

7 - (Künstliche) Intelligenz

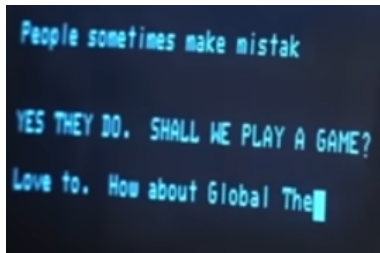
Marianne Maertens

Technische Universität Berlin

Wintersemester 2025/2026

Ziele & Themen

1. Was ist Intelligenz?
2. Künstliche Intelligenz und maschinelles Lernen
3. Anthropomorphismus - Eigenschaft menschlichen Denkens
4. Wissenschaftlichkeit: Kritisches Denken



Ziele & Themen

1. Was ist Intelligenz?
2. Künstliche Intelligenz und maschinelles Lernen
3. Anthropomorphismus - Eigenschaft menschlichen Denkens
4. Wissenschaftlichkeit: Kritisches Denken



Was ist Intelligenz?

Lesen Sie das Kapitel aus dem Buch “Intelligente Täuschung”

- Wie definiert der Autor KI?
- Wie definiert er Intelligenz?
- Welchen Unterschied sieht der Autor zwischen der Intelligenz von Maschinen und Menschen?
- Welche “Zutaten” sieht der Autor als notwendig an, um eine menschenähnliche allgemeine Intelligenz in Computern nachzubilden?



Wie definiert der Autor KI?

- KI wird attestiert, wenn Computerprogramme Aufgaben übernehmen, für die Menschen eine gewisse Intelligenz benötigen
- eine Fähigkeit gilt so lange als Ausdruck von Intelligenz, bis ein Computer sie ausführen kann
- gerade jene Fähigkeiten gelten als Ausdruck spezifisch menschlicher Intelligenz, die ein Computer nicht beherrscht

Wie definiert der Autor KI?

- KI wird attestiert, wenn Computerprogramme Aufgaben übernehmen, für die Menschen eine gewisse Intelligenz benötigen
- eine Fähigkeit gilt so lange als Ausdruck von Intelligenz, bis ein Computer sie ausführen kann
- gerade jene Fähigkeiten gelten als Ausdruck spezifisch menschlicher Intelligenz, die ein Computer nicht beherrscht

⇒ echte KI kann es nicht geben

Wie wird Intelligenz definiert? Woran erinnert Sie das?

Wie wird Intelligenz definiert? Woran erinnert Sie das?

- intuitiv: “We know it when we see it”

Wie wird Intelligenz definiert? Woran erinnert Sie das?

- intuitiv: “We know it when we see it”
- Psychologie: das, was ein Intelligenztest misst

Wie wird Intelligenz definiert? Woran erinnert Sie das?

- intuitiv: “We know it when we see it”
- Psychologie: das, was ein Intelligenztest misst
 1. große Menge von Aufgaben, die vermeintlich Intelligenz erfordern

Wie wird Intelligenz definiert? Woran erinnern Sie das?

- intuitiv: “We know it when we see it”
- Psychologie: das, was ein Intelligenztest misst
 1. große Menge von Aufgaben, die vermeintlich Intelligenz erfordern
 2. Aufgaben, die von wenigen Personen gelöst werden können, sind schwerer als Aufgaben, die von vielen Personen gelöst werden können

Wie wird Intelligenz definiert? Woran erinnert Sie das?

- intuitiv: “We know it when we see it”
- Psychologie: das, was ein Intelligenztest misst
 1. große Menge von Aufgaben, die vermeintlich Intelligenz erfordern
 2. Aufgaben, die von wenigen Personen gelöst werden können, sind schwerer als Aufgaben, die von vielen Personen gelöst werden können→ Personen, die mehr schwere Aufgaben lösen können, sind per Definition intelligenter

Wie wird Intelligenz definiert? Woran erinnert Sie das?

- intuitiv: “We know it when we see it”
- Psychologie: das, was ein Intelligenztest misst
 1. große Menge von Aufgaben, die vermeintlich Intelligenz erfordern
 2. Aufgaben, die von wenigen Personen gelöst werden können, sind schwerer als Aufgaben, die von vielen Personen gelöst werden können

→ Personen, die mehr schwere Aufgaben lösen können, sind per Definition intelligenter
- stichprobenabhängig - Teststichprobe, Validierungsstichprobe (Validität!)
- aufgabenabhängig

Intelligenztests

- Wortanalogien
(verbal/semantisch/konzeptuell)

Lehrer : Kreide = Soldat : ?

