

Künstliche Intelligenz: Revolution oder Risiko?

## Teil 2: Generative KI verstehen – Von Neuronen zu ChatGPT

Philipp Gerber

30. November 2023



**itcc.li**

# IT Crowd Club Liechtenstein

informieren • austauschen • vernetzen

- Gegründet 2011
- 50 Mitglieder



**itcc.li**

# Generative KI



Welcome to

EEEEEE LL IIII ZZZZZZZ AAAAA  
EE LL II ZZ AA AA  
-----

## ELIZA-Effekt

Computer, die Texte generieren können,  
werden von Menschen oft als  
intelligent und empathisch  
wahrgenommen,  
auch wenn diese nur begrenzte und  
mechanische Antworten generieren.

Eliza  
The c  
This

ELIZA:  
YOU:  
ELIZA:  
YOU:  
ELIZA:  
YOU:  
ELIZA:  
YOU:  
ELIZA:

YOU: It's true. I am unhappy.

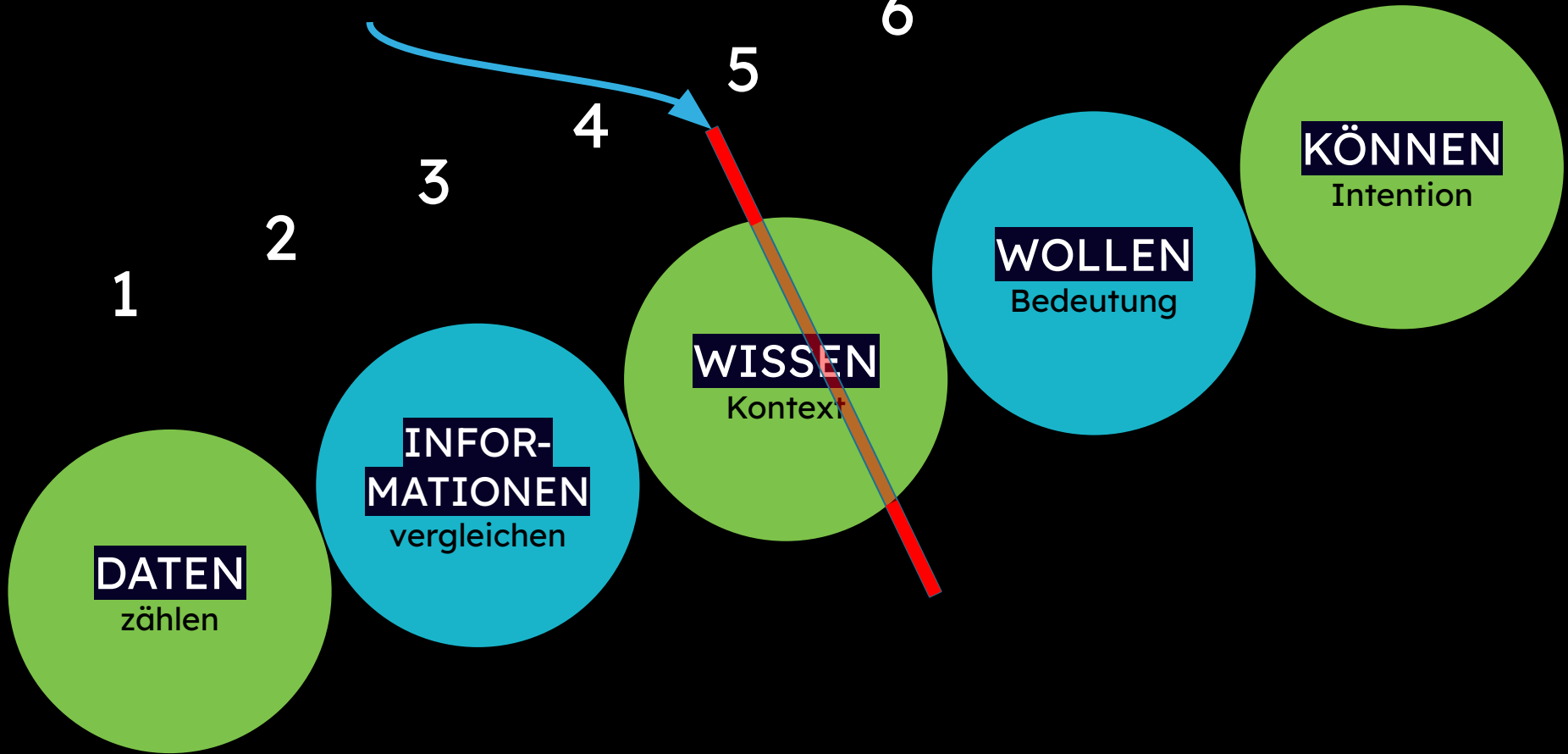
ELIZA: Can you explain what made you unhappy ?

