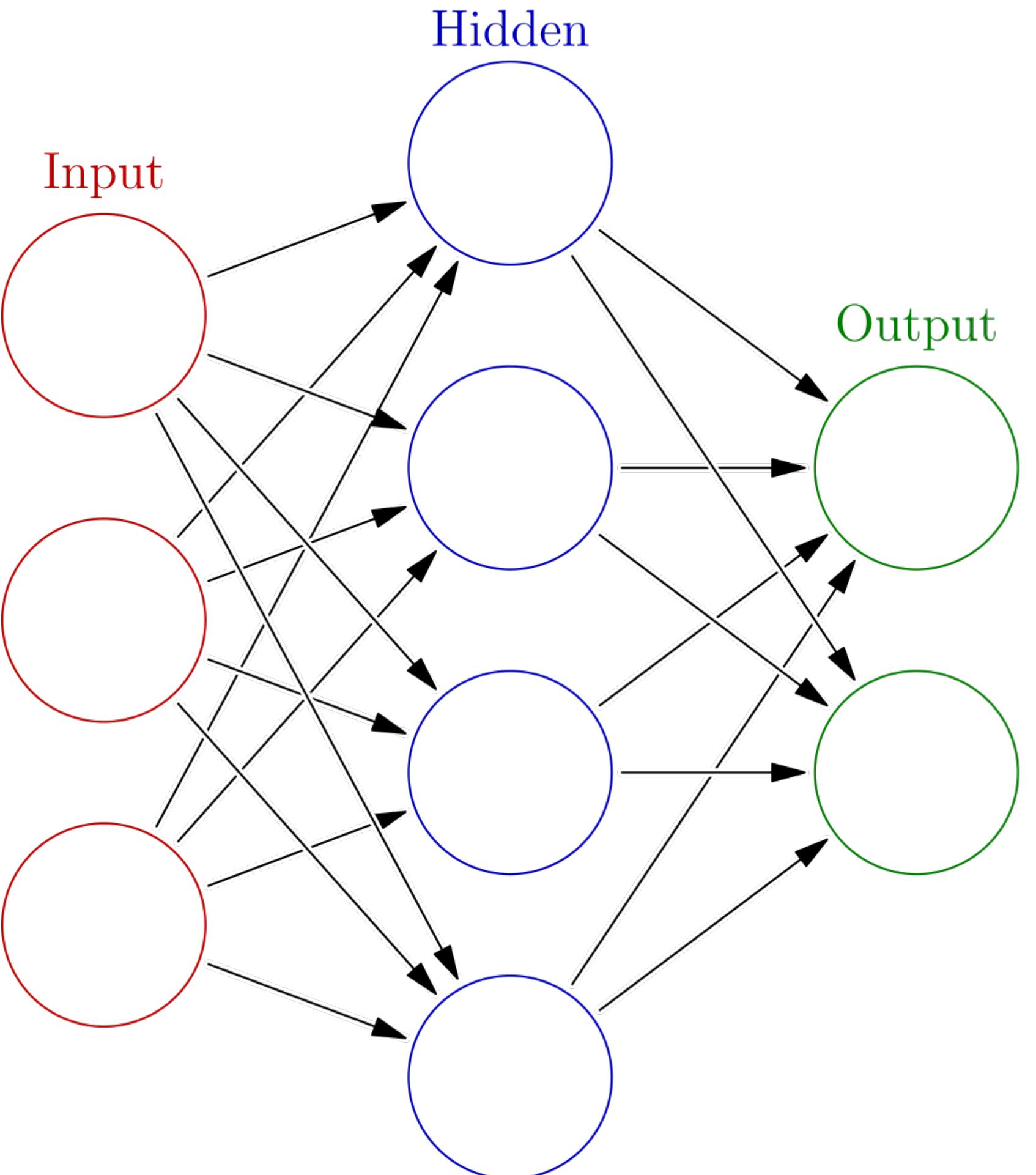
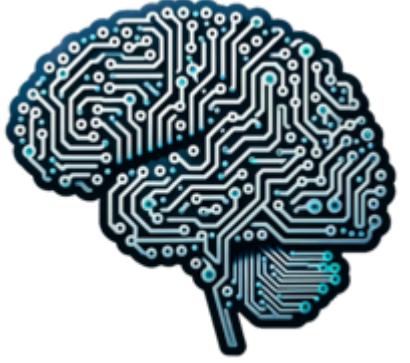
Wie sind **Neuronale Netze** im **Computer** aufgebaut?



- Neuronale Netze bestehen aus Eingangs-, Hidden- und Ausgangsschicht
- Eingangsebene nimmt Eingangsdaten, z.B. Pixel in einem Bild, in Form eines Vektors auf
- Daten wandern über die gewichteten Verbindung von einer Neuronenschicht zur Nächsten
- Je nach Lernmethode werden Ausgangsdaten behandelt

# Neuronale Netze

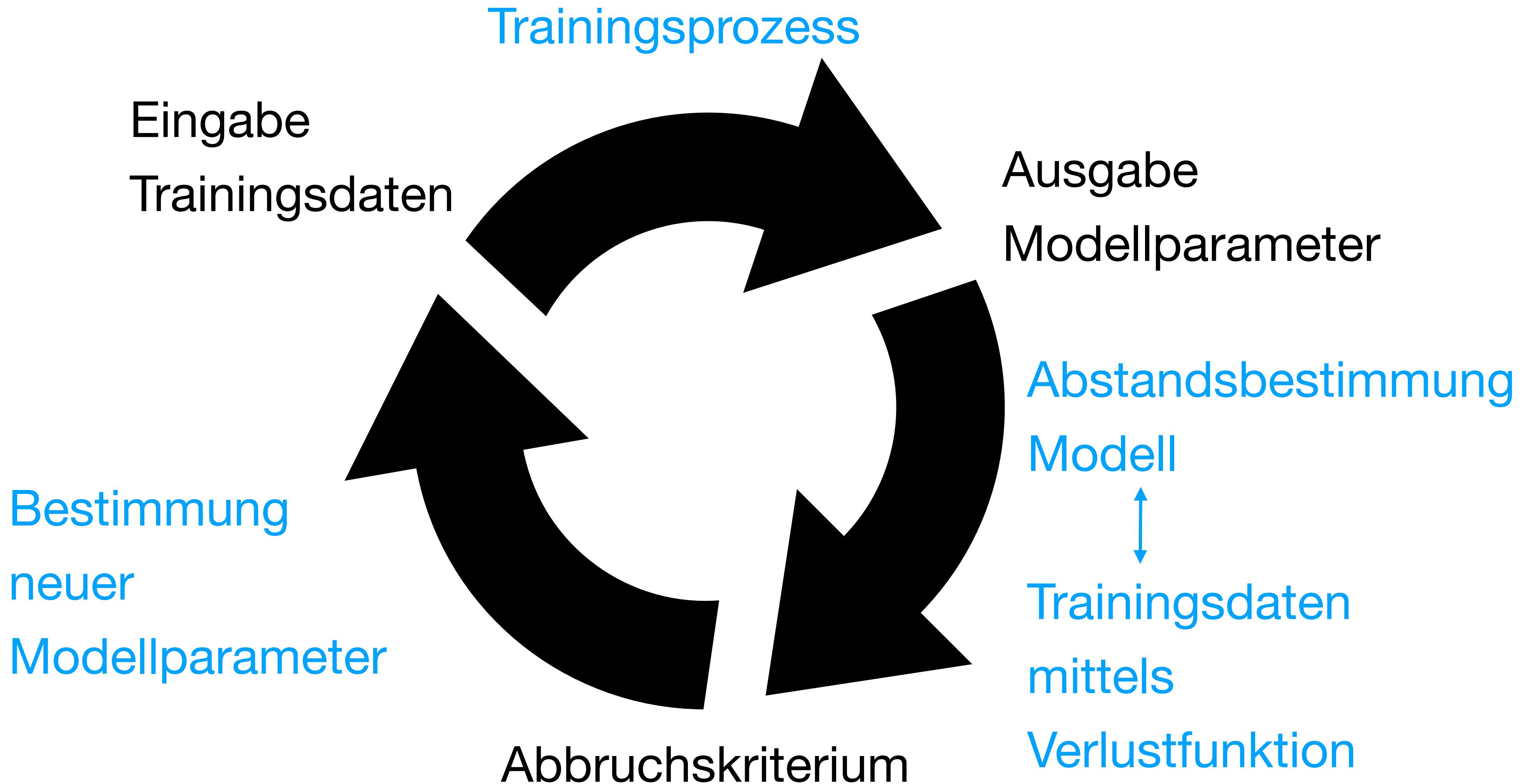
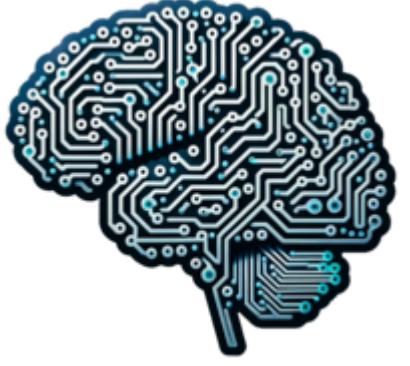
## Wie funktioniert das **Lernen** beim **Computer** ?



- Neuronale Netze lernen dadurch, dass die Gewichtungen durch einen repetitiven Prozess eingestellt werden
- Für die Einstellung der Gewichtungen werden immer die aktuellen Ausgangsdaten herangezogen
- Der Vergleich wird ermittelt über die aktuellen Vorhersagen des Modells mit den durch Menschen zur Verfügung gestellten Ausgangsdaten

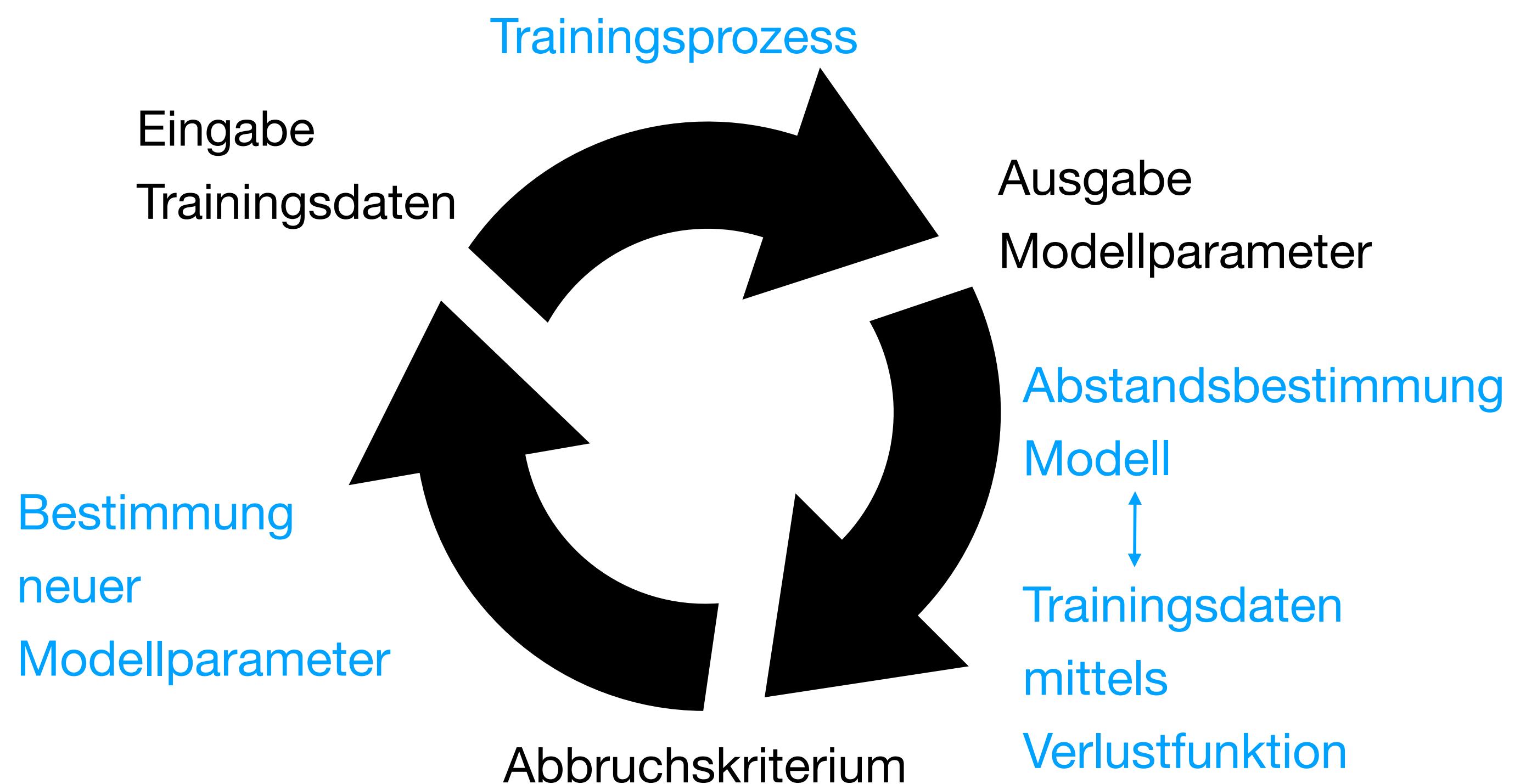
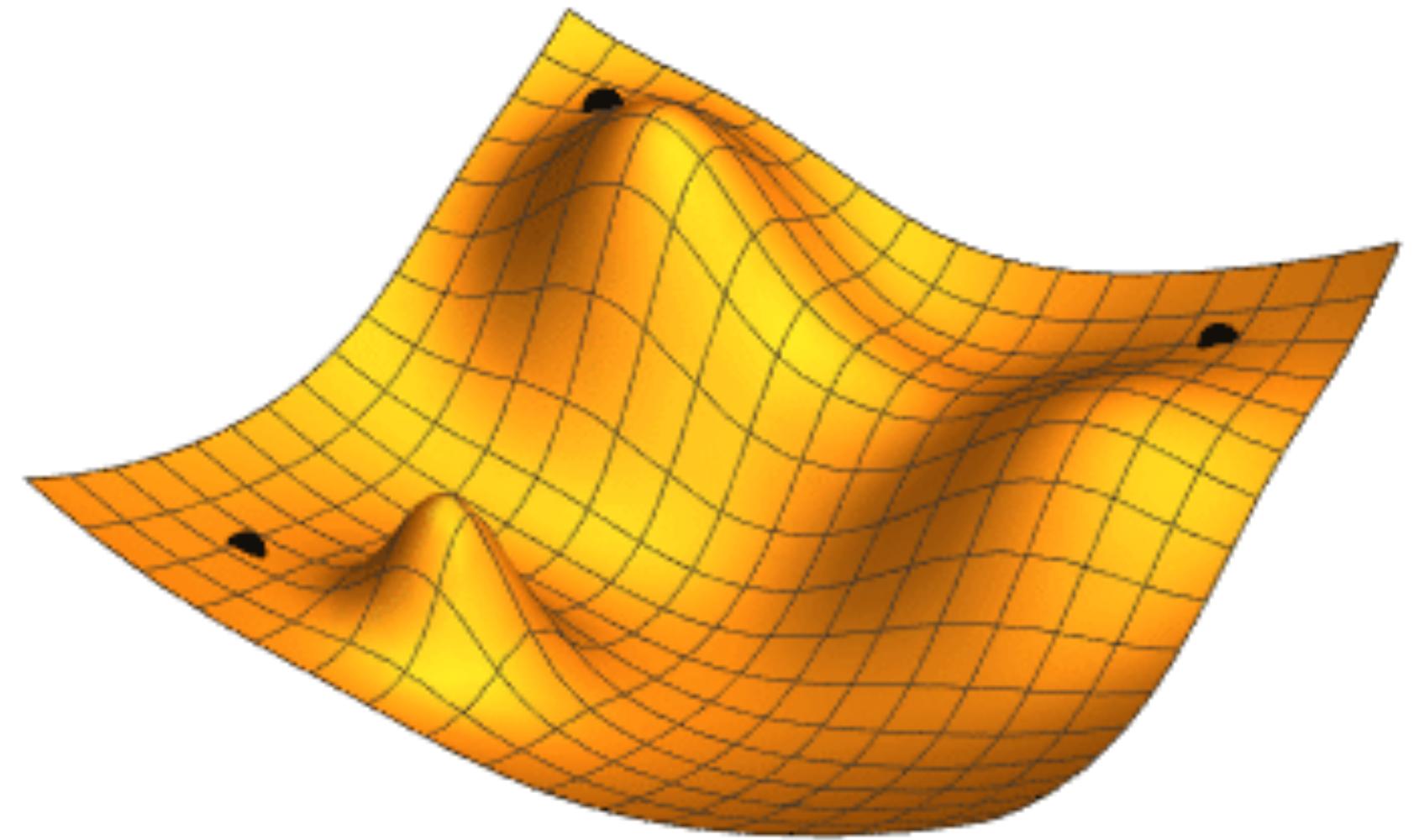
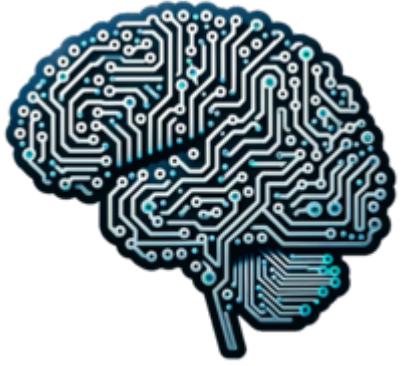
# Neuronale Netze