Sonne : Planet = Erde : ?

Hernandez-Orallo et al. (2016) <https://www.sciencedirect.com/science/article/pii/S0004370215001538>

Intelligenztests

- Wortanalogien
(verbal/semantisch/konzeptuell)

Lehrer : Kreide = Soldat : ?

Sonne : Planet = Erde : ?

(Planeten umkreisen Sonne)

Hernandez-Orallo et al. (2016) <https://www.sciencedirect.com/science/article/pii/S0004370215001538>

Intelligenztests

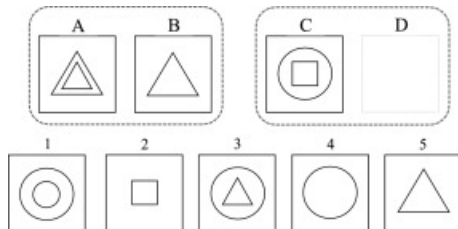
- Wortanalogien
(verbal/semantisch/konzeptuell)
- Zahlenfolgen
(mathematisch/numerisch)

1,1,2,3,5,8,?

Hernandez-Orallo et al. (2016) <https://www.sciencedirect.com/science/article/pii/S0004370215001538>

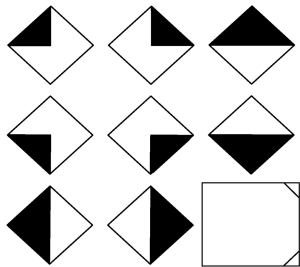
Intelligenztests

- Wortanalogien
(verbal/semantisch/konzeptuell)
- Zahlenfolgen
(mathematisch/numerisch)
- Geometrische Analogien



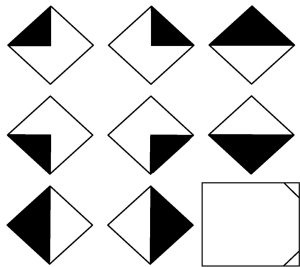
Intelligenztests

- Wortanalogien
(verbal/semantisch/konzeptuell)
- Zahlenfolgen
(mathematisch/numerisch)
- Geometrische Analogien
- Räumliches Schlussfolgern



Intelligenztests

- Wortanalogien
(verbal/semantisch/konzeptuell)
- Zahlenfolgen
(mathematisch/numerisch)
- Geometrische Analogien
- Räumliches Schlussfolgern



- allgemeine Intelligenz: Fähigkeit viele verschiedene (mehr oder weniger sinnvolle) Testaufgaben erfolgreich zu bearbeiten

Welchen Unterschied sieht der Autor zwischen der Intelligenz von Maschinen und Menschen?

Welche “Zutaten” sieht der Autor als notwendig an, um eine menschenähnliche allgemeine Intelligenz in Computern nachzubilden?

Welchen Unterschied sieht der Autor zwischen der Intelligenz von Maschinen und Menschen?

- Schachprogramme spielen nur Schach, Taschenrechner rechnen nur, Programme zur Analyse von Röntgenbilder analysieren Röntgenbilder, ...
- in allem teils besser als Menschen, aber Inselbegabungen

Welche “Zutaten” sieht der Autor als notwendig an, um eine menschenähnliche allgemeine Intelligenz in Computern nachzubilden?

Welchen Unterschied sieht der Autor zwischen der Intelligenz von Maschinen und Menschen?

- Schachprogramme spielen nur Schach, Taschenrechner rechnen nur, Programme zur Analyse von Röntgenbilder analysieren Röntgenbilder, ...
- in allem teils besser als Menschen, aber Inselbegabungen
- “Erfahrungen” im Schach nicht übertragbar auf Poker
- Menschen setzen flexibel unterschiedliche Strategien bei verschiedenen Problemen ein bzw. übertragen eine Strategie auf verschiedene Probleme

Welche “Zutaten” sieht der Autor als notwendig an, um eine menschenähnliche allgemeine Intelligenz in Computern nachzubilden?