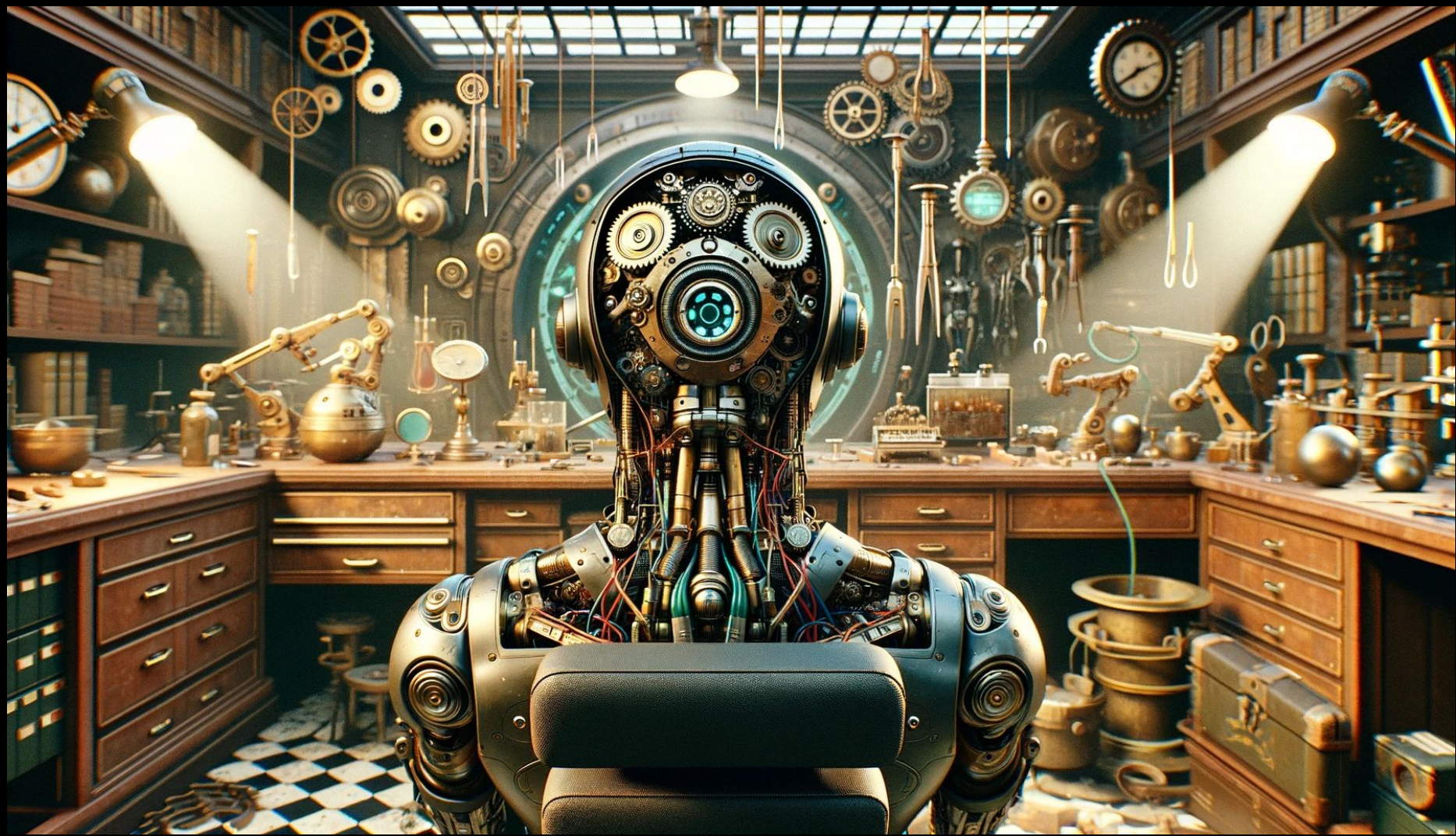
YOU: █

re ?



# ChatGPT?





DALL-E3



**Was ist Künstliche Intelligenz?**

**Wie lernt eine Künstliche Intelligenz?**

**Kernkomponenten generativer KI**



# Beispiel ChatGPT

## Prompt (Eingabe):

Berechne mit Hilfe des 3. Kepler'schen Gesetzes die grosse Halbachse

- für Saturn (Umlaufzeit 29,46 Jahre) und
- für Neptun (Umlaufzeit 164,8 Jahre).

(Tipp: Verwende die Information, dass die Erde in einem Jahr um die Sonne kreist und die grosse Halbachse 1AE beträgt).



## ChatGPT

Das 3. Kepler'sche Gesetz besagt, dass das Verhältnis zwischen der Umlaufzeit eines Planeten um die Sonne und der grossen Halbachse seiner Umlaufbahn im Quadrat proportional ist:

$$(T1 / T2)^2 = (a1 / a2)^3$$

wobei T die Umlaufzeit und a die grosse Halbachse ist.

Für Saturn gilt:

$$T = 29,46 \text{ Jahre}$$

$$a = ?$$

Durch Umstellung des 3. Kepler'schen Gesetzes ergibt sich für a1:

$$a1 = (T1^2)^{(1/3)}$$

Setzt man für T1 die Umlaufzeit von Saturn ein, ergibt sich:

$$a1 = (29,46^2)^{(1/3)} = 9,539 \text{ AE}$$

# Beispiel ChatGPT

Setzt man für T2 die Umlaufzeit von Neptun ein, ergibt sich:

$$a_2 = (164,8^2)^{1/3} = 30,07 \text{ AE}$$

Das bedeutet, dass die grosse Halbachse der Umlaufbahn von Neptun etwa 30,07 Astronomische Einheiten beträgt.

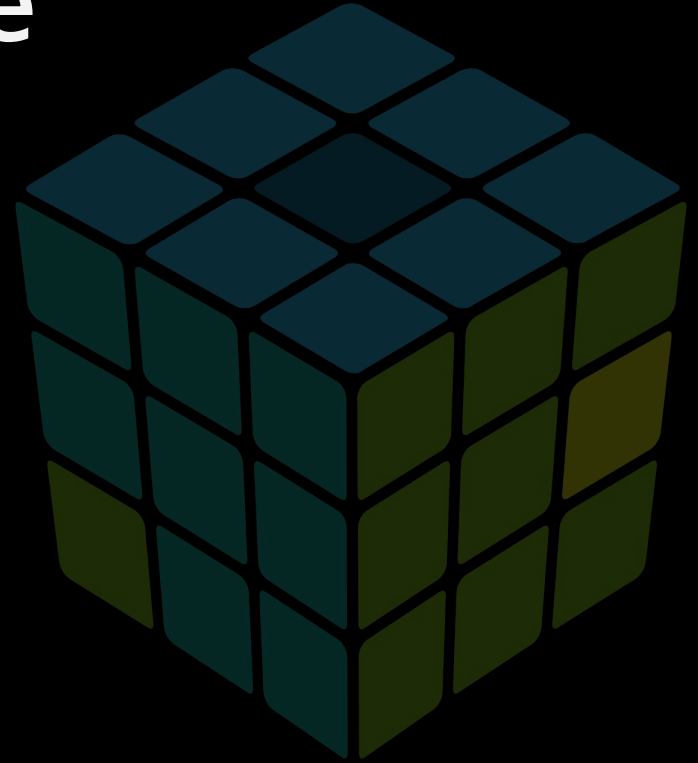
## **Prompt (Eingabe):**

Erstelle ein passendes Bild, kreativ und künstlerisch.



DALL-E3

**Was ist Künstliche  
Intelligenz?**



# Was ist Künstliche Intelligenz?

Intelligenz – intelligent:

von lateinisch *intellegere*  
verstehen, erfassen, erkennen

Künstliche Intelligenz (KI)

ist die Eigenschaft eines IT-Systems,  
»menschenähnliche«, intelligente  
Verhaltensweisen  
zu zeigen.