## Kreisprozess des Lernens



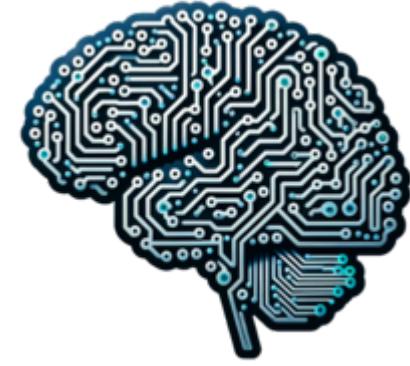
# Neuronale Netze

## Kreisprozess des Lernens

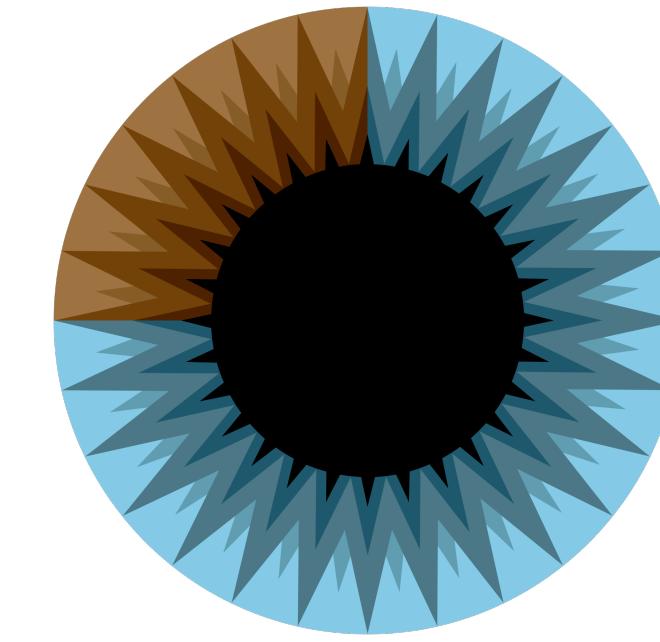


# Neuronale Netze

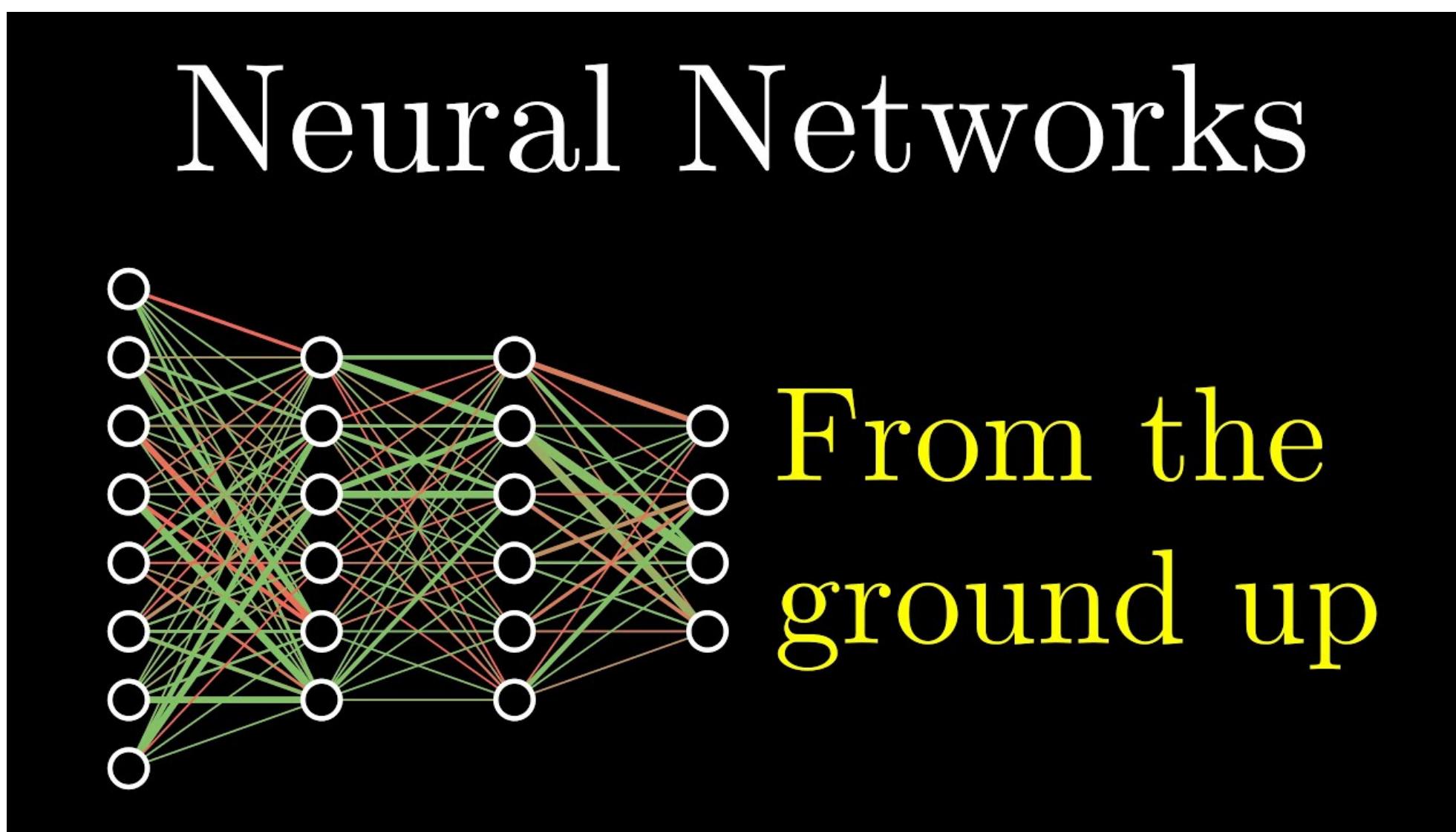
## Empfehlung zum Thema



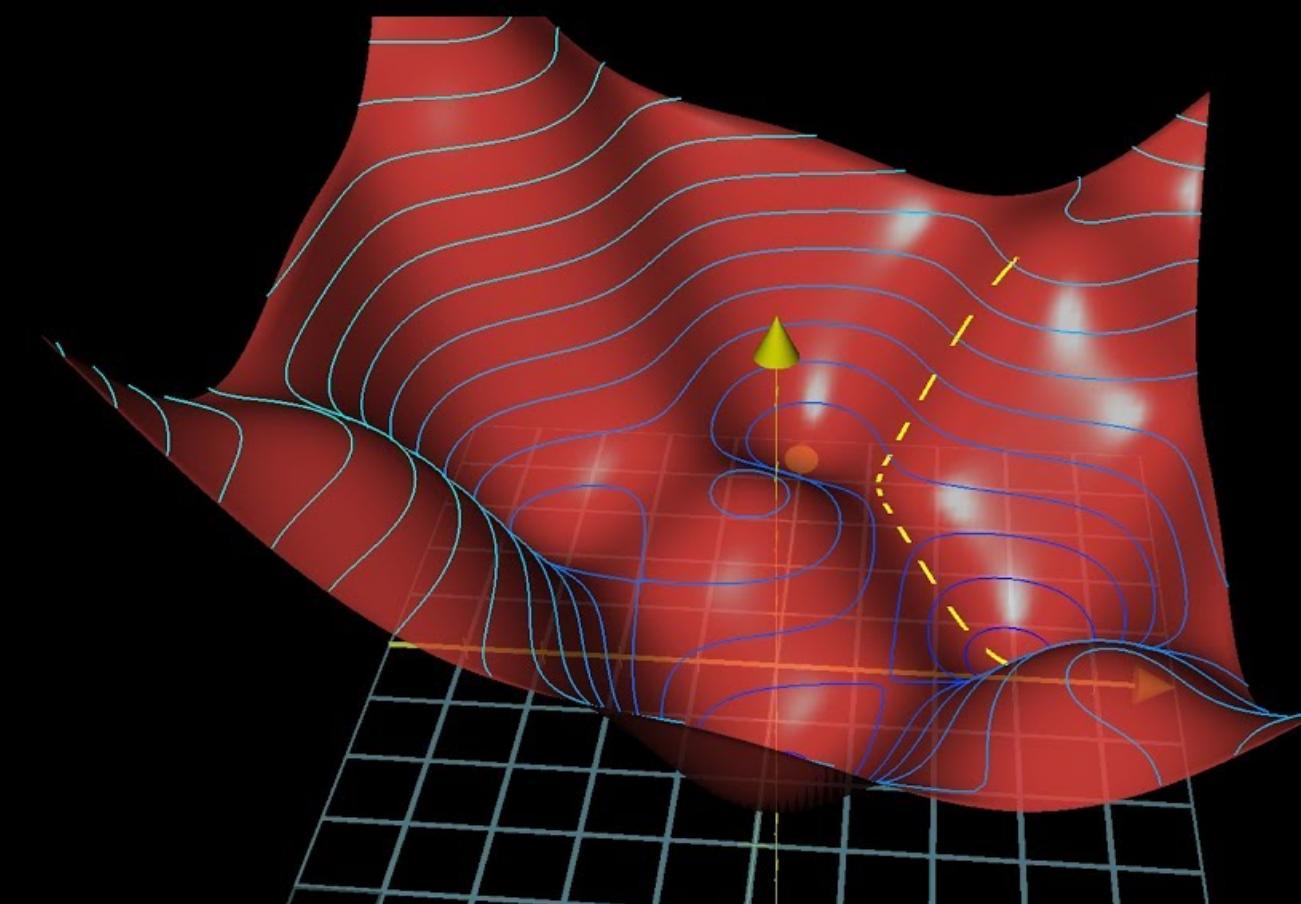
Grant Sanderson



3Blue1Brown



How machines learn





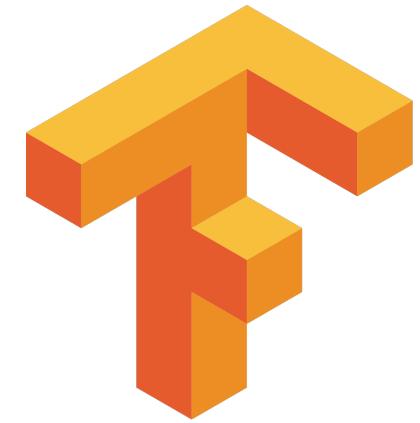
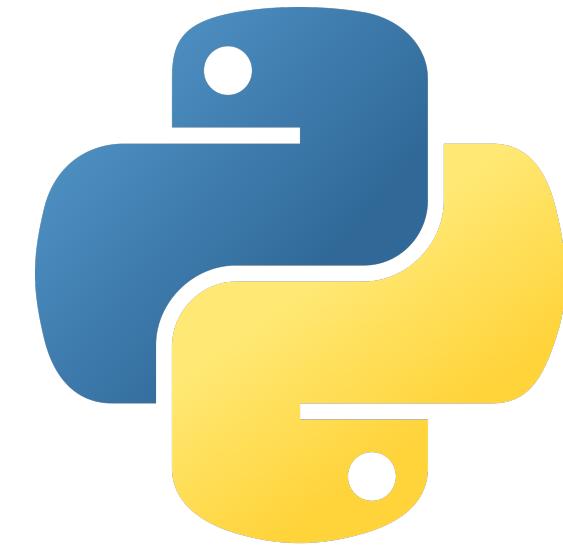
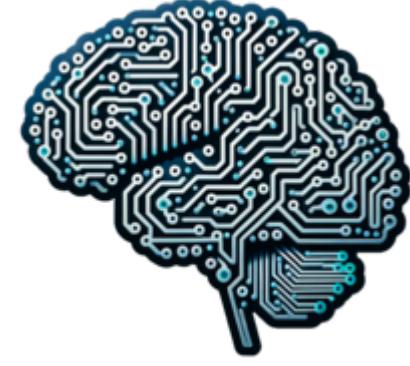
# **Trainieren eines NN**

## **Überblick und Beispiel**

**Dr. Frank Zimmer**

# Trainiere eines Neuronale Netzes

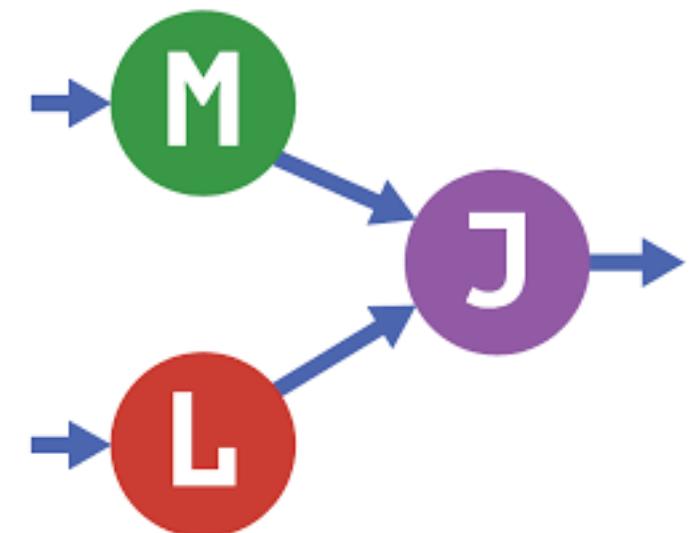
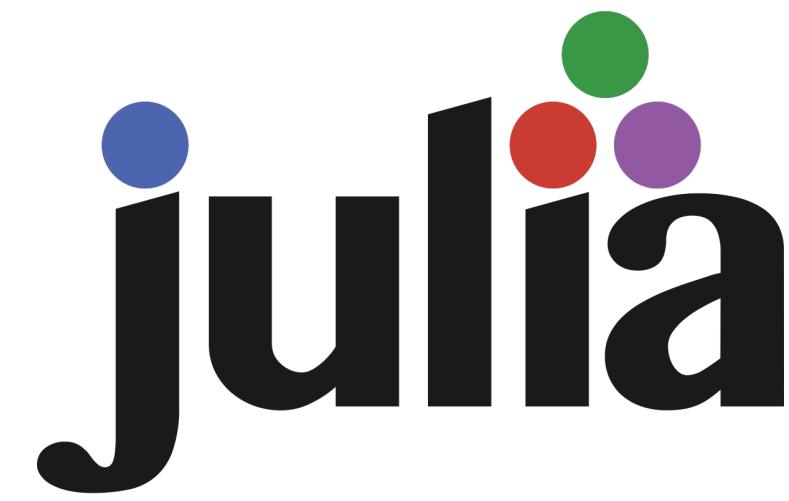
## Technologischer Überblick