Welchen Unterschied sieht der Autor zwischen der Intelligenz von Maschinen und Menschen?

- Schachprogramme spielen nur Schach, Taschenrechner rechnen nur, Programme zur Analyse von Röntgenbilder analysieren Röntgenbilder, ...
- in allem teils besser als Menschen, aber Inselbegabungen
- “Erfahrungen” im Schach nicht übertragbar auf Poker
- Menschen setzen flexibel unterschiedliche Strategien bei verschiedenen Problemen ein bzw. übertragen eine Strategie auf verschiedene Probleme

Welche “Zutaten” sieht der Autor als notwendig an, um eine menschenähnliche allgemeine Intelligenz in Computern nachzubilden?

- Suchalgorithmen
- Lernalgorithmen
- Sprachmodelle



erkläre mir kurz technisch wie Du funktionierst

Ich bin ein **großes neuronales Sprachmodell (LLM)**, genauer ein **Transformermodell**.

Technisch läuft es so:

1. Eingabe → Tokenisierung

Dein Text wird in kleine Einheiten (Tokens) zerlegt.

2. Transformer-Architektur

Diese besteht vor allem aus *Self-Attention*-Schichten. Sie berechnen für jedes Token, wie stark es auf andere Tokens im Kontext „achten“ soll. Das ergibt eine Kontextrepräsentation deines Textes.

3. Nächste-Token-Vorhersage

Das Modell wurde darauf trainiert, immer das wahrscheinlichste nächste Token vorherzusagen, Milliarden Mal, auf sehr großen Textmengen.

4. Gewichtete Wahrscheinlichkeiten → Antwort

Die Wahrscheinlichkeiten über alle möglichen nächsten Tokens werden genutzt, um Wort für Wort eine Antwort zu generieren.

5. Kein Abruf von Fakten aus einer Datenbank

Alles Wissen liegt in den trainierten Gewichten des neuronalen Netzes, nicht in expliziten Einträgen.

Künstliche Intelligenz

- ChatGPT (**G**enerative **P**retrained **T**ransformer) ist ein großes Sprachmodell - Large Language Model (**LLM**)

Künstliche Intelligenz

- ChatGPT (**G**enerative **P**retrained **T**ransformer) ist ein großes Sprachmodell - Large Language Model (**LLM**)
- LLMs sind Transformermodelle, eine bestimmte Architektur von Deep Neural Networks (**DNN**), die auf Sprachverarbeitung trainiert worden

Künstliche Intelligenz

- ChatGPT (**G**enerative **P**retrained **T**ransformer) ist ein großes Sprachmodell - Large Language Model (**LLM**)
- LLMs sind Transformermodelle, eine bestimmte Architektur von Deep Neural Networks (**DNN**), die auf Sprachverarbeitung trainiert worden
- Deep Neural Networks (Tiefe Neuronale Netzwerke) sind **künstliche neuronale Netzwerke** mit vielen Schichten

Künstliche Intelligenz

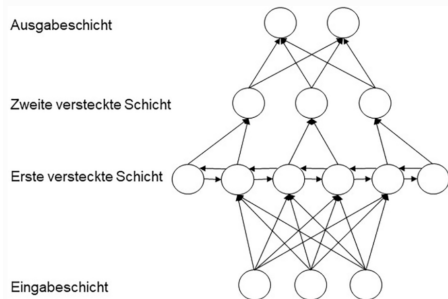
- ChatGPT (**G**enerative **P**retrained **T**ransformer) ist ein großes Sprachmodell - Large Language Model (**LLM**)
- LLMs sind Transformermodelle, eine bestimmte Architektur von Deep Neural Networks (**DNN**), die auf Sprachverarbeitung trainiert worden
- Deep Neural Networks (Tiefe Neuronale Netzwerke) sind **künstliche neuronale Netzwerke** mit vielen Schichten
- Künstliche Neuronale Netzwerke sind eine Form des **maschinellen Lernens**