[\[Deutsches Forschungszentrum für künstliche Intelligenz\]](#)

# Was ist Künstliche Intelligenz?

Künstliche Intelligenz (KI)

ist die Fähigkeit einer Maschine,

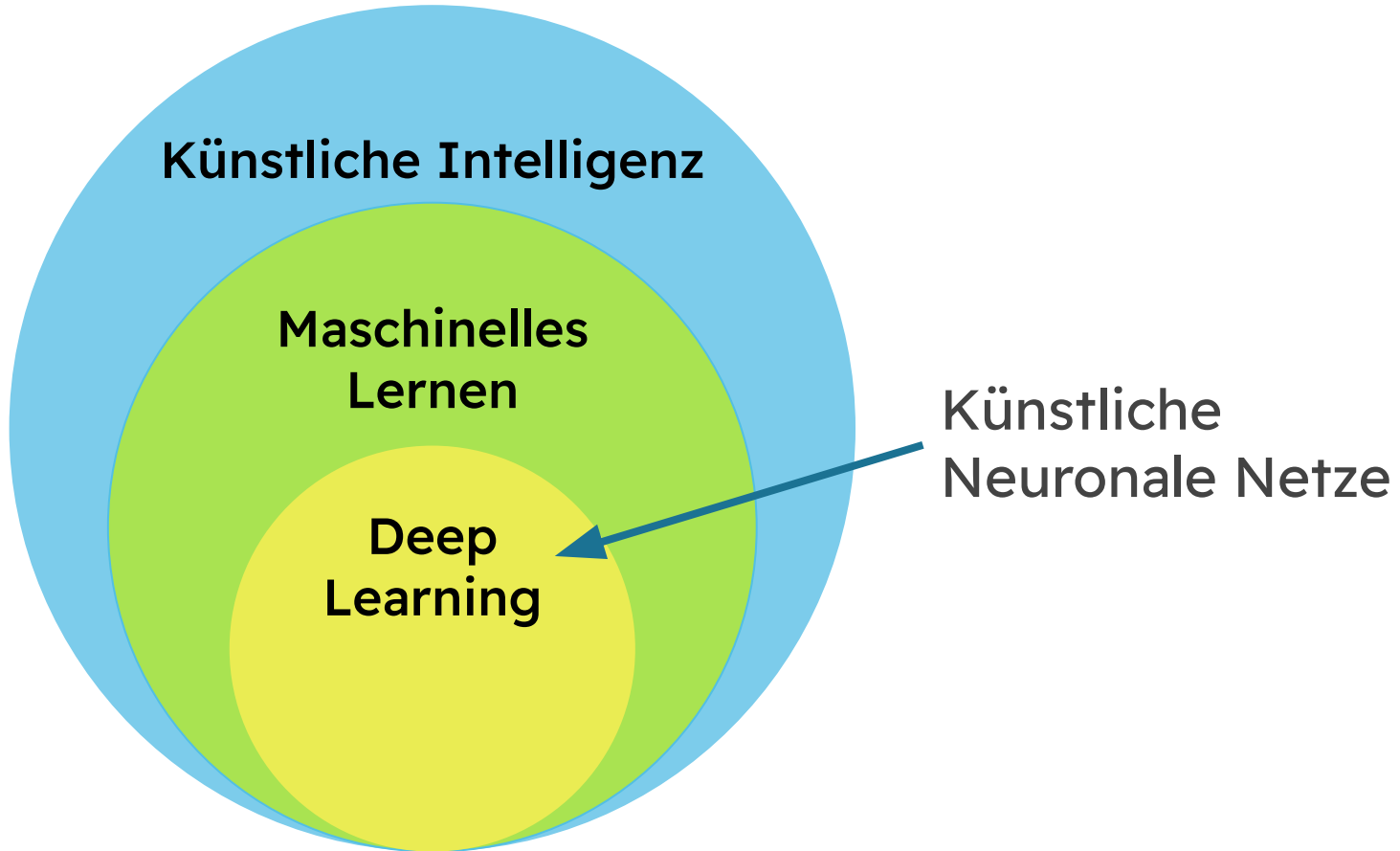
*menschliche Fähigkeiten* wie

logisches Denken, Lernen, Planen und Kreativität

zu **imitieren**.

[[Europäisches Parlament](#)]

# Künstliche Intelligenz

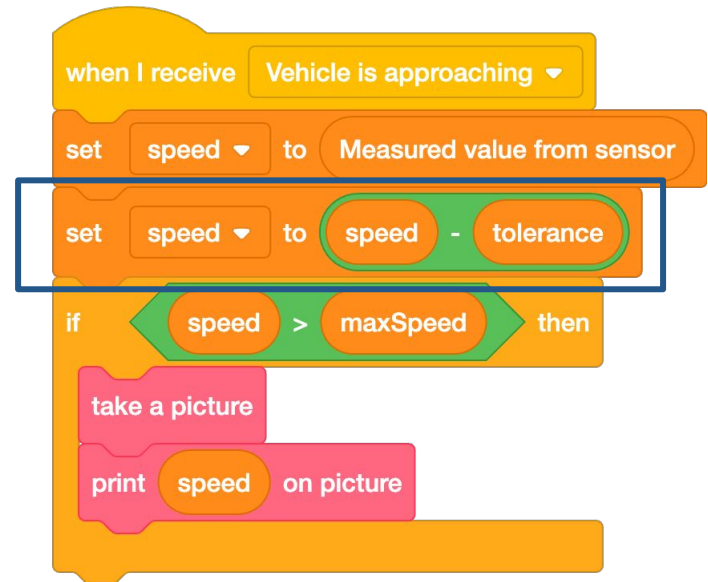
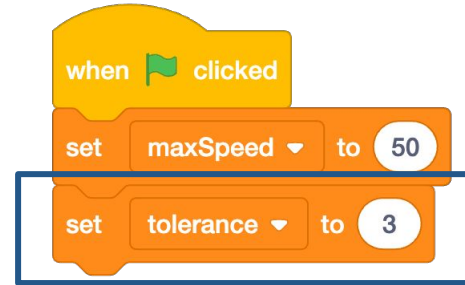




# Herkömmliche Programme vs. KI Systeme

- Festgelegte Regeln  
“wenn-dann”
- Definiertes Umfeld
- Parameter
- Fehlerkorrekturen möglich

Beispiel:  
Geschwindigkeitsüberwachung




# Herkömmliche Programme vs. KI Systeme

## Programm

- Festgelegte Regeln
- Definiertes Umfeld
- Parameter
- Fehlerkorrekturen möglich

## KI System

- Ziel, Randbedingungen
- Flexibles Umfeld
-  "lernen" durch optimieren
- Fehlerkorrekturen?