Tensorflow



PyTorch



### Python

- **Vorteile:**

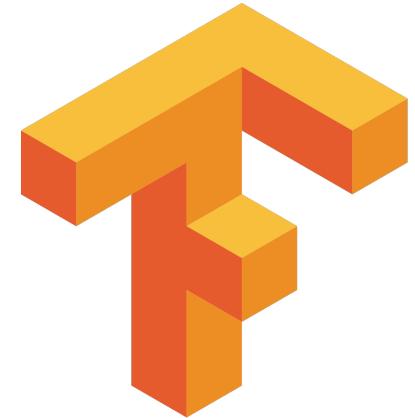
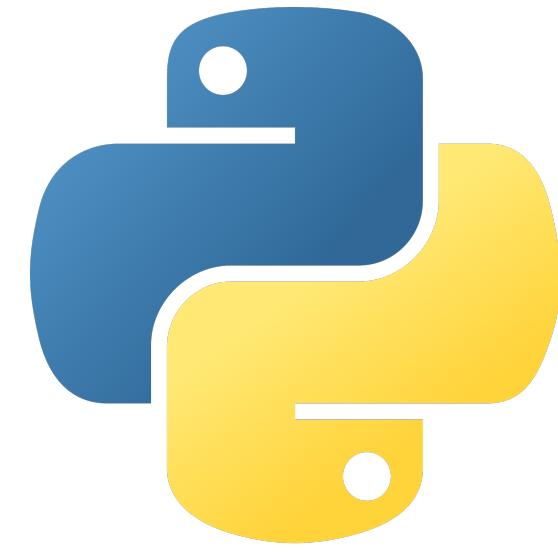
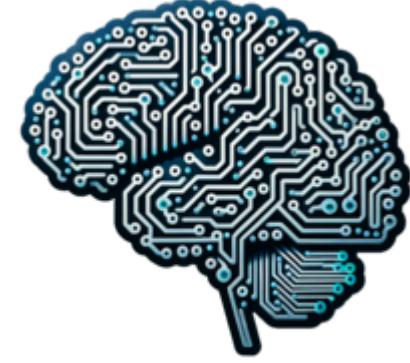
- Etablierte Programmiersprache mit großem Ökosystem
- Etablierte Frameworks

- **Nachteile:**

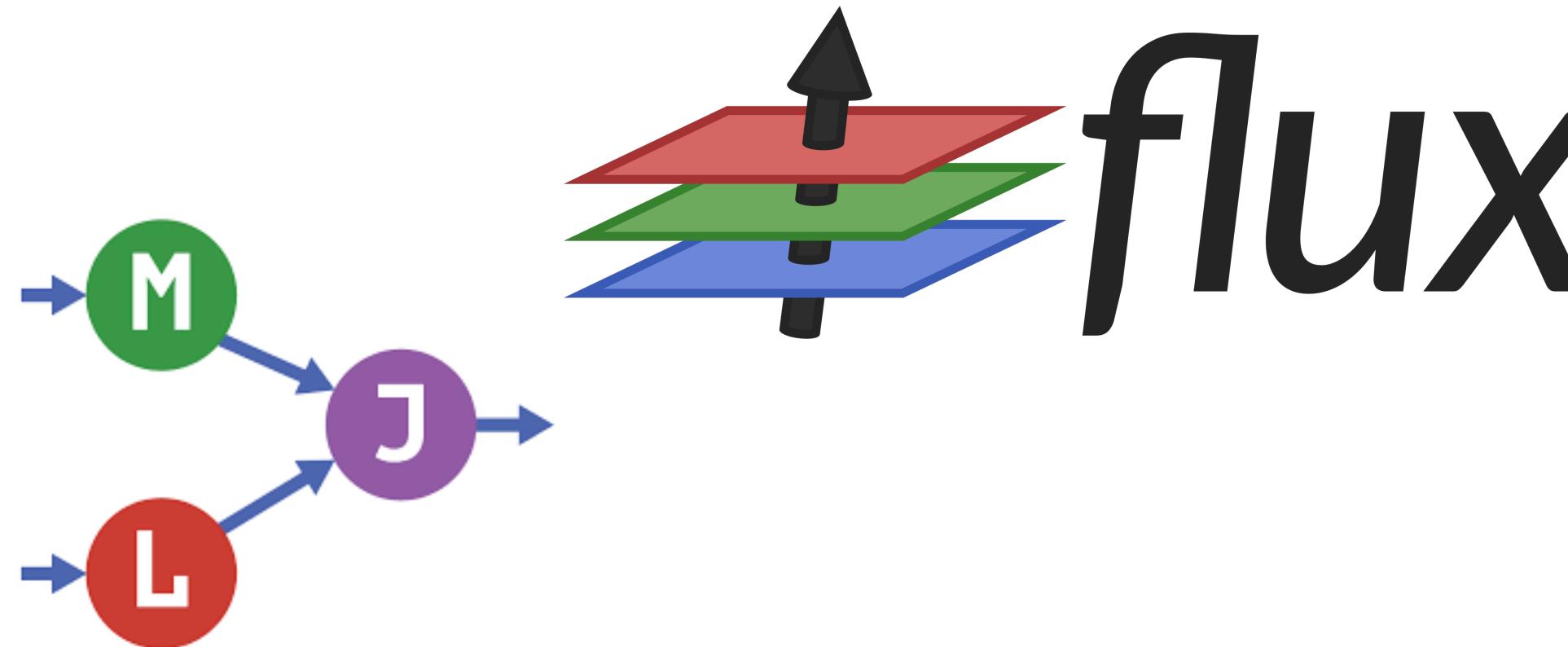
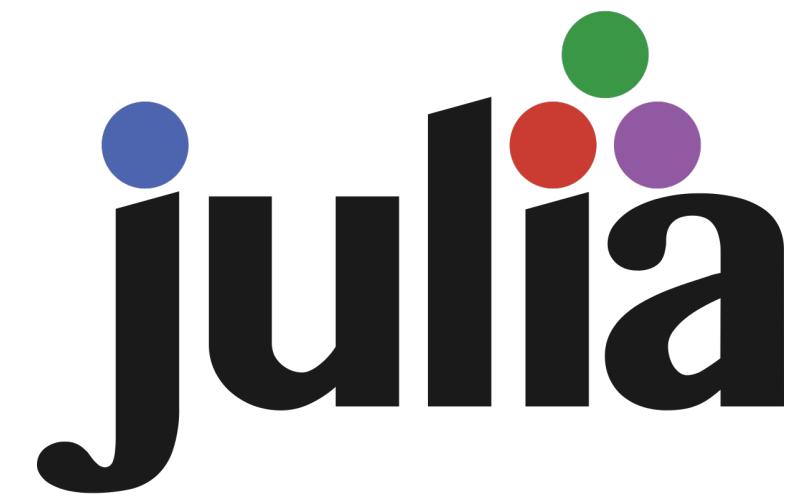
- Langsam - Interpretierte Sprache
- Keine Typsicherheit

# Trainiere eines Neuronale Netzes

## Technologischer Überblick



Tensorflow



### Julia

- **Vorteile:**

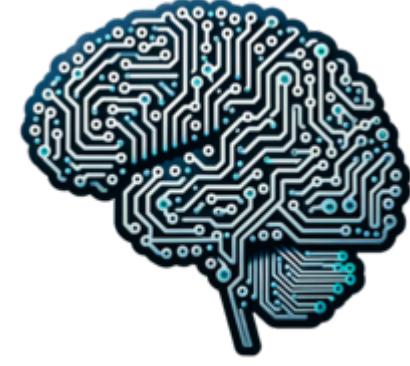
- Hohe Verarbeitung Geschwindigkeit
- Dynamisches Ökosystem

- **Nachteile:**

- Relativ junge Programmiersprache
- Kein großes Ökosystem

# Trainiere eines Neuronale Netzes

## Trainieren mit MNIST-Datenbank



0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3  
4 4 4 4 4 4 4 4 4 4 4 4 4 4 4  
5 5 5 5 5 5 5 5 5 5 5 5 5 5 5  
6 6 6 6 6 6 6 6 6 6 6 6 6 6 6  
7 7 7 7 7 7 7 7 7 7 7 7 7 7 7  
8 8 8 8 8 8 8 8 8 8 8 8 8 8 8  
9 9 9 9 9 9 9 9 9 9 9 9 9 9 9

### MNIST-Datennbank

- Öffentliche Datenbank für handgeschriebene Ziffern
- Trainingssatz 60.000 Beispiele
- Testdaten 10.000 Beispiele
- Typisches Einsteigerbeispiel für maschinelles Lernen
- Paket für Julia verfügbar

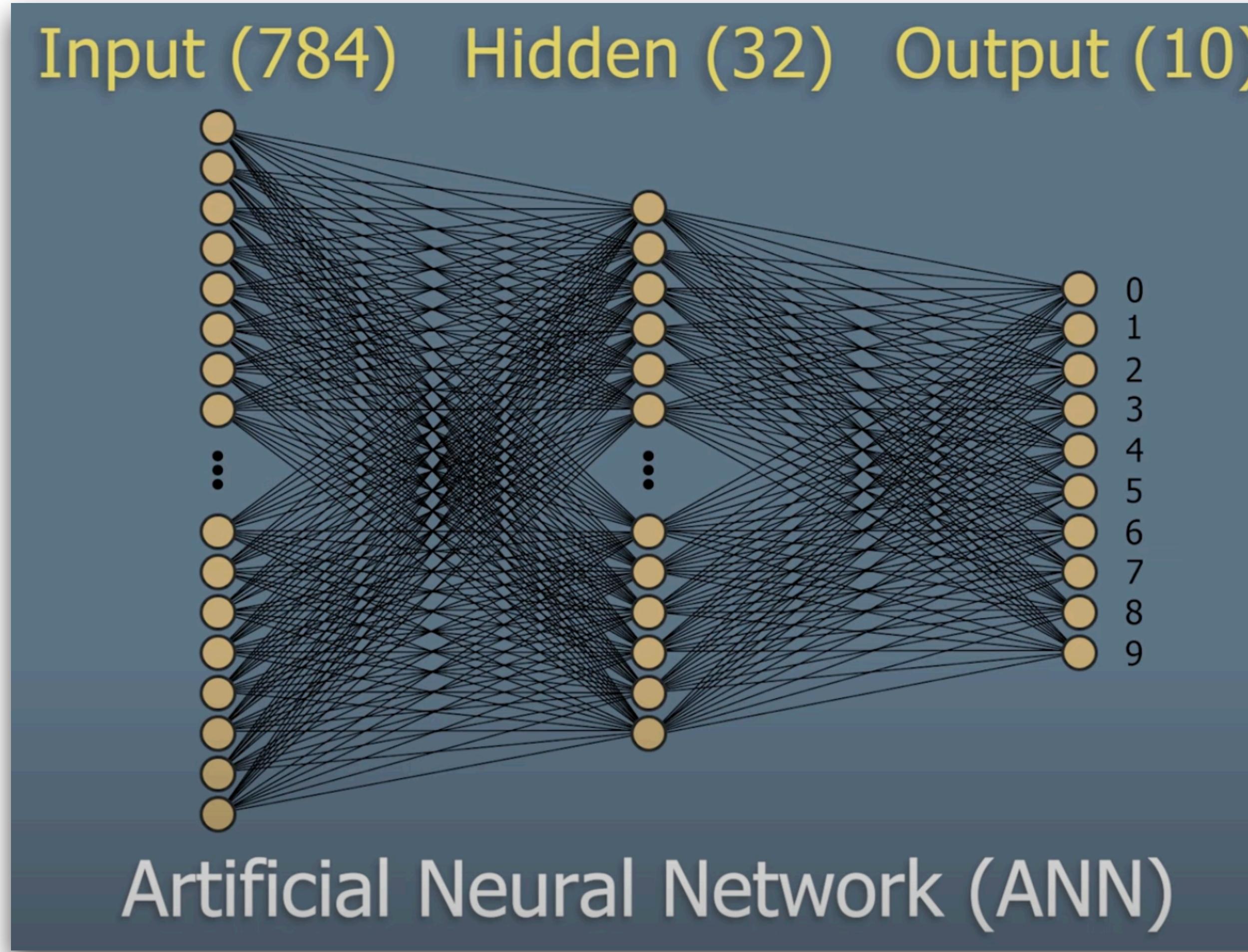
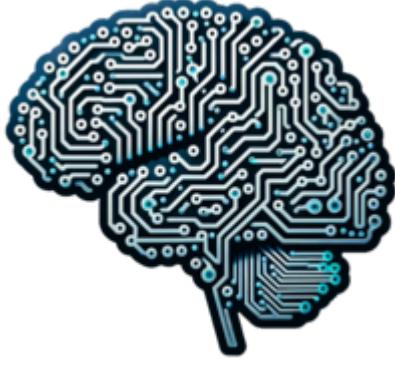


# **Code - Beispiel mit Julia - Flux**

**Dr. Frank Zimmer**

# Trainiere eines Neuronale Netzes

## Trainieren mit MNIST-Datenbank

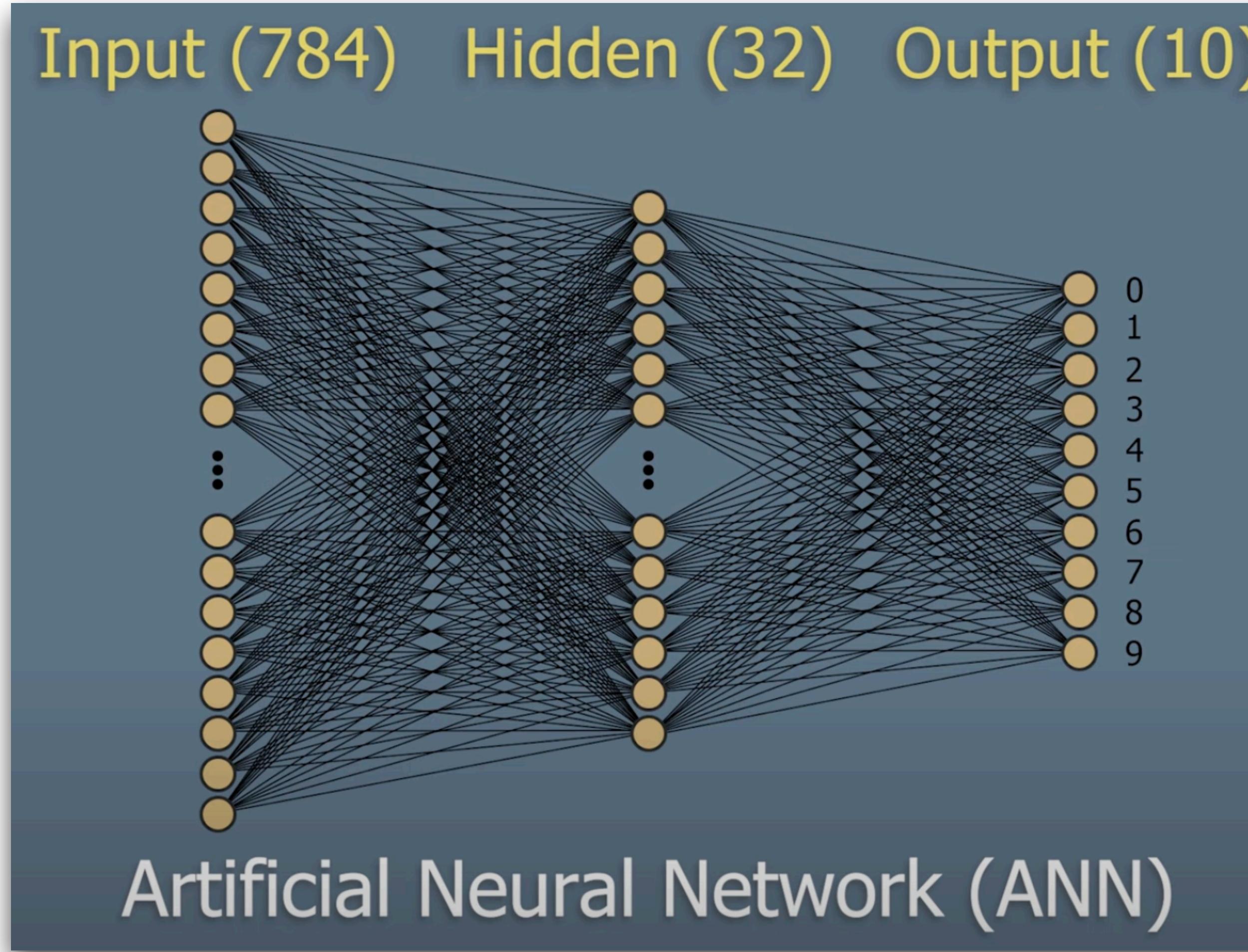
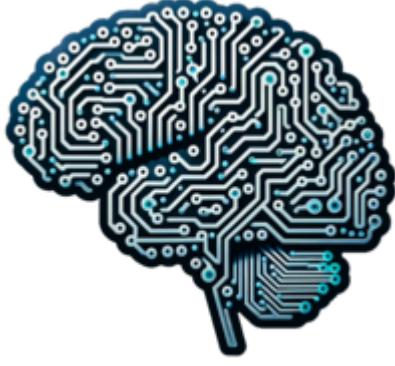