Künstliche Intelligenz

- ChatGPT (**G**enerative **P**retrained **T**ransformer) ist ein großes Sprachmodell - Large Language Model (**LLM**)
- LLMs sind Transformermodelle, eine bestimmte Architektur von Deep Neural Networks (**DNN**), die auf Sprachverarbeitung trainiert worden
- Deep Neural Networks (Tiefe Neuronale Netzwerke) sind **künstliche neuronale Netzwerke** mit vielen Schichten
- Künstliche Neuronale Netzwerke sind eine Form des **maschinellen Lernens**
- Maschinelles Lernen: Oberbegriff für Verfahren, bei denen Computer aus Daten Muster lernen

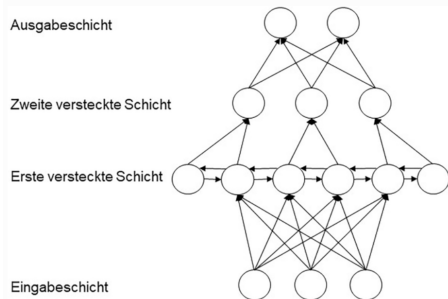
Architektur tiefer neuronaler Netze

- künstliche neuronale Netze sind mathematische Modelle: gerichtete Graphen, bestehend aus Knoten (Neuronen) und Kanten (den gewichteten Verbindungen zwischen den Neuronen)



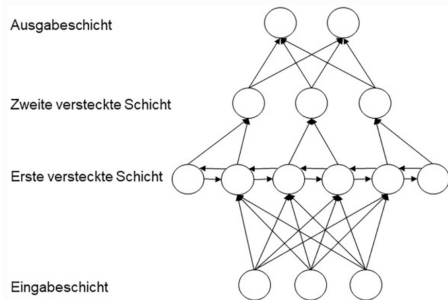
Architektur tiefer neuronaler Netze

- künstliche neuronale Netze sind mathematische Modelle: gerichtete Graphen, bestehend aus Knoten (Neuronen) und Kanten (den gewichteten Verbindungen zwischen den Neuronen)
- Abstraktion von Nervenzellen (Neuronen)



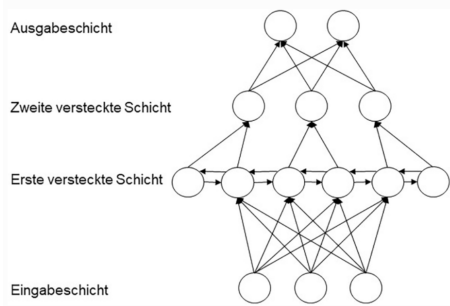
Architektur tiefer neuronaler Netze

- künstliche neuronale Netze sind mathematische Modelle: gerichtete Graphen, bestehend aus Knoten (Neuronen) und Kanten (den gewichteten Verbindungen zwischen den Neuronen)
- Abstraktion von Nervenzellen (Neuronen)
- in verschiedenen Schichten organisiert



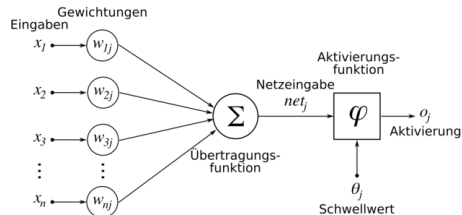
Architektur tiefer neuronaler Netze

- künstliche neuronale Netze sind mathematische Modelle: gerichtete Graphen, bestehend aus Knoten (Neuronen) und Kanten (den gewichteten Verbindungen zwischen den Neuronen)
- Abstraktion von Nervenzellen (Neuronen)
- in verschiedenen Schichten organisiert
- Schichten haben unterschiedliche Funktionen