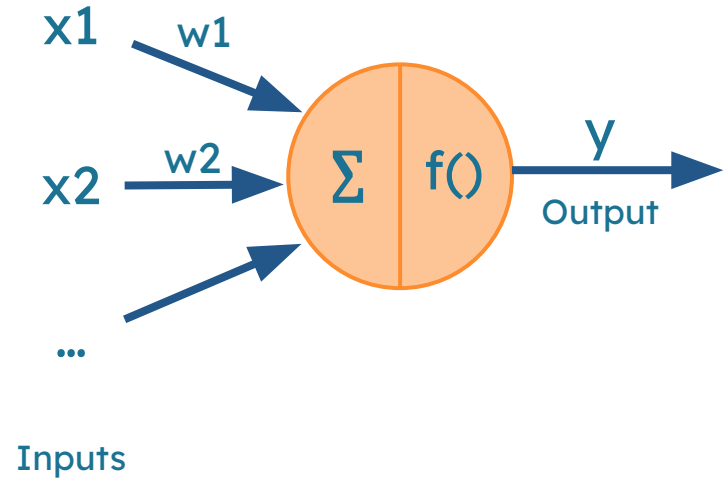
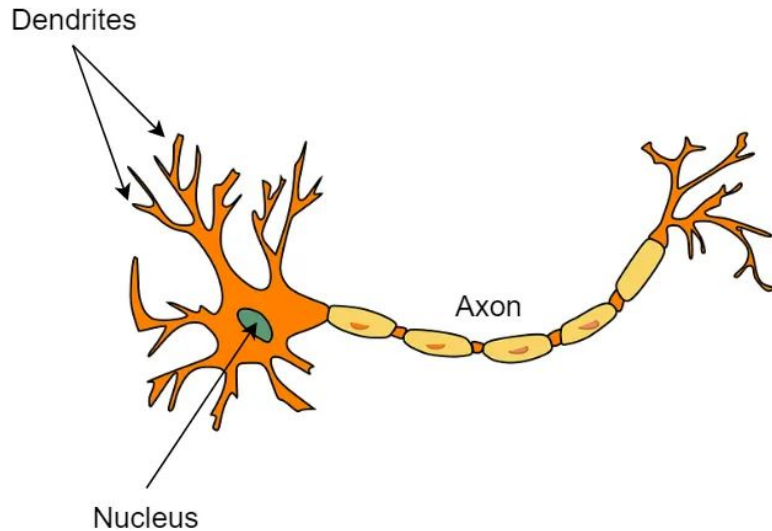
Wie lernt eine Künstliche  
Intelligenz?