Artificial Neural Network (ANN)

```
model = Chain(  
    Dense(28 * 28, 32, relu),  
    Dense(32, 10),  
    softmax  
)  
  
# Eingangsneuronen: 28*28=784  
# Hidden-Layer-Neuronen: 32  
# Ausgangsneuronen: 10
```

# Trainiere eines Neuronale Netzes

## Trainieren mit MNIST-Datenbank



```
model = Chain(  
    Dense(28 * 28, 32, relu),  
    # 25120 Parameter  
    Dense(32, 10),  
    # 330 Parameter  
    softmax  
)  
  
# Parameter 25450 = 25120 + 330  
  
# 25120 = 784*32+32  
# = #Wichtungen + #Biases  
#  
# 330 = 32*10+10
```

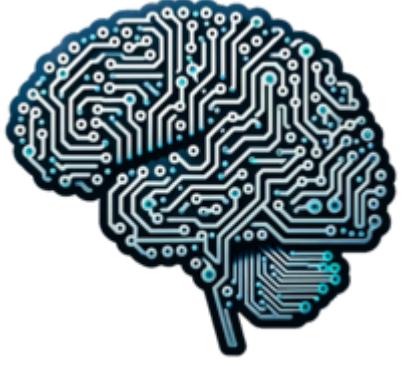


# KI - Anwendungen Kommerzieller Ansatz

**Dr. Frank Zimmer**

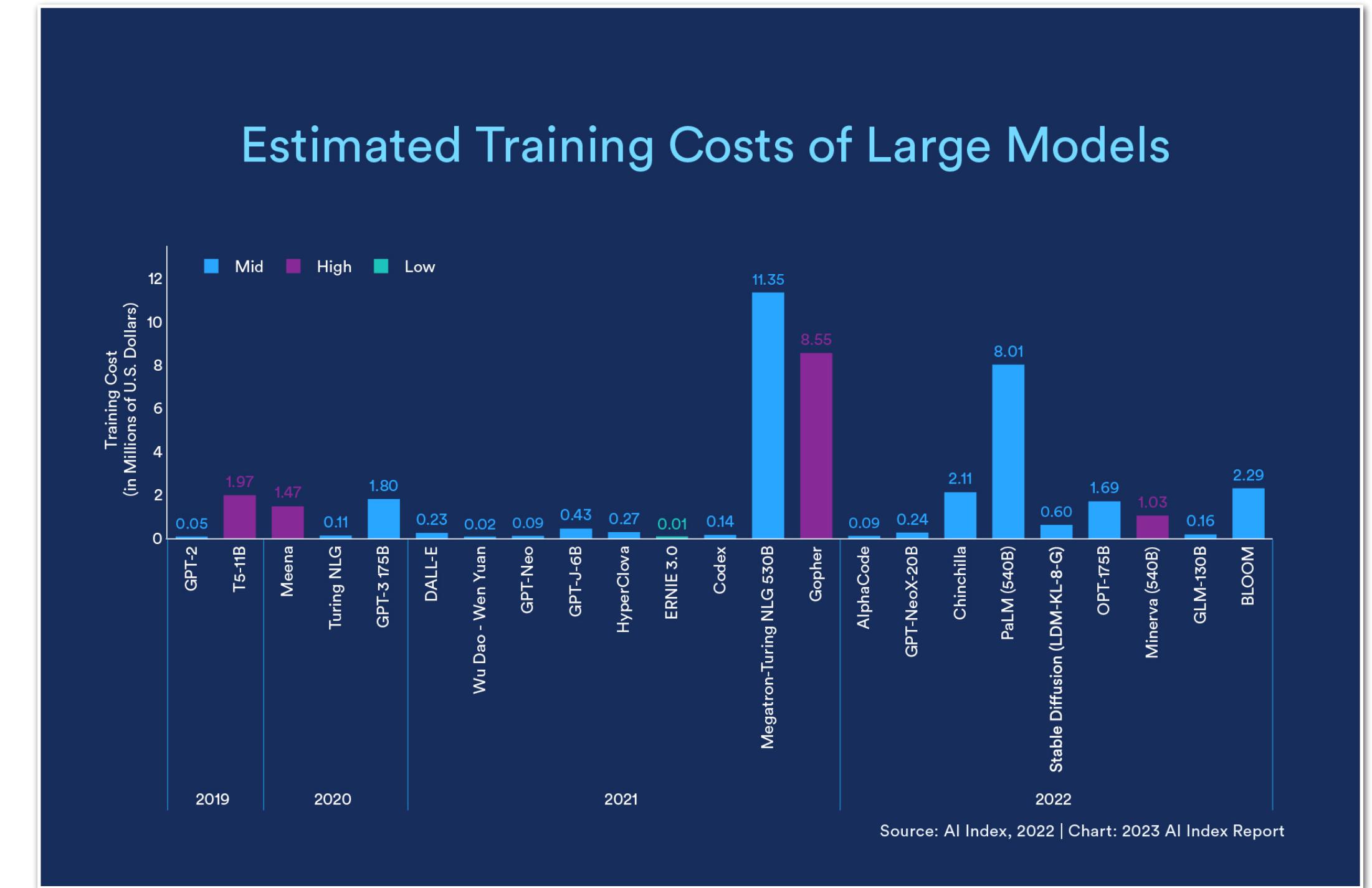
# Large Language Models

## Was sind große Sprachmodelle?



### Eigenschaften

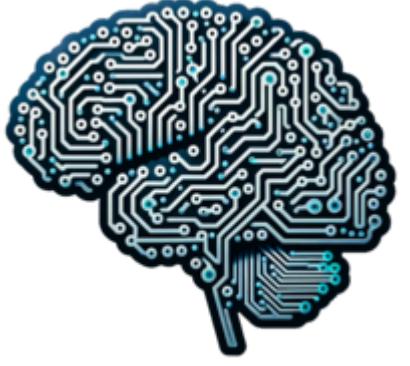
- Sind Neuronale Netze
- Zeichnen sich durch Eigenschaft aus, dass sie natürliche Sprache **klassifizieren und generieren** können
- Erlangen diese Fähigkeit durch Erlernen von „statistische“ Zusammenhängen in **riesigen Textdatenmengen**
- Benutzen Self-supervised oder Semi-Supervised Learning



Trainieren kostet viel Geld

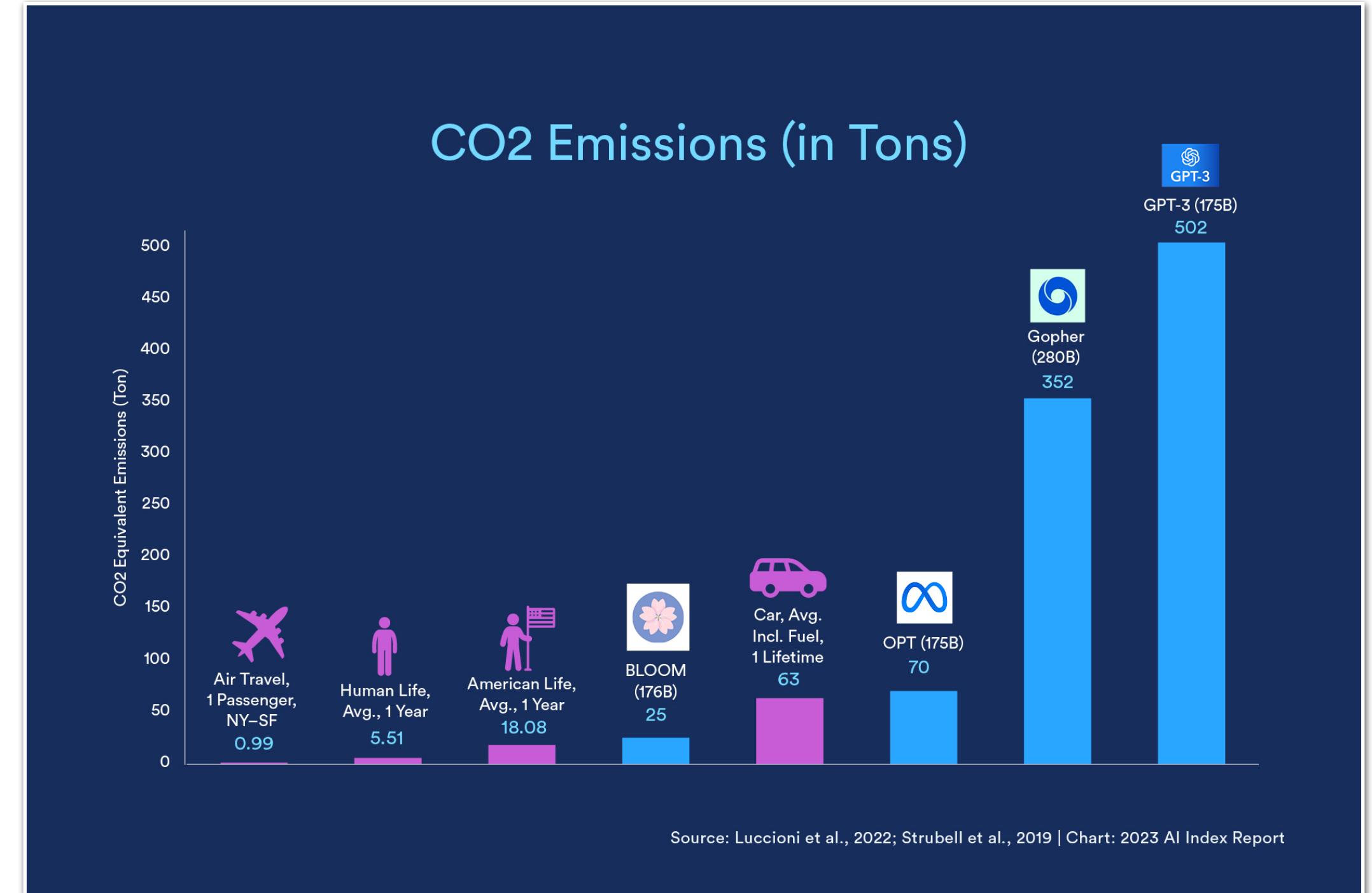
# Large Language Models

## Was sind große Sprachmodelle?



### Eigenschaften

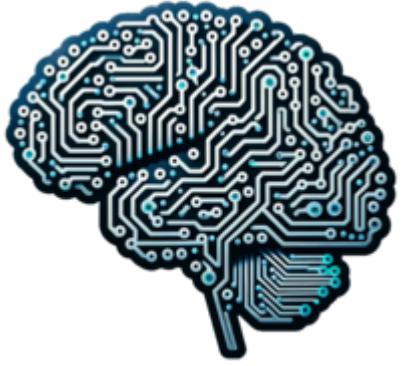
- Sind Neuronale Netze
- Zeichnen sich durch Eigenschaft aus, dass sie natürliche Sprache **klassifizieren und generieren** können
- Erlangen diese Fähigkeit durch Erlernen von „statistische“ Zusammenhängen in **riesigen Textdatenmengen**
- Benutzen Self-supervised oder Semi-Supervised Learning



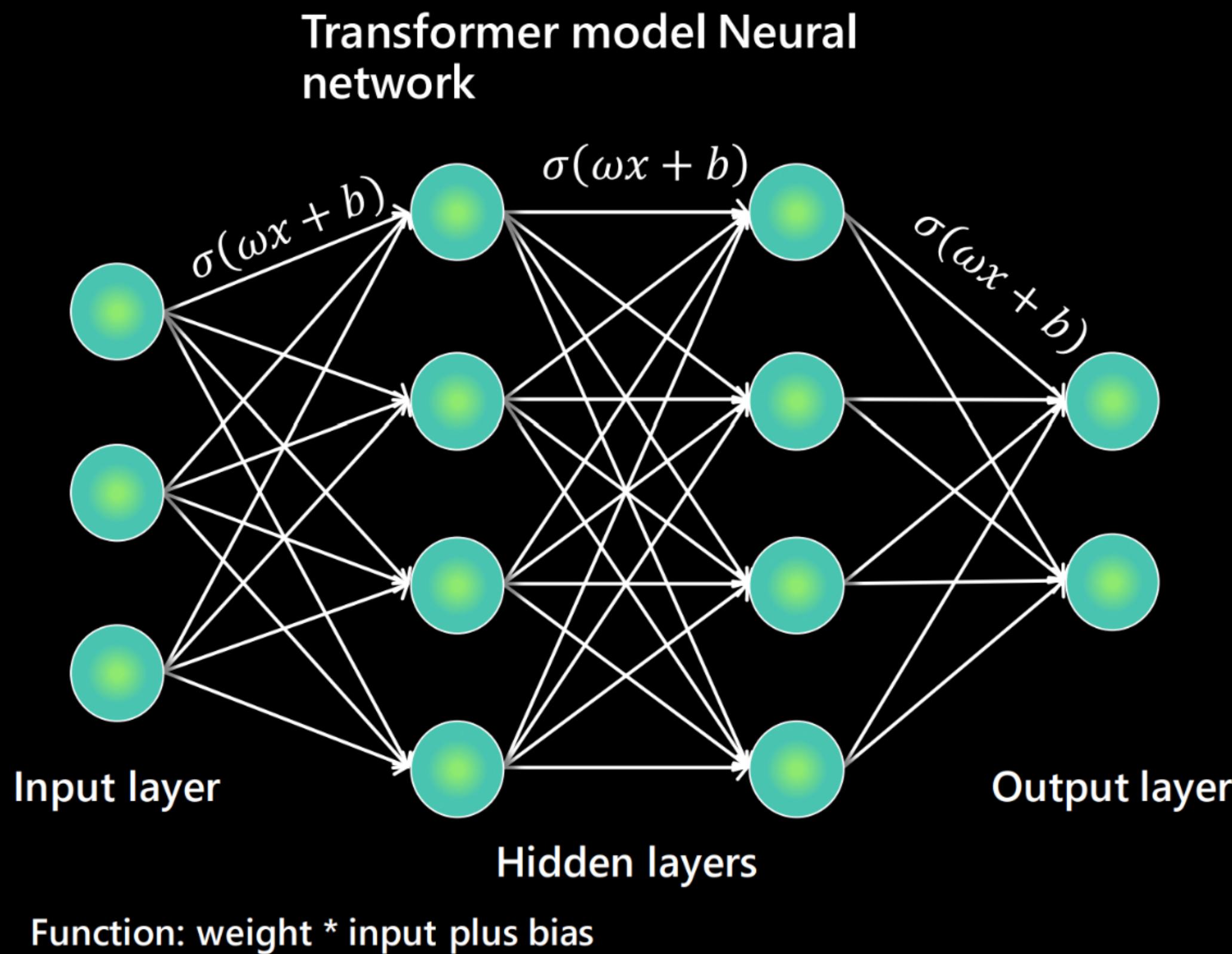
Trainieren belastet die Umwelt

# Large Language Models

Wie groß sind die Sprachmodelle?