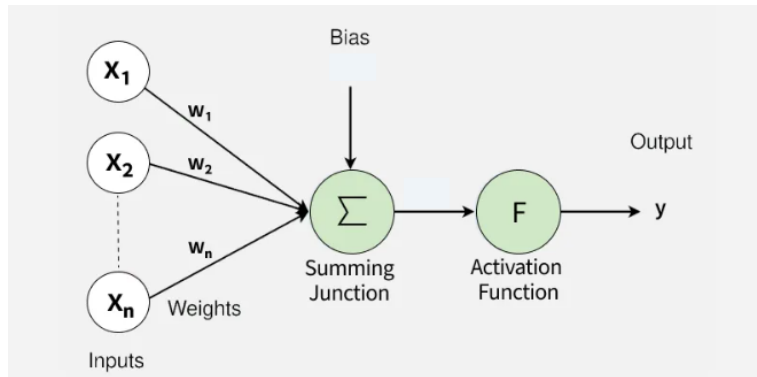


Architektur tiefer neuronaler Netze

- künstliche neuronale Netze sind mathematische Modelle: gerichtete Graphen, bestehend aus Knoten (Neuronen) und Kanten (den gewichteten Verbindungen zwischen den Neuronen)
- Abstraktion von Nervenzellen (Neuronen)
- in verschiedenen Schichten organisiert
- Schichten haben unterschiedliche Funktionen
- jedes Neuron bekommt Zahlen als Eingabe, multipliziert sie mit Gewichten, addiert einen Bias, und schickt das Ergebnis durch eine Aktivierungsfunktion weiter

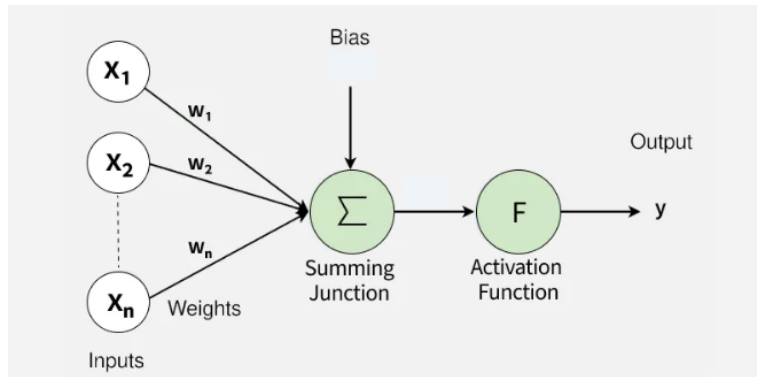


Lernen in tiefen neuronalen Netzen



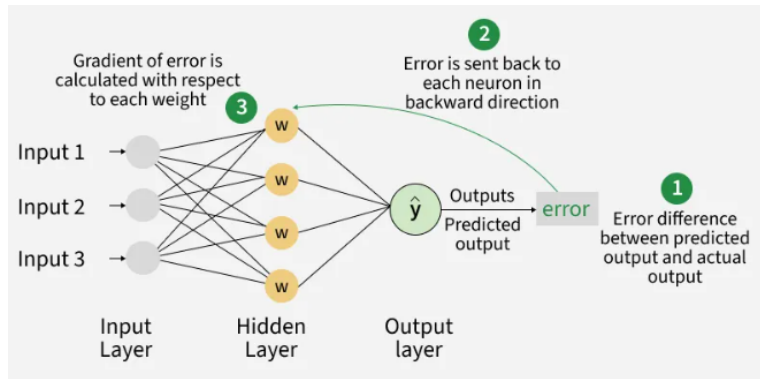
1. Vorwärtsthroughlauf
Input wird Schicht für Schicht verarbeitet
→ Vorhersage (Output)

Lernen in tiefen neuronalen Netzen



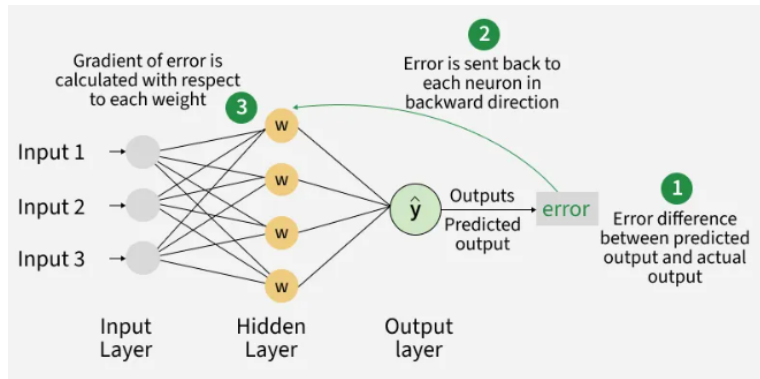
1. Vorwärtsthroughlauf
Input wird Schicht für Schicht verarbeitet
→ Vorhersage (Output)
2. Vorhersage wird mit Zielwert verglichen
→ Fehler (loss)