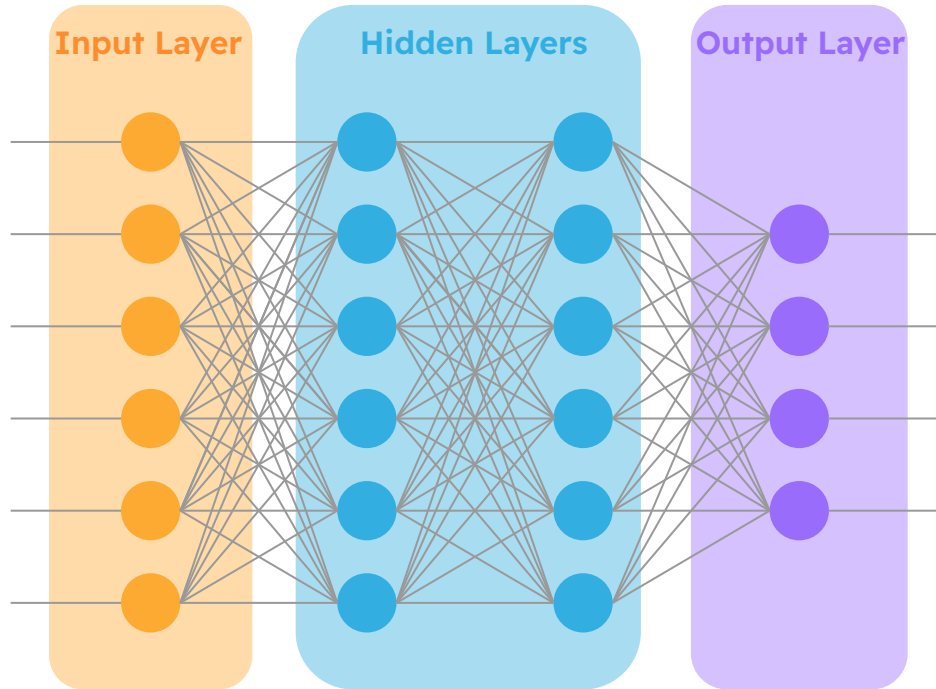


DALL-E3

# Neuronen und Künstliche Neuronen

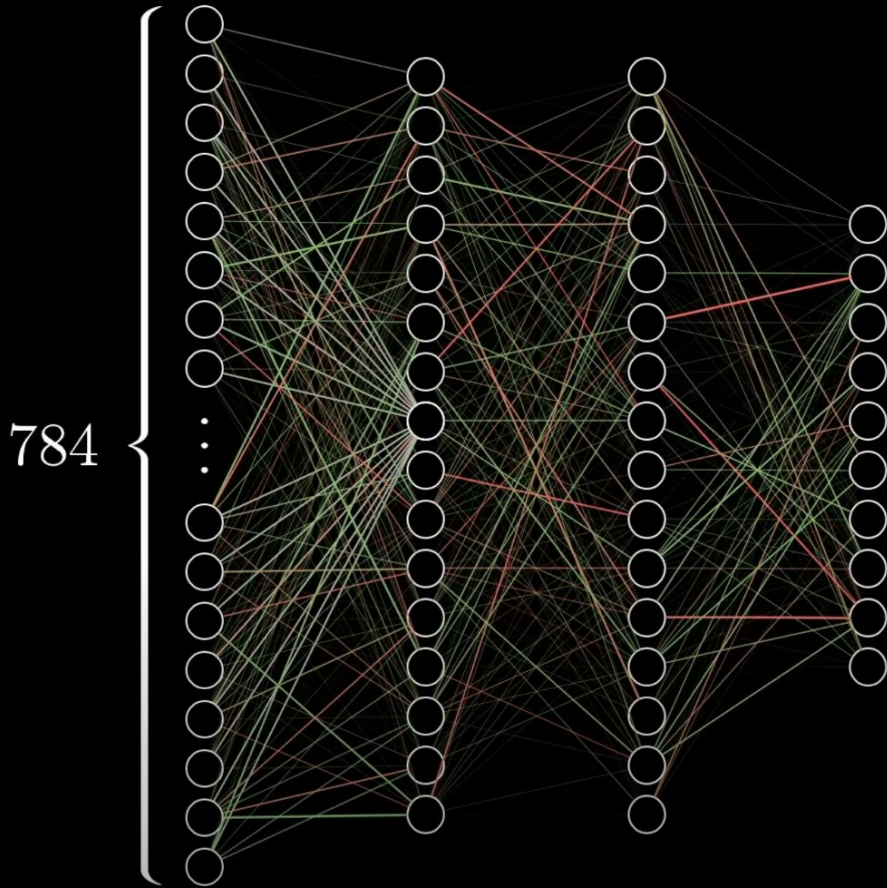


# Künstliches Neuronales Netz



Klassifikation  
oder  
Regression





$$784 \times 16 + 16 \times 16 + 16 \times 10$$

weights

$$16 + 16 + 10$$

biases

**13,002**

Learning  $\rightarrow$  Finding the right weights and biases





Epoch  
000,000

Learning rate  
0.03

Activation  
Tanh

Regularization  
None

Regularization rate  
0

Problem type  
Classification

### DATA

Which dataset do you want to use?



Ratio of training to test data: 50%



Noise: 0



Batch size: 10



REGENERATE

### FEATURES

Which properties do you want to feed in?

$X_1$

$X_2$

$X_1^2$

$X_2^2$

$X_1 X_2$

$\sin(X_1)$

$\sin(X_2)$

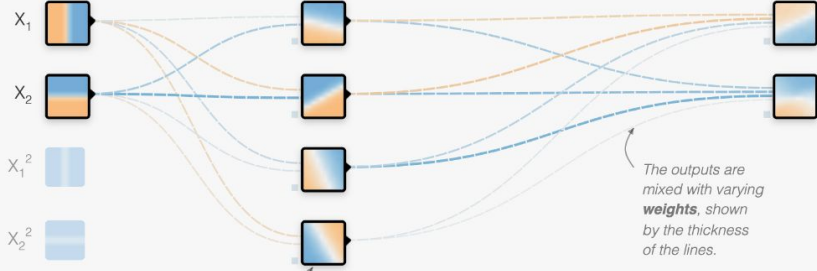
### 2 HIDDEN LAYERS

+ -

4 neurons

+ -

2 neurons



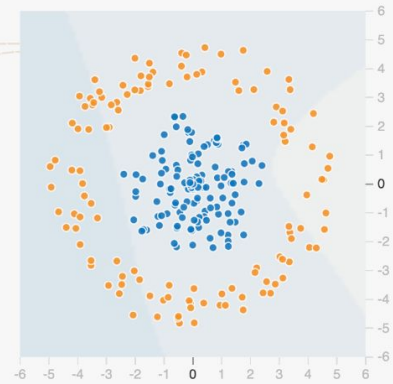
This is the output from one neuron. Hover to see it larger.

The outputs are mixed with varying weights, shown by the thickness of the lines.

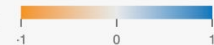
### OUTPUT

Test loss 0.510

Training loss 0.505



Colors shows data, neuron and weight values.