## How large are they?



BERT Large - 2018

**345M**

GPT2 - 2019

**1.5B**

GPT3 - 2020

**175B**

Turing Megatron NLG

2021

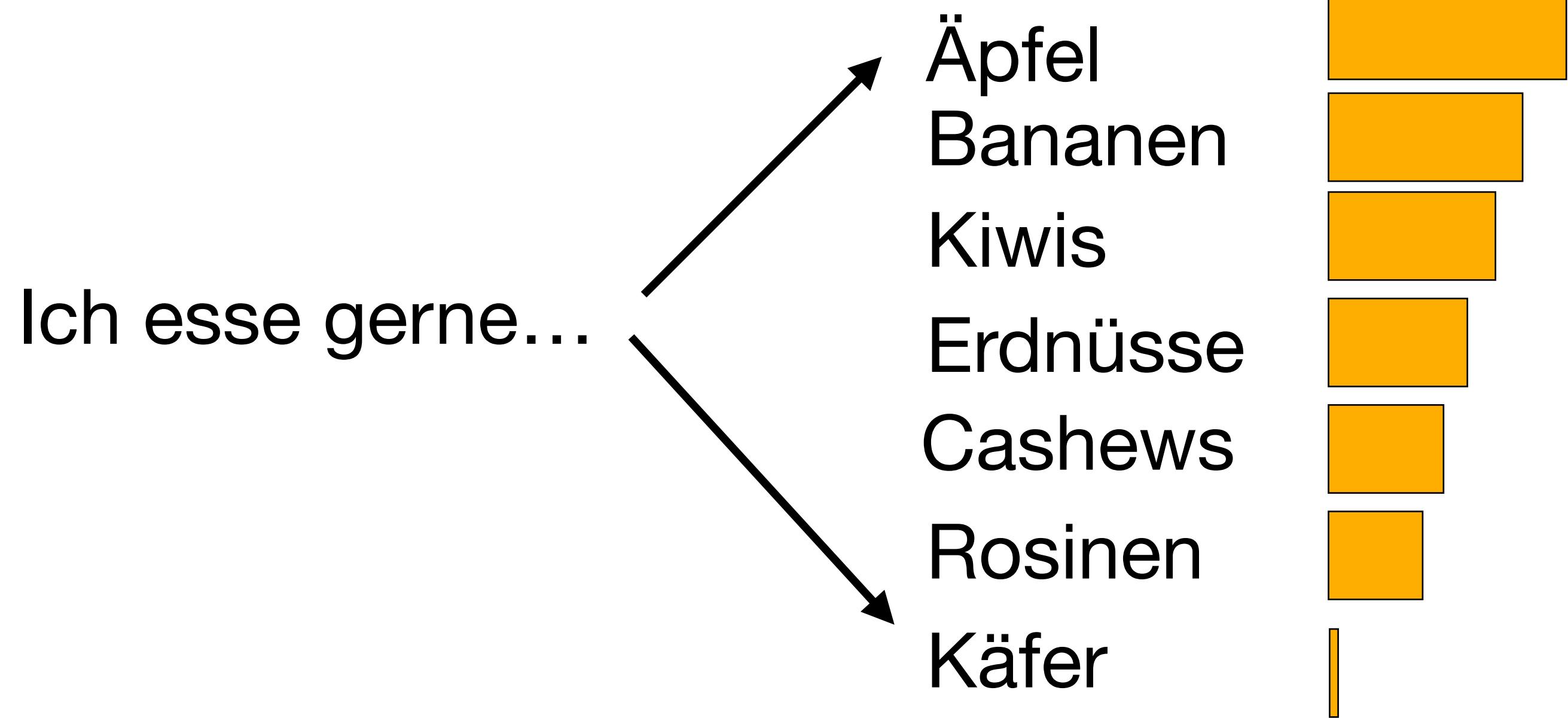
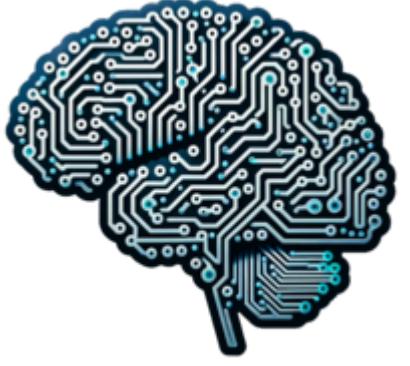
**530B**

GPT4 – 2023

**1.4T** (estimated)

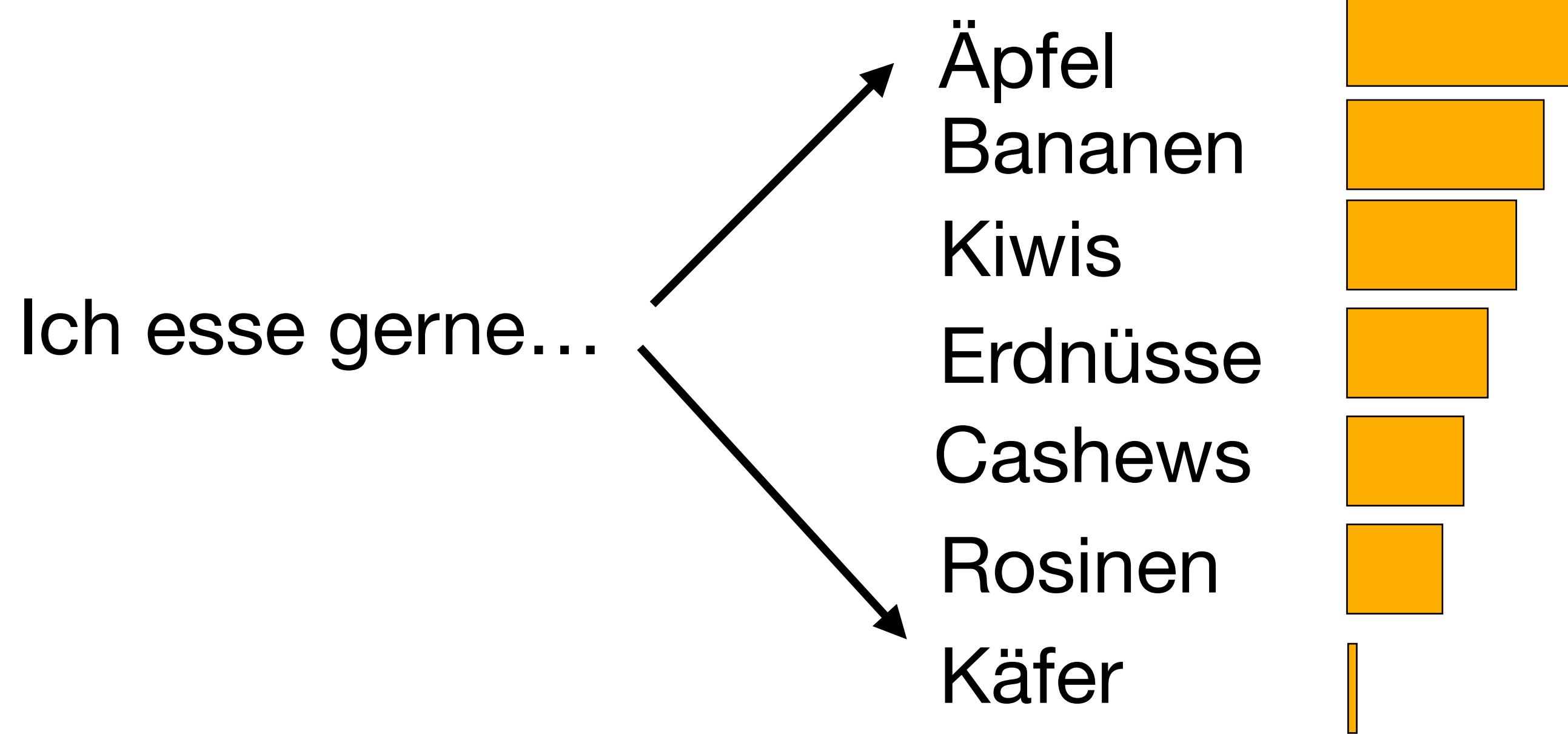
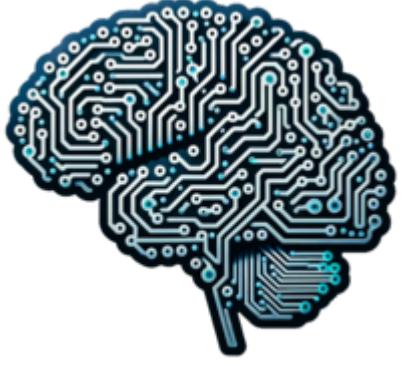
# Large Language Models

## Wie funktionieren Sprachmodelle?



# Large Language Models

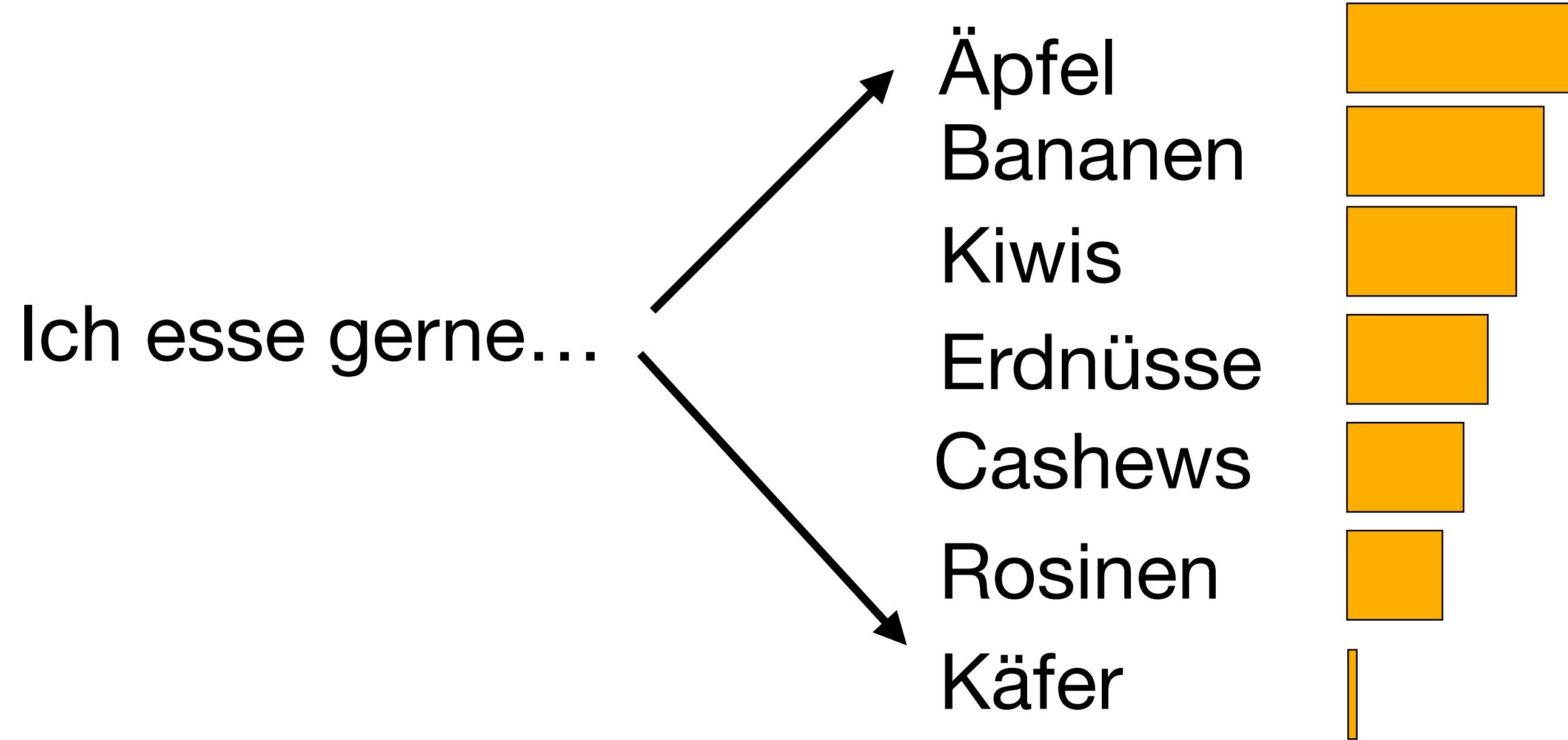
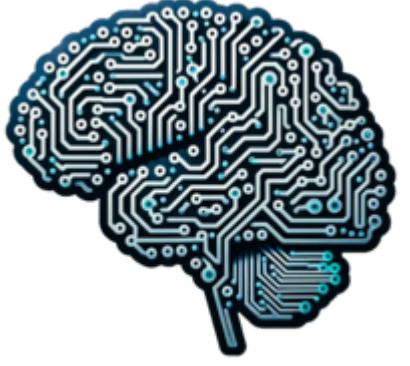
## Wie funktionieren Sprachmodelle?



- LLMs verarbeiten Eingaben und bilden dadurch einen **Kontext** auf.

# Large Language Models

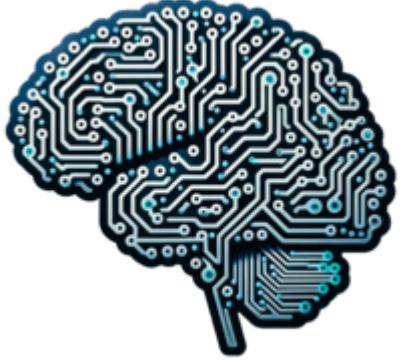
## Wie funktionieren Sprachmodelle?



- LLMs verarbeiten Eingaben und bilden dadurch einen **Kontext** auf.
- LLMs sagen das nächste „Wort“ (Token) bei basierend auf dem Kontext vorher

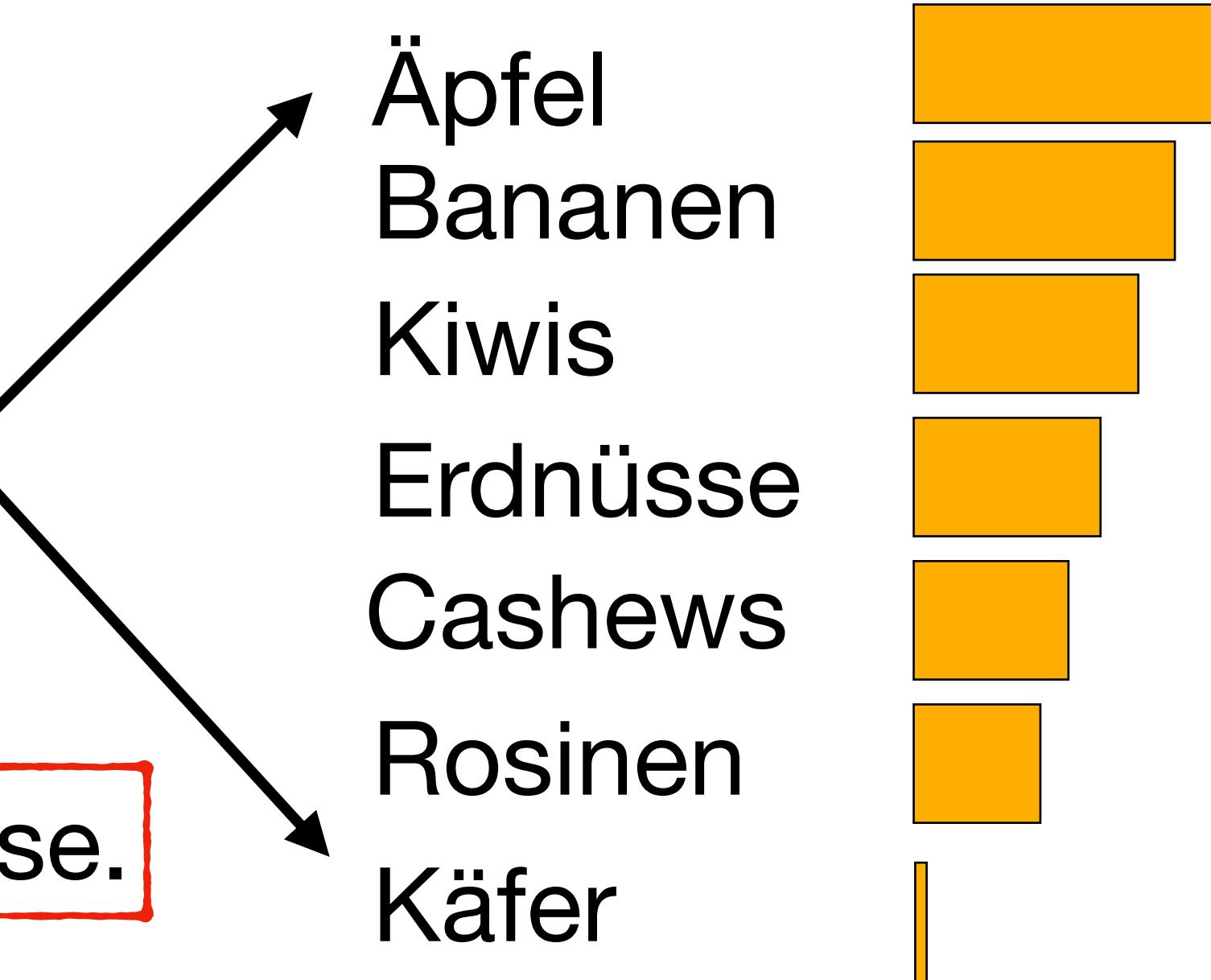
# Large Language Models

## Wie funktionieren Sprachmodelle?



Ich esse gerne...