Lernen in tiefen neuronalen Netzen



1. Vorwärtsthroughlauf
Input wird Schicht für Schicht verarbeitet
→ Vorhersage (Output)
2. Vorhersage wird mit Zielwert verglichen
→ Fehler (loss)

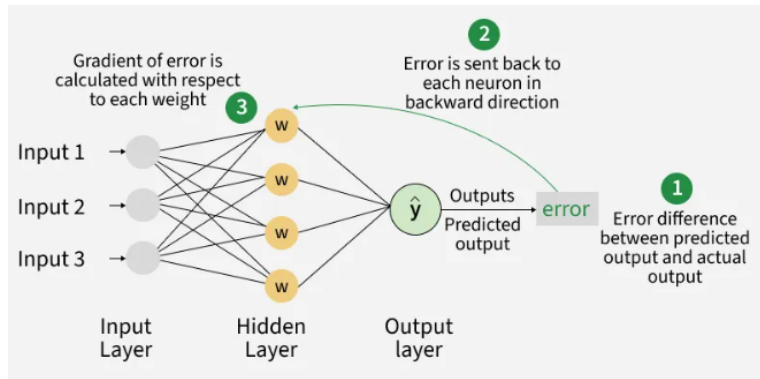
Lernen in tiefen neuronalen Netzen



1. Vorwärtsthroughlauf
Input wird Schicht für Schicht verarbeitet
→ Vorhersage (Output)
2. Vorhersage wird mit Zielwert verglichen
→ Fehler (loss)
3. Rückwärtsthroughlauf (Backpropagation)

- Netz berechnet wie stark jedes Gewicht und Bias zum Fehler beiträgt, dazu verwendet es Ableitungen (Gradienten) der Aktivierungsfunktion und des Fehlers

Lernen in tiefen neuronalen Netzen



1. Vorwärtsthroughlauf
Input wird Schicht für Schicht verarbeitet
→ Vorhersage (Output)
2. Vorhersage wird mit Zielwert verglichen
→ Fehler (loss)
3. Rückwärtsthroughlauf (Backpropagation)

- Netz berechnet wie stark jedes Gewicht und Bias zum Fehler beiträgt, dazu verwendet es Ableitungen (Gradienten) der Aktivierungsfunktion und des Fehlers
- Gewichte und Bias werden so angepasst, dass der Fehler minimiert wird

Quelle: <https://www.geeksforgeeks.org/deep-learning/the-role-of-weights-and-bias-in-neural-networks/>

Was wird gelernt?

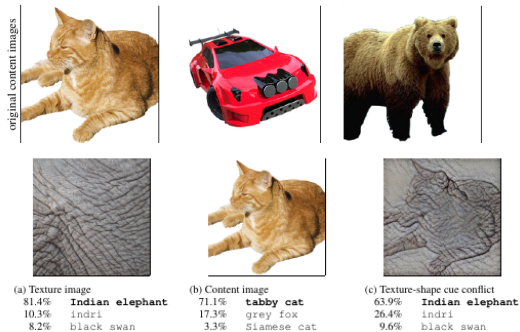
- Ziel: Objekterkennung



Geirhos et al., 2019

Was wird gelernt?

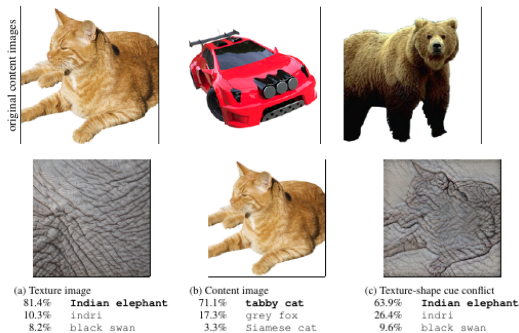
- Ziel: Objekterkennung
- Transfertests! sonst nicht klar, was gelernt wird



Geirhos et al., 2019

Was wird gelernt?