Show test data

Discretize output

# Herkömmliche Programme vs. KI Systeme

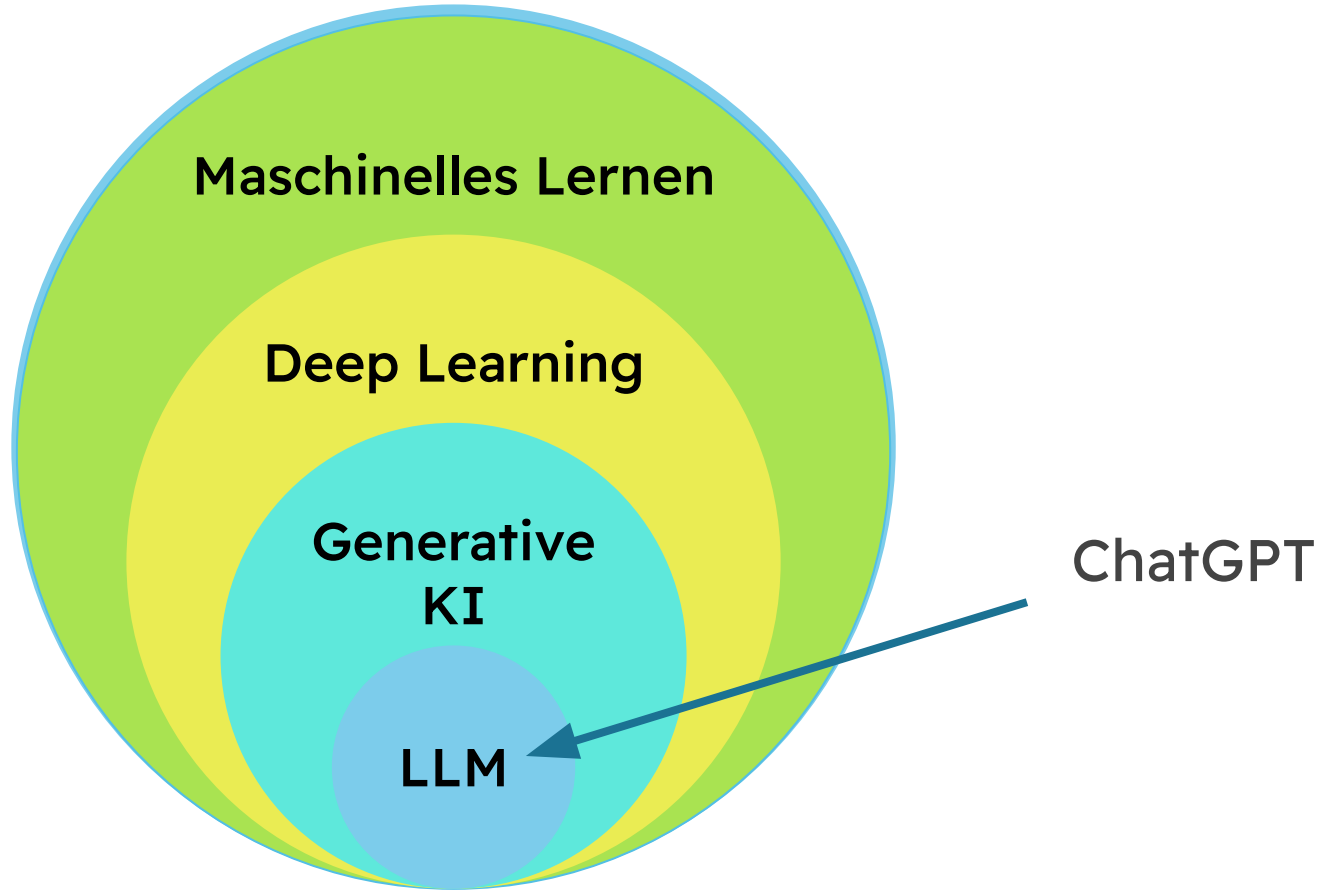
## Programm

- Festgelegte Regeln
- Definiertes Umfeld
- Parameter
- Fehlerkorrekturen möglich

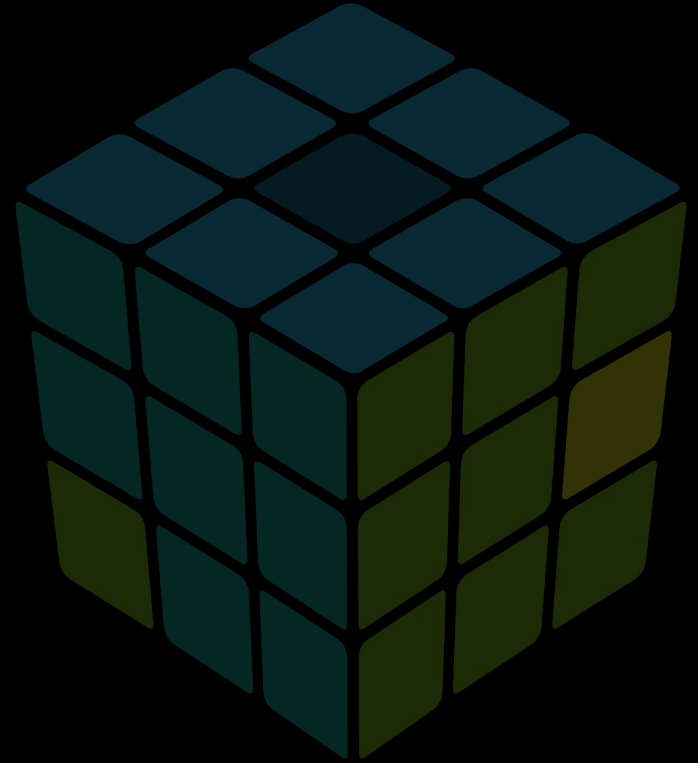
## KI System

- Ziel, Randbedingungen
- Flexibles Umfeld
- “lernen” durch optimieren
- **Fehlerkorrekturen?**

# Generative KI, Large Language Modelle



# Kernkomponenten Generativer KI



# Kernkomponenten Generativer KI

Trainingsdaten



Bedeutung



Verknüpfung



# Trainingsdaten – Text und Tokens



Tokens	Characters
104	363

KI verstehen: Von Neuronen zu ChatGPT  
Do., 30. November 2023, 19:00 Uhr

In diesem Vortrag erfährst du alles über die Grundlagen der KI, vom Neuronalen Netz bis zum LLM (Large language model). Du wirst lernen, wie generative KI-Systeme funktionieren und wie sie trainiert werden. Im Mittelpunkt der Diskussion steht die Frage, ob ChatGPT wirklich intelligent ist.

TEXT    TOKEN IDS

# LLM – Trainingsdaten (GPT3)

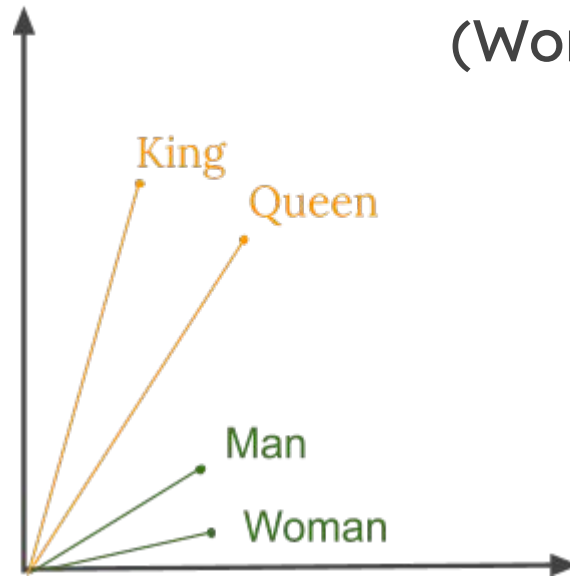
Token: Wort, Teil eines Wortes oder Satzzeichen

- Wikipedia: 3 Milliarden Token
- Books1: 12 Milliarden Token
- Books2: 55 Milliarden Token
- Common Crawl: 410 Milliarden Token
- WebText2: 19 Milliarden Token

Trainiertes Neuronales Netz:  
**175 Milliarden Parameter**

# Bedeutung und Beziehung von Wörtern

Worteinbettung  
(Word embedding)





# Kontext oder Verknüpfung von Wörtern


## Der Schlüssel zum Textverständnis

Attention

As aliens entered our planet and began to colonize earth a certain group of extraterrestrials ...

# Was macht ein LLM wie **GPT** aus?

- Trainingsdaten mit Milliarden Token
- Bedeutung (**Word embedding**)
- Verknüpfung (**Attention**)
- + viele Wochen Training



**G**enerative  
**P**re-trained  
**T**ransformer

Eine KI zur **Textvervollständigung**

# GPT und Feinabstimmung $\Rightarrow$ ChatGPT

①

**Trainieren**

Dialoge führen.

②

**Lernen**

Was sind gute  
Antworten?