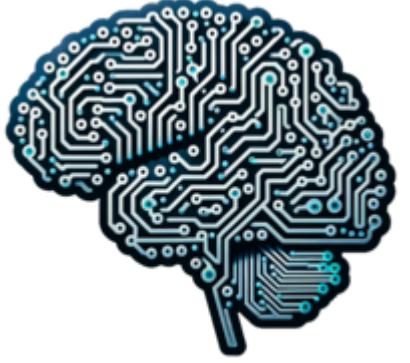
Person mag lieber Nüsse.



- LLMs verarbeiten Eingaben und bilden dadurch einen **Kontext** auf.
- LLMs sagen das nächste „Wort“ (Token) bei basierend auf dem Kontext vorher

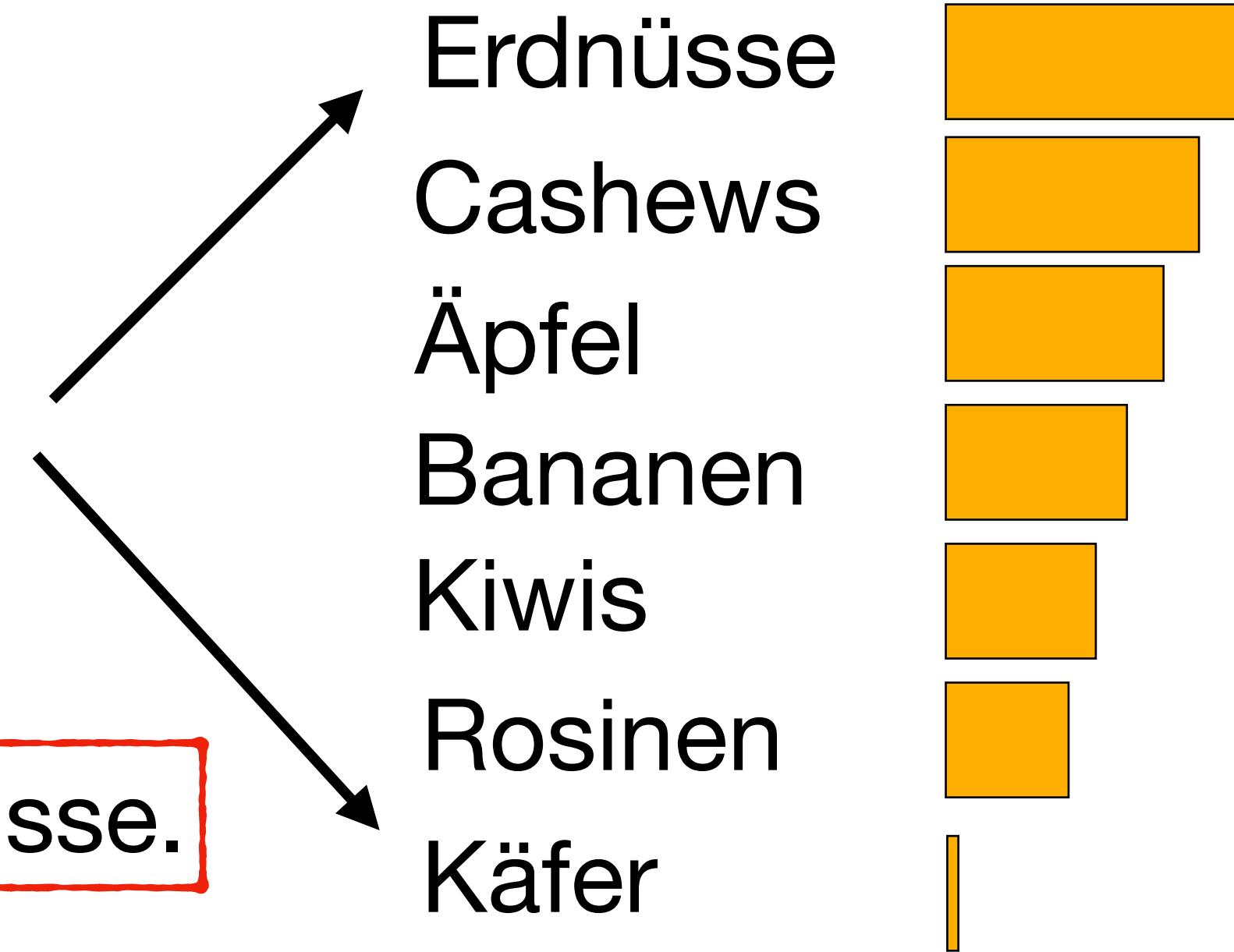
# Large Language Models

## Wie funktionieren Sprachmodelle?



Ich esse gerne...

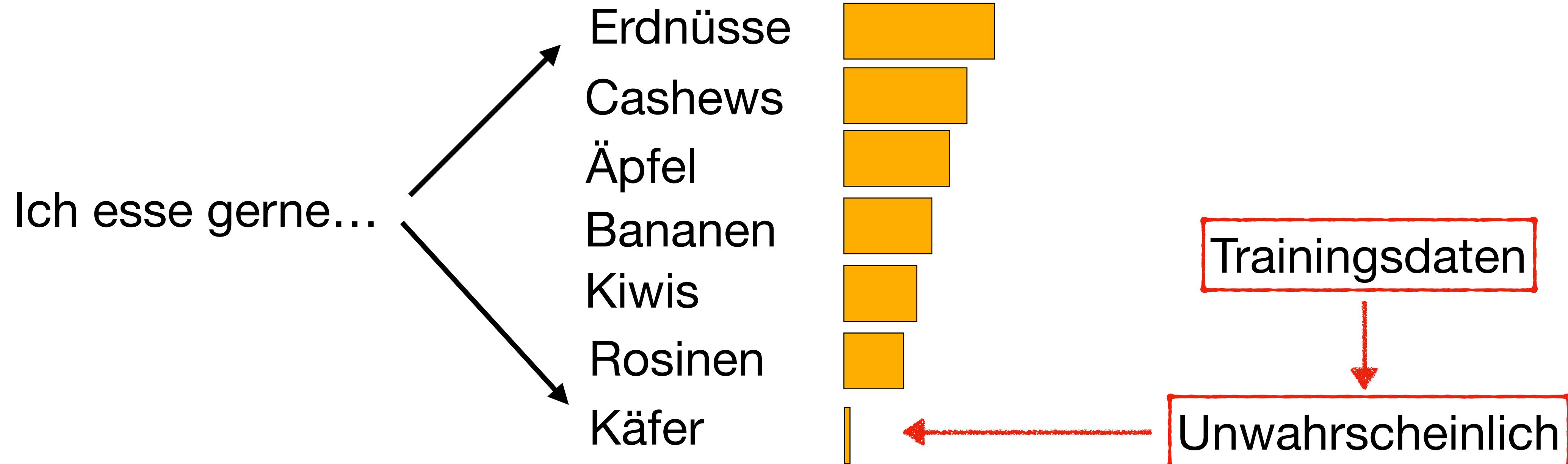
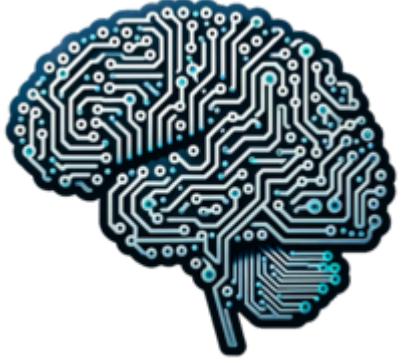
Person mag lieber Nüsse.



- LLMs verarbeiten Eingaben und bilden dadurch einen **Kontext** auf.
- LLMs sagen das nächste „Wort“ (Token) bei basierend auf dem Kontext vorher

# Large Language Models

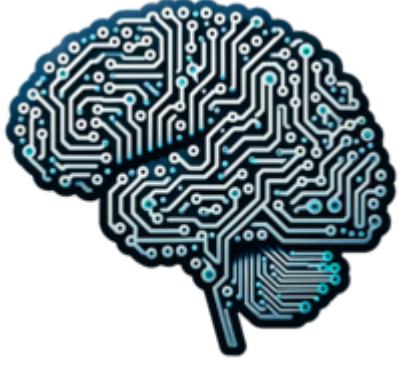
## Wie funktionieren Sprachmodelle?



- LLMs verarbeiten Eingaben und bilden dadurch einen **Kontext** auf.
- LLMs sagen das nächste „Wort“ (Token) bei basierend auf dem Kontext vorher
- LLMs nutzen dabei ihre trainierten Wichtungen bei ihren Antworten.

# Kommerzielle LLMs

## Welche Anbieter gibt es?

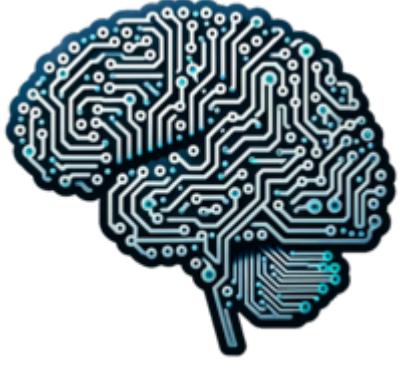


The screenshot shows the OpenAI website with the ChatGPT interface. On the left, there's a large yellow "ChatGPT" logo with a yellow dot. Below it, the text "Get instant answers, find creative inspiration, learn something new." is displayed. On the right, there's a mobile phone interface showing a conversation between "Me" and "ChatGPT". "Me" asks for brunch suggestions, and "ChatGPT" responds with a list of 10 items. Above the phone, there's a header with the OpenAI logo and links for Research, API, ChatGPT, Safety, Company, Search, Log in, and Try ChatGPT. To the right of the phone, there's a desktop interface for "ChatGPT 3.5" where "Me" asks for feedback on a short story. "ChatGPT" provides a detailed analysis and three points for improvement. At the bottom, there's a message input field and a note about ChatGPT making mistakes.

<https://openai.com/chatgpt>

# Kommerzielle LLMs

## Welche **Anbieter** gibt es?

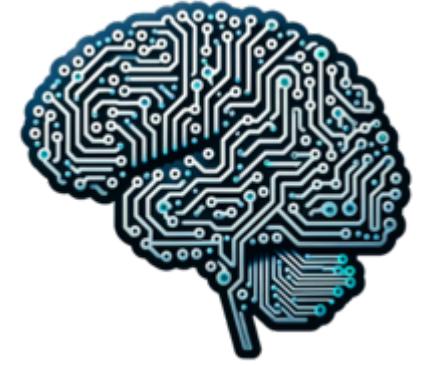


The screenshot shows the homepage of the Google DeepMind Gemini website. At the top, there is a navigation bar with links: Google DeepMind, About, Technologies (which is underlined), Impact, and Discover. The main feature is a large, bold text "Welcome to the Gemini era" in white and light orange. Below this, there is a button labeled "Chat with Gemini →". At the bottom, there is a footer bar with links: The Gemini era (underlined), Gemini 1.5, Gemini 1.0, Safety, Gemini Apps, and Build with Gemini.

<https://gemini.google.com>

# Kommerzielle LLMs

## Welche Anbieter gibt es?



\* Claude

Build With Claude ↗

### Think visually with Claude

Get started with your email below

Email address

Continue with email

OR

Continue with Google

By continuing, you agree to Anthropic's [Consumer Terms](#) and [Acceptable Use Policy](#), and acknowledge their [Privacy Policy](#).

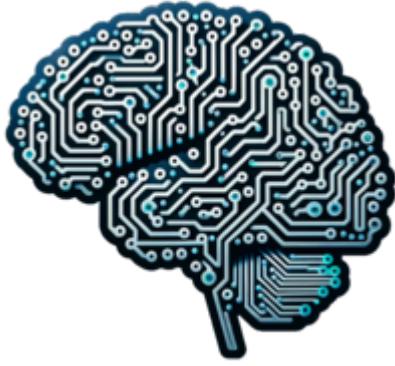
Claude is a next generation AI assistant built for work and trained to be safe, accurate, and secure.

BY ANTHROPIC

<https://claude.ai>

# Kommerzielle LLMs

## Welche Anbieter gibt es?



### Microsoft Copilot

UNTERNEHMEN

#### Microsoft Copilot für Microsoft 365

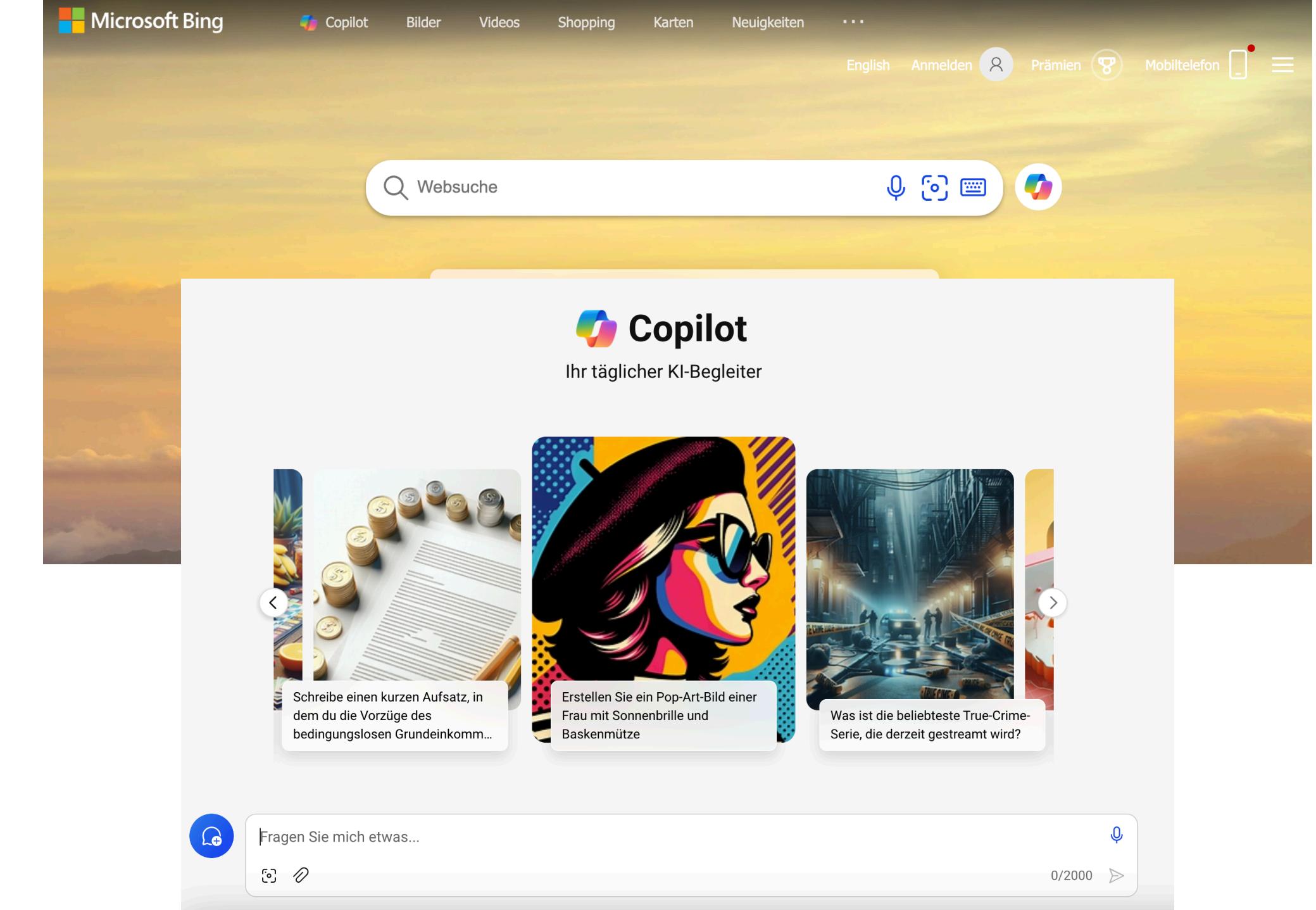
Sparen Sie Zeit und konzentrieren Sie sich auf das, was wirklich zählt – mit KI in Microsoft 365 für Unternehmen.\*

[Preise anzeigen](#) [Kaufoptionen](#)

\* Erfordert ein Microsoft 365 Business Standard- oder Business Premium-Abonnement.