- Ziel: Objekterkennung
- Transfertest! sonst nicht klar, was gelernt wird
- Erkennen von Zusammenhängen vs. Erklären von Zusammenhängen

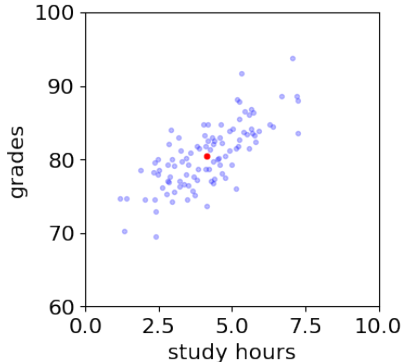


Geirhos et al., 2019

Lineare Regression als Beispiel für Maschinelles Lernen

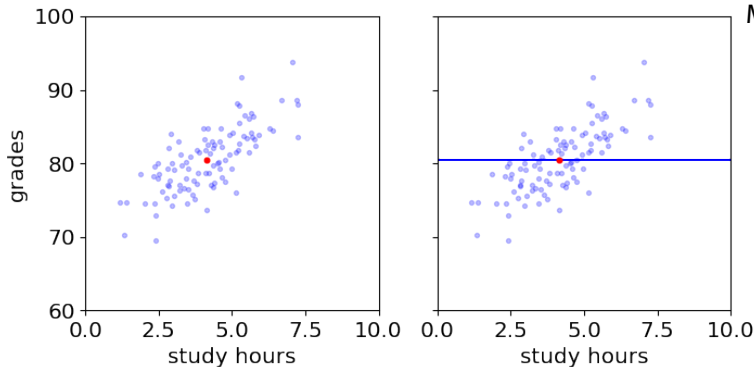
Vorhersage von Daten

Gegeben ist ein Datensatz mit Lernzeiten ('study hours') und erreichten Abschlussnoten ('grades'). Gesucht ist ein Modell, das die Daten gut beschreibt und erlaubt, die Note anhand der Lernzeit vorherzusagen.



Vorhersage von Daten

Gegeben ist ein Datensatz mit Lernzeiten ('study hours') und erreichten Abschlussnoten ('grades'). Gesucht ist ein Modell, das die Daten gut beschreibt und erlaubt, die Note anhand der Lernzeit vorherzusagen.

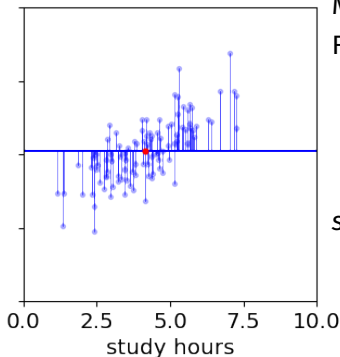
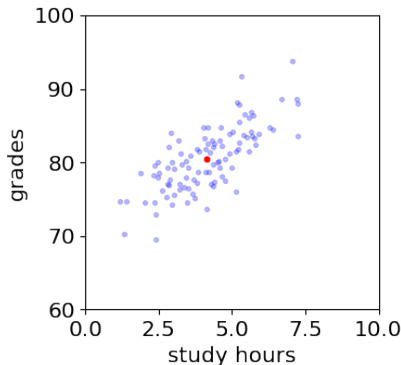


Modell 1: Mittelwert

$$\bar{y} = \frac{1}{n} \sum_{i=1} y_i = 80.5$$

Vorhersage von Daten

Gegeben ist ein Datensatz mit Lernzeiten ('study hours') und erreichten Abschlussnoten ('grades'). Gesucht ist ein Modell, das die Daten gut beschreibt und erlaubt, die Note anhand der Lernzeit vorherzusagen.



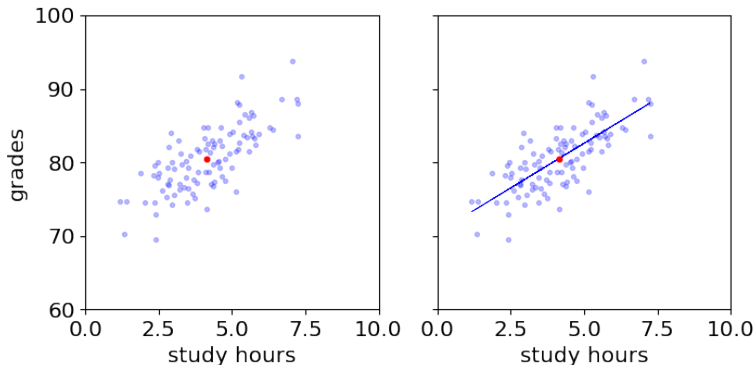
Modell 1: Mittelwert
Fehler: Standardabweichung

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = 80.5$$

$$s_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} = 4.4$$

Vorhersage von Daten

Gegeben ist ein Datensatz mit Lernzeiten ('study hours') und erreichten Abschlussnoten ('grades'). Gesucht ist ein Modell, das die Daten gut beschreibt und erlaubt, die Note anhand der Lernzeit vorherzusagen.