③

**Trainieren**

Dialoge mit  
guten Antworten.

# Large Language Modelle – Risiken

- Halluzinationen  
Falsche Informationen und Desinformation
- Bias  
Diskriminierung und Vorurteile
- Sicherheits- und Datenschutzverletzungen
- etc.

# GPT und Feinabstimmung $\Rightarrow$ ChatGPT

②

## Lernen

Was sind gute  
Antworten?

③

## Trainieren

Dialoge mit  
guten Antworten.

## “Fehlerkorrekturen”

oder vermeiden von

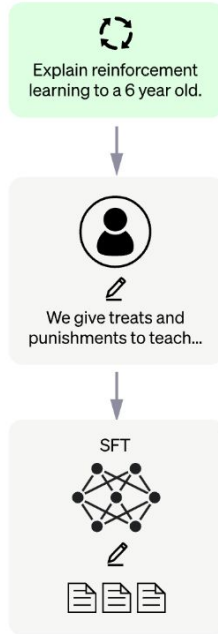
- Halluzinationen
- Diskriminierung
- Vorurteilen
- gefährlichen  
Informationen

aber auch Zensur

### Step 1

## Collect demonstration data and train a supervised policy.

A prompt is sampled from our prompt dataset.



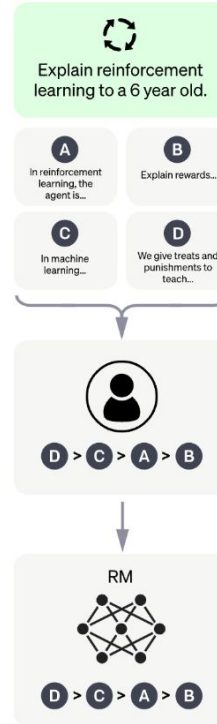
A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.

### Step 2

## Collect comparison data and train a reward model.

A prompt and several model outputs are sampled.



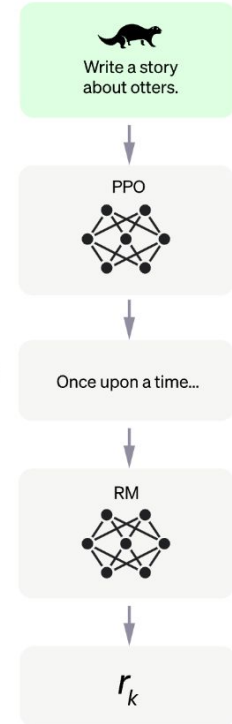
A labeler ranks the outputs from best to worst.

This data is used to train our reward model.

### Step 3

## Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

A new prompt is sampled from the dataset.



The PPO model is initialized from the supervised policy.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.

# ChatGPT und Bard in der Praxis

OpenAI:

<https://chat.openai.com/>

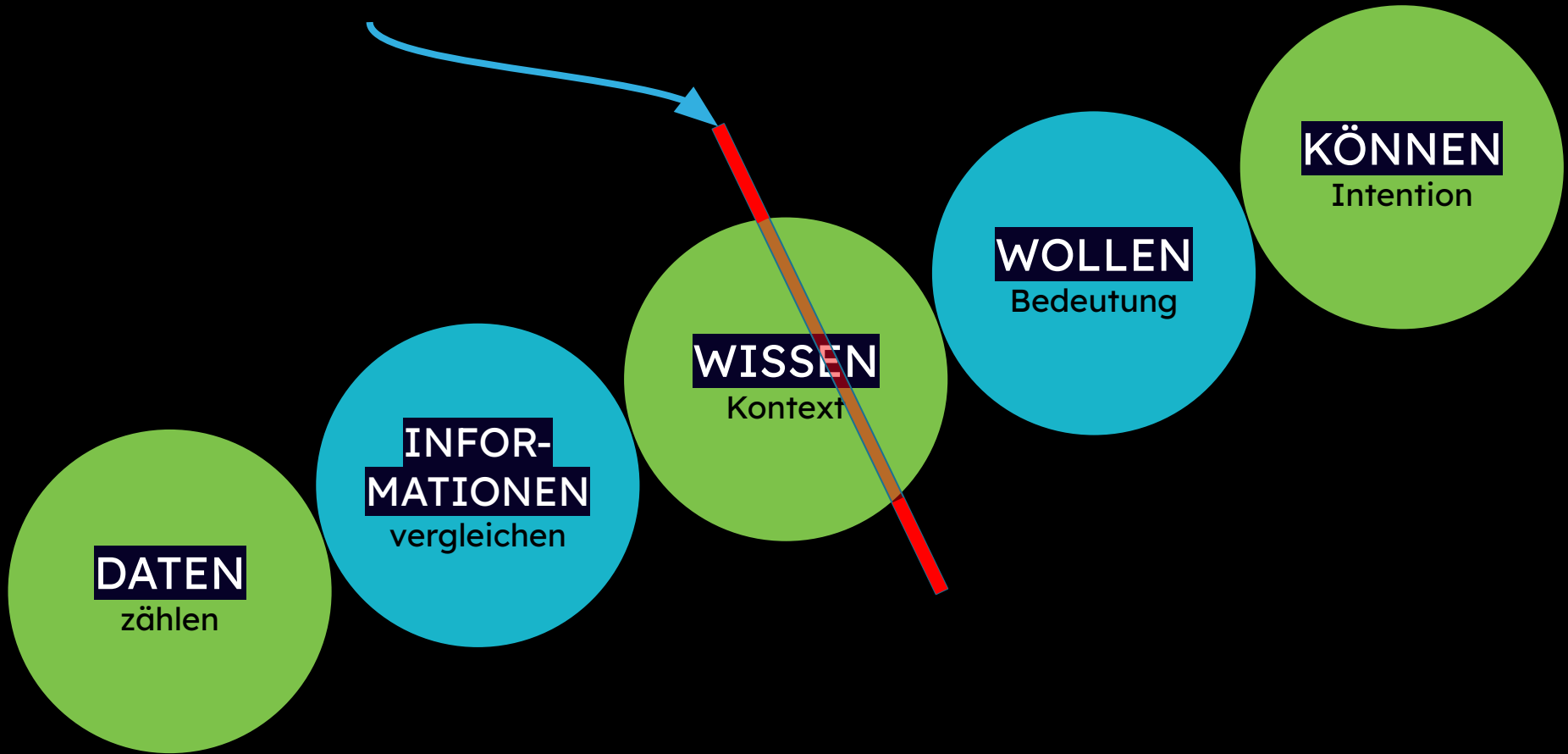


Google Bard

<https://bard.google.com/>

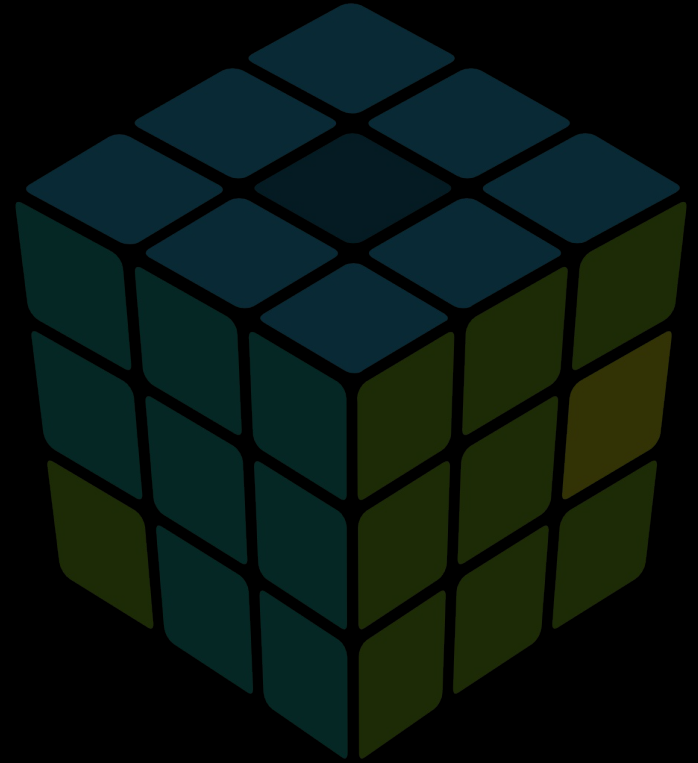


# ChatGPT?





Fragen?



# KI: Revolution oder Risiko?

Do., 29. Februar 2024

## 4. ChatGPT selbst gemacht: Ein Leitfaden

Du möchtest deine eigene KI-Software entwickeln? In diesem Workshop lernst du, wie du mit Open-Source-Tools und -Ressourcen ein eigenes generatives KI-Modell erstellst.

Wir zeigen dir, wie du die Grundmodelle trainierst.

Anmelden: [www.itcc.li/ki](http://www.itcc.li/ki)

Do., 25. Januar 2024

## 3. Die Kunst des Prompt Engineerings

In diesem Vortrag erfährst du, wie du die Ausgabe einer KI beeinflussen und sie für deine eigenen Bedürfnisse anpassen kannst.

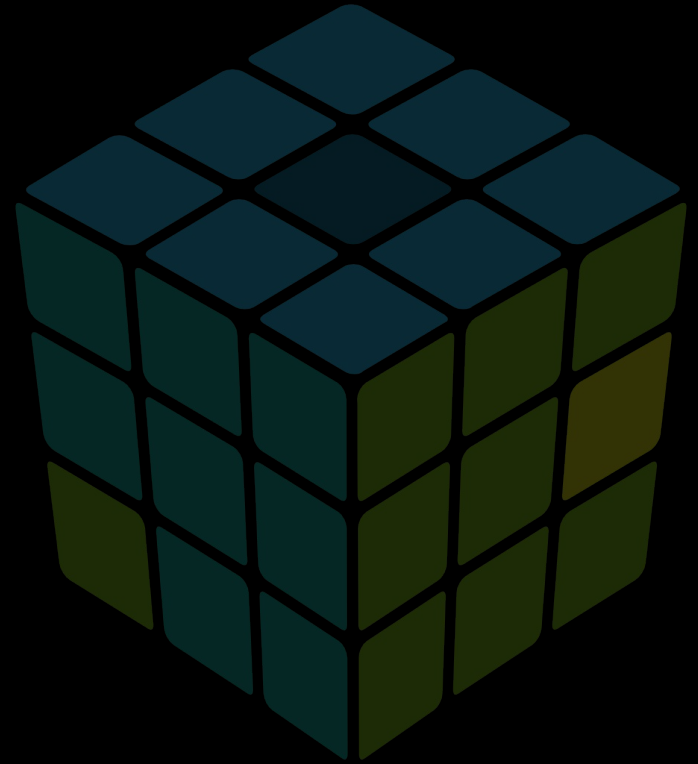
Wir zeigen dir, wie Prompts strukturiert werden und wie du mithilfe von Prompt-Hacking-Techniken die Ergebnisse verbessern und Einschränkungen umgehen kannst.

Anmelden: [www.itcc.li/ki](http://www.itcc.li/ki)

Vielen Dank!

Mitglied werden

<https://itcc.li>



# Links

## OpenHPI

- [Künstliche Intelligenz und maschinelles Lernen für Einsteiger](#)
- [ChatGPT: Was bedeutet generative KI für unsere Gesellschaft?](#)

## 3Blue1Brown

- [3Blue1Brown : Neural Networks](#)

# Links – Einführung KI

- [Künstliche Intelligenz verstehen](#) (Youtube Playlist)
- [Neural Networks - The Nature of Code](#) (Youtube Playlist)
- [Understanding neural networks](#)  
→ [TensorFlow Playground](#)

## Links – LLM, Transformers

- [Introduction to Large Language Models](#)
- [Introduction to Generative AI](#)
- [Transformers: The model behind GPT, BERT, and T5](#)
- [Transformers Neural Network: A step by step explanation](#)
- [Let's build GPT: from scratch, in code, spelled out.](#)  
[Intro to Large Language Models](#)

# Generative KI

[Björn Ommer – Generative KI: schöne neue Welt?](#)

[How AI Image Generators Work \(Stable Diffusion / Dall-E\) - Computerphile](#)

[How Stable Diffusion Works \(AI Image Generation\)](#)

[So entsteht ein KI Bild in midjourney](#)

# Buch – Deep Learning illustriert

Jon Krohn / Grant Beyleveld / Aglaé Bassens

## Deep Learning illustriert

Eine anschauliche Einführung in Machine Vision, Natural Language Processing und Bilderzeugung für Programmierer und Datenanalysten

[dpunkt.de](https://dpunkt.de)