### Integration in Office 365



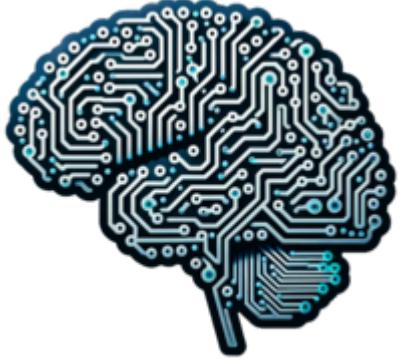
The screenshot shows the Microsoft Bing search interface. At the top, there's a navigation bar with links for Copilot, Bilder, Videos, Shopping, Karten, Neuigkeiten, and more. Below the bar is a search bar with the placeholder "Websuche". To the right of the search bar are icons for microphone, camera, and keyboard. The main content area features a large image of a sunset. Overlaid on this are several cards from the Copilot feature. One card says "Copilot Ihr täglicher KI-Begleiter". Another card shows a woman wearing sunglasses and asks: "Schreibe einen kurzen Aufsatz, in dem du die Vorzüge des bedingungslosen Grundeinkomm...". A third card shows a cityscape and asks: "Erstellen Sie ein Pop-Art-Bild einer Frau mit Sonnenbrille und Baskenmütze". A fourth card asks: "Was ist die beliebteste True-Crime-Serie, die derzeit gestreamt wird?". At the bottom, there's a chat input field with the placeholder "Fragen Sie mich etwas..." and a character icon.

### Integration in Bing

<https://www.bing.com/chat>

# Kommerzielle LLMs

## Programmieren mit KI



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



<https://www.tabnine.com/>



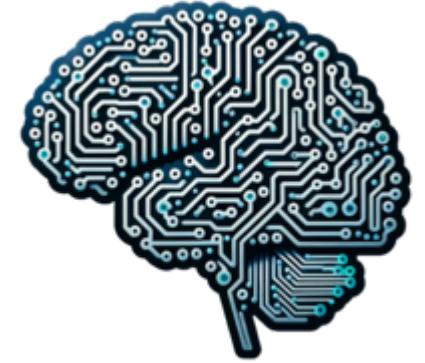
# KI - Anwendungen

## Open source Ansatz

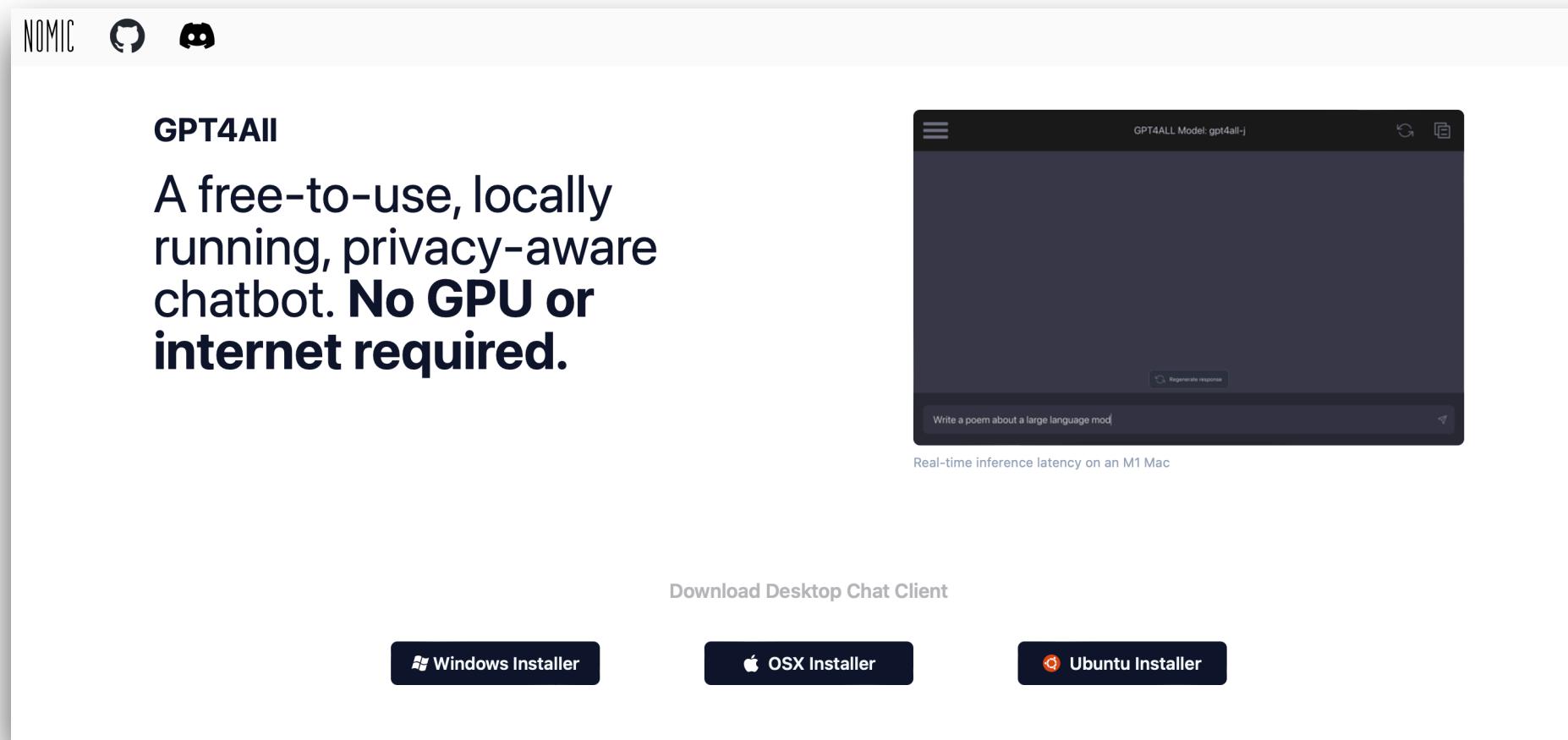
**Dr. Frank Zimmer**

# Open Source Ansatz

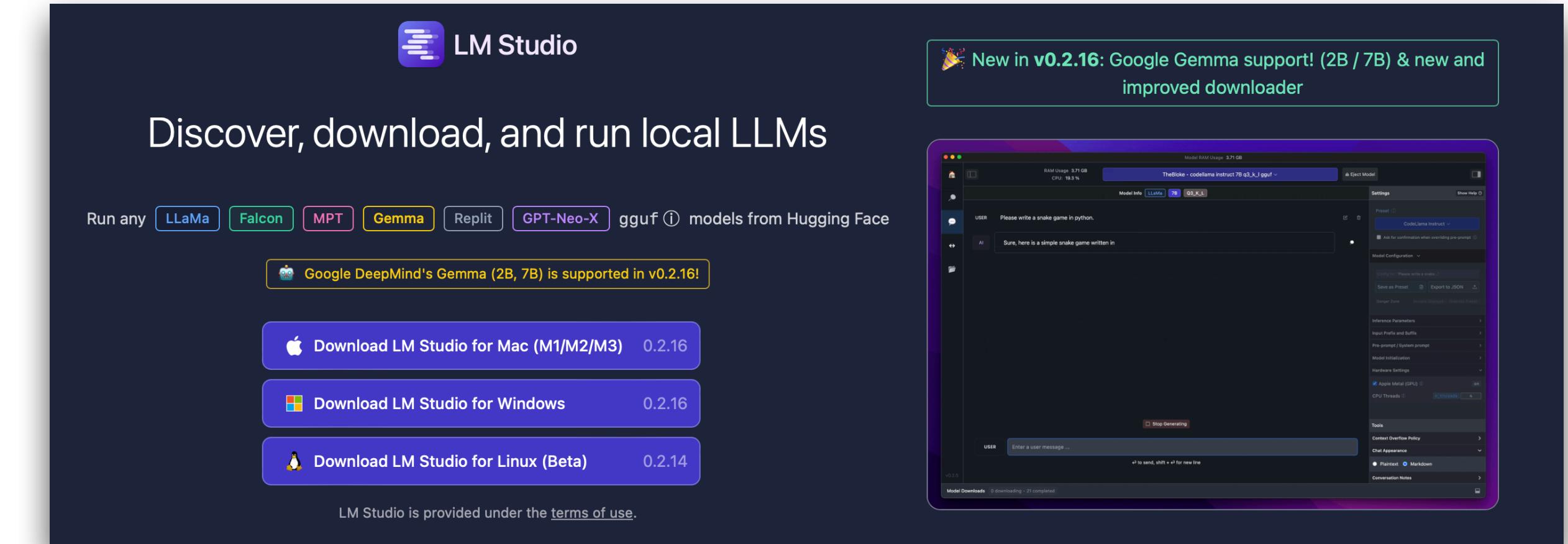
Wie kann man einfach **LLMs lokal** auf dem **Computer laufen lassen?**



<https://gpt4all.io/>



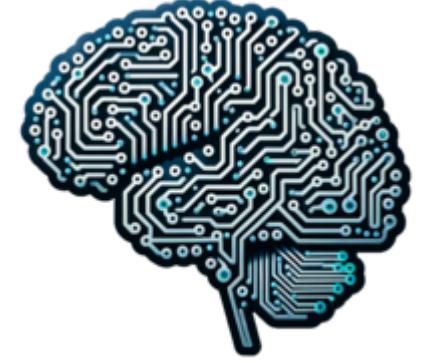
<https://lmstudio.ai/>



- **Unterstützte Betriebssysteme:** Windows, macOS, Linux
- **Ausreichend RAM (16GB+)** notwendig
- **Ausreichend Speicherplatz** für Modelle notwendig (mehrere GB pro Model)

# Open Source Ansatz

Wie kann man einfach **LLMs lokal** auf dem **Computer laufen lassen?**



<https://gpt4all.io/>

The screenshot shows the GPT4ALL website. On the left, there's a section for "GPT4ALL" with text: "A free-to-use, locally running, privacy-aware chatbot. **No GPU or internet required.**" Below this is a screenshot of the desktop chat client interface, which is a dark-themed window with a text input field and a "Request response" button. At the bottom, there are download links for "Windows Installer", "OSX Installer", and "Ubuntu Installer".

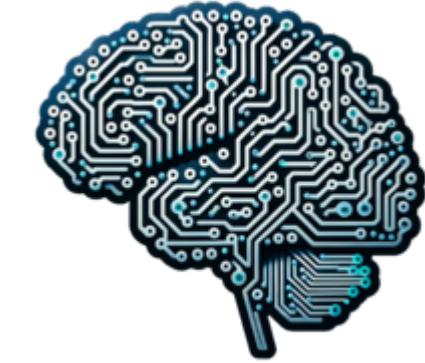
<https://lmstudio.ai/>

The screenshot shows the LM Studio website and its desktop application. The website has a dark header with the LM Studio logo and a banner about Google DeepMind's Gemma support. It features a main heading "Discover, download, and run local LLMs" and a section for "Run any" models like LLaMa, Falcon, MPT, Gemma, Replit, GPT-Neo-X, and gguf. Below this are download links for "Mac (M1/M2/M3)", "Windows", and "Linux (Beta)". The desktop application window on the right shows a user interface with tabs for "Model Info", "LM", and "OKX". It displays a conversation between a user asking for a snake game in Python and an AI responding with the code. The application has various settings and toolbars at the top and bottom.

- **Vorteile:** keine Programmierkenntnisse für Nutzung notwendig
- **Nachteile:** Bindung an Beschränkung durch vortrainierte Modelle

# Open Source Ansatz

Wie kann man einfach **LLMs lokal** auf dem **Computer laufen lassen?**



<https://github.com/imartinez/privateGPT>

PRIVATE GPT

Mode

Query Docs

Search in Docs LLM Chat

Upload a File

Ingested Files

File name

bitcoin.pdf

Chapter1-DigitalCurrency\_DavidLee.p

LLM: local

According to David Lee's document, what is the major benefit of Bitcoin, and where in the document is it explained?

The major benefit of Bitcoin, as explained in the Handbook of Digital Currency, is its ability to enable fast transactions at low costs. This is mentioned in section 1.5.1 under the subheading "Freedom of payments".

Type a message... Submit

Retry Undo Clear

- **Vorteile:**

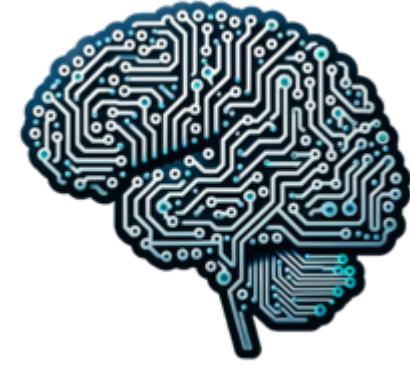
- Erlaubt Verarbeitung von eigenen Dokumenten
- Entwicklung noch sehr aktiv

- **Nachteile:**

- Aufsetzen benötigt Erfahrung beim Umgang mit Python
- Entwicklung noch sehr aktiv

# Hugging Face

## Online Community für KI-Projekte



 Hugging Face

[Models](#) [Datasets](#) [Spaces](#) [Posts](#) [Docs](#) [Pricing](#) [Log In](#) [Sign Up](#)

**NEW** Create Assistants in HuggingChat



**The AI community building the future.**

The platform where the machine learning community collaborates on models, datasets, and applications.

Tasks   Libraries   Datasets   Languages   Licenses   Other

Filter Tasks by name

Multimodal

- Text-to-Image
- Image-to-Text
- Text-to-Video
- Visual Question Answering
- Document Question Answering
- Graph Machine Learning

Computer Vision

- Depth Estimation
- Image Classification
- Object Detection
- Image Segmentation
- Image-to-Image
- Unconditional Image Generation
- Video Classification
- Zero-Shot Image Classification

Natural Language Processing

- Text Classification
- Token Classification
- Table Question Answering
- Question Answering
- Zero-Shot Classification
- Translation
- Summarization
- Conversational
- Text Generation
- Text2Text Generation
- Sentence Similarity

Audio

- Text-to-Speech
- Automatic Speech Recognition
- Audio-to-Audio
- Audio Classification
- Voice Activity Detection

Tabular

- Tabular Classification
- Tabular Regression

Reinforcement Learning

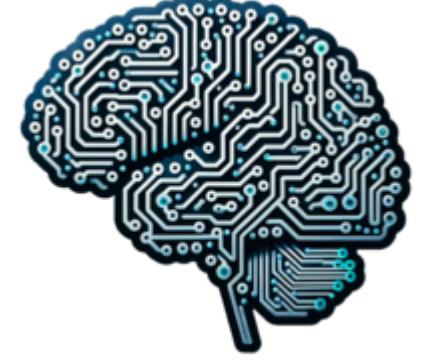
- Reinforcement Learning
- Robotics

Models 469,541 [Filter by name](#)

- meta-llama/Llama-2-70b**  
Text Generation • Updated 4 days ago • ± 25.2k • ❤ 64
- stabilityai/stable-diffusion-xl-base-0.9**  
Updated 6 days ago • ± 2.01k • ❤ 393
- openchat/openchat**  
Text Generation • Updated 2 days ago • ± 1.3k • ❤ 136
- llyasviel/ControlNet-v1-1**  
Updated Apr 26 • ❤ 1.87k
- cerspense/zeroscope\_v2\_XL**  
Updated 3 days ago • ± 2.66k • ❤ 334
- meta-llama/Llama-2-13b**  
Text Generation • Updated 4 days ago • ± 328 • ❤ 64
- tiiuae/falcon-40b-instruct**  
Text Generation • Updated 27 days ago • ± 288k • ❤ 899
- WizardLM/WizardCoder-15B-V1.0**  
Text Generation • Updated 3 days ago • ± 12.5k • ❤ 332
- CompVis/stable-diffusion-v1-4**  
Text-to-Image • Updated about 17 hours ago • ± 448k • ❤ 5.72k
- stabilityai/stable-diffusion-2-1**  
Text-to-Image • Updated about 17 hours ago • ± 782k • ❤ 2.81k
- Salesforce/xgen-7b-8k-inst**  
Text Generation • Updated 4 days ago • ± 6.18k • ❤ 57

# Hugging Face

Online Community für KI-Projekte



A screenshot of the Hugging Face website homepage. The top navigation bar includes links for 'Models', 'Datasets', 'Spaces', 'Docs', 'Pricing', 'Log In', and 'Sign Up'. A search bar is located at the top left. The main content area features a dark background with white text. It includes a 'Create Assistants in HuggingChat' button, a smiling emoji, and the tagline 'The AI community building the future.' Below this, there's a section about the platform being the place where the machine learning community collaborates on models, datasets, and applications. On the right side, there's a list of various AI models and their details, such as 'meta-llama/Llama-2-70b', 'stabilityai/stable-diffusion-xl-base-0.9', and 'meta-llama/Llama-2-13b'. The entire screenshot is framed by a red border.

<https://huggingface.co/>

Interessante Beispiele unter Spaces

<https://huggingface.co/spaces/chaenig/Webcam-Object-Recognition-Yolo>

<https://huggingface.co/spaces/jbilcke-hf/ai-comic-factory>

<https://huggingface.co/spaces/hilamanor/audioEditing>

<https://huggingface.co/spaces/stabilityai/TripoSR>

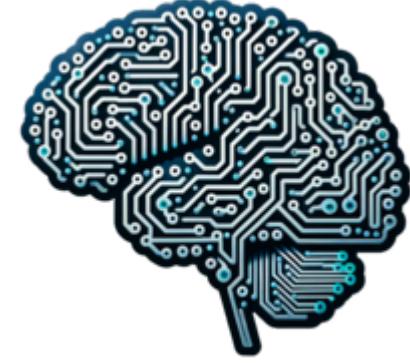


# Zukunft mit künstlicher Intelligenz

**Dr. Frank Zimmer**

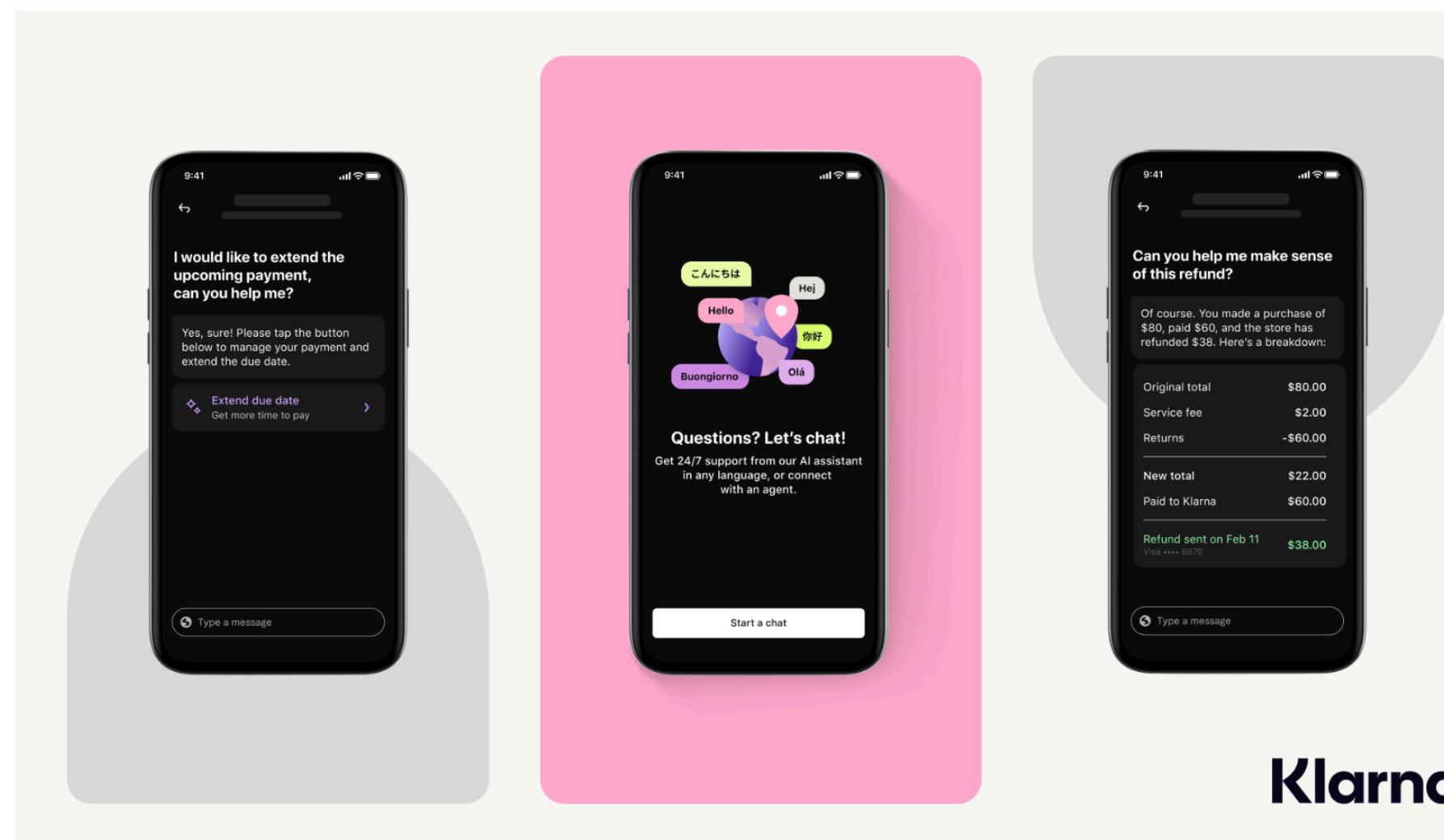
# Auswirkungen

## Wie wird KI die **Arbeitswelt** verändern?



General News · 27 Feb 2024

### Klarna AI assistant handles two-thirds of customer service chats in its first month



**Klarna**

#### Vorteile:

- Langweilige Arbeiten können abgegeben werden
- Kleinere Unternehmen können ohne Aufwand größere Profite erwirtschaften.

#### Nachteile:

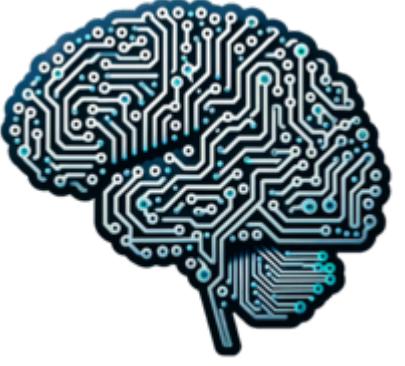
- Jobverlust für viele Menschen
- Unsichere Zukunft → Wer ist betroffen?

New York, NY – February 27, 2024 – [Klarna](#) today announced its AI assistant powered by OpenAI. Now live globally for 1 month, the numbers speak for themselves:

- The AI assistant has had **2.3 million conversations, two-thirds of Klarna's customer service chats**
- **It is doing the equivalent work of 700 full-time agents**
- It is on par with human agents in regard to customer satisfaction score
- It is more accurate in errand resolution, **leading to a 25% drop in repeat inquiries**
- Customers now **resolve their errands in less than 2 mins compared to 11 mins previously**
- It's available in **23 markets, 24/7** and communicates in more than **35 languages**
- **It's estimated to drive a \$40 million USD in profit improvement to Klarna in 2024**

# Auswirkungen

## Wie wird KI die **Arbeitswelt** verändern?



GR1



H1

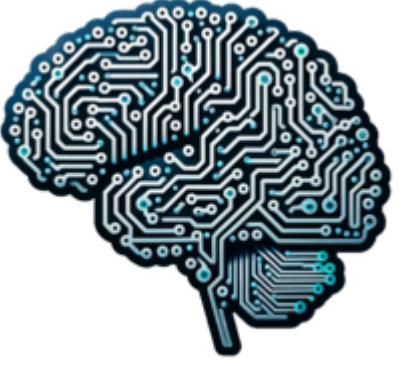
Tesla Optimus

- **Ziele:**

- Nutzung von Robotern in Umgebungen nicht für Mensch geeignet sind
- Nutzung von Robotern bei einfachen und repetitiven Arbeiten

# Auswirkungen

## Wie wird KI die **Arbeitswelt** verändern?



GR1



Tesla Optimus



H1

### Vorteile:

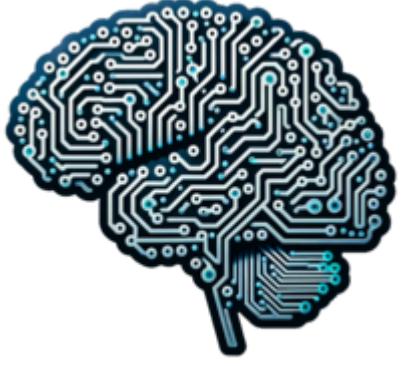
- Arbeiten wo Menschen nicht arbeiten können
- Unterstützung von alten und gebrechlichen Menschen
- Könnten Anfang sich auch selbst zu bauen

### Nachteile:

- Ersetzung Menschen in verschiedenen Jobs
- Einsatz als Soldaten

# Auswirkungen

## Wie wird KI die Welt verändern?



## Moore's Law for Everything

by Sam Altman · March 16, 2021

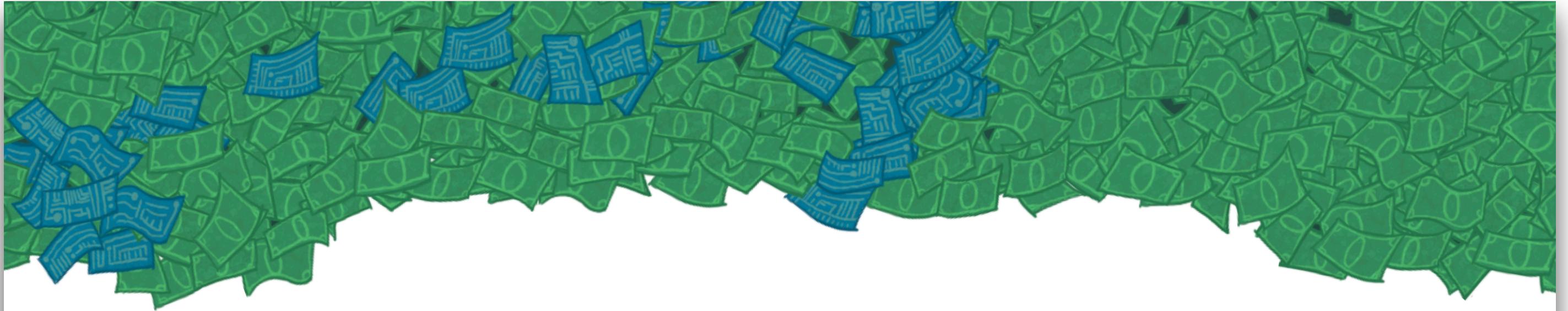
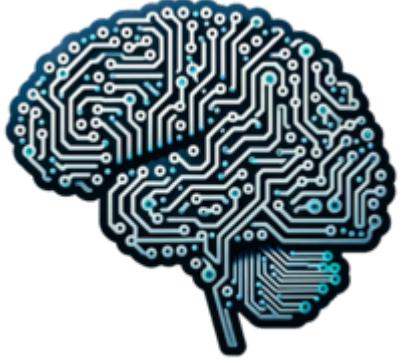
My work at OpenAI reminds me every day about the magnitude of the socioeconomic change that is coming sooner than most people believe. Software that can think and learn will do more and more of the work that people now do. Even more power will shift from labor to capital. If public policy doesn't adapt accordingly, most people will end up worse off than they are today.

We need to design a system that embraces this technological future and

<https://moores.samaltman.com/>

# Auswirkungen

## Wie wird KI die Welt verändern?



### Moore's Law for Everything

by Sam Altman · March 16, 2021

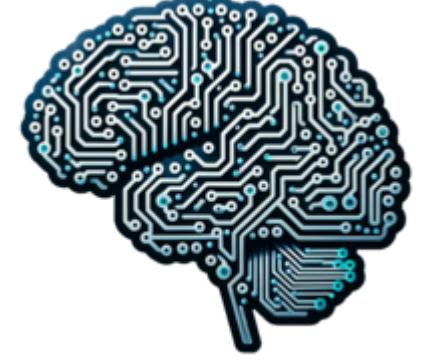
My work at OpenAI reminds me every day about the magnitude of the socioeconomic change that is coming sooner than most people believe. Software that can think and learn will do more and more of the work that people now do. Even more power will shift from labor to capital. If public policy doesn't adapt accordingly, most people will end up worse off than they are today.

We need to design a system that embraces this technological future and

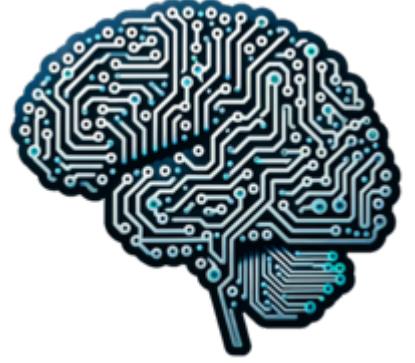
**Was denken Sie?**

AGO SOLVOU SIC!

<https://moores.samaltman.com/>



Intelligence  
Deep Halluzinationen  
Texterkennung Assistenten  
Supervised Bilderkennung  
Learning  
Workspace Edge  
Fahren Chatbot  
Bard Robotik Home  
Reinforcement Smart Gesichtserkennung  
Artificial Neuronale Spracherkennung  
General chat Virtuelle Generative  
GPT Autonomes Netze  
Expertensysteme



Intelligence  
Deep Halluzinationen  
Texterkennung Assistenten  
Supervised Bilderkennung  
Learning Edge  
Workspace Ethik  
Fahren Chatbot  
Bard Robotik  
Reinforcement  
KI Artificial  
Neuronale Spracherkennung  
Machine Virtuelle  
Chat General  
Generative Schwarmintelligenz  
Smart Gesichtserkennung  
Autonomes GPTNetze  
Gesichtserkennung Expertensysteme