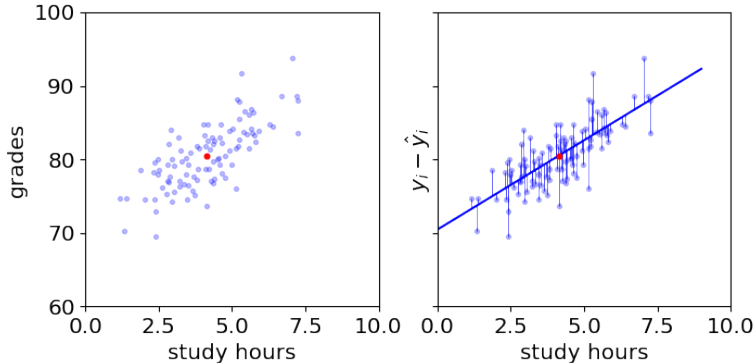


Modell 2: Lineare Regression

$$\hat{y} = b * x + a$$

Vorhersage von Daten

Gegeben ist ein Datensatz mit Lernzeiten ('study hours') und erreichten Abschlussnoten ('grades'). Gesucht ist ein Modell, das die Daten gut beschreibt und erlaubt, die Note anhand der Lernzeit vorherzusagen.



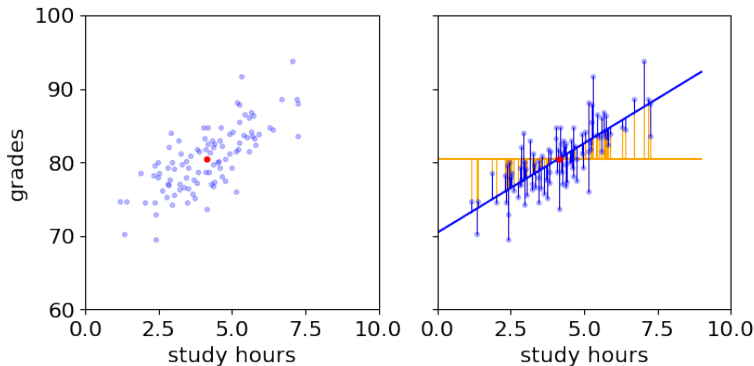
Modell 2: Lineare Regression
Fehler: Residuen

$$\hat{y} = b * x + a$$

$$s_{y\hat{y}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} = 2.9$$

Vorhersage von Daten

Gegeben ist ein Datensatz mit Lernzeiten ('study hours') und erreichten Abschlussnoten ('grades'). Gesucht ist ein Modell, das die Daten gut beschreibt und erlaubt, die Note anhand der Lernzeit vorherzusagen.



Modell 2: Lineare Regression
Fehler: Residuen

$$\hat{y} = b * x + a$$

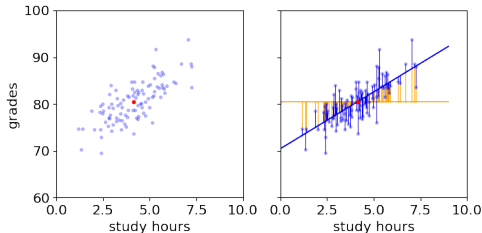
$$s_{y\hat{y}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} = 2.9$$

$$s_y = 4.4$$

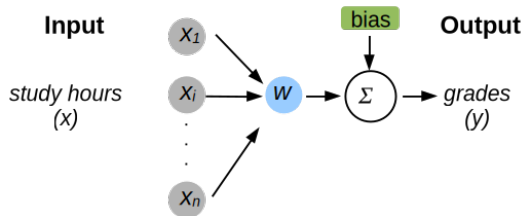
Lineare Regression und maschinelles Lernen

Gesucht ist ein Modell, das Variable (y) aus Variable (x) vorhersagt

Ziel: minimiere die Abstände zwischen der Vorhersage und den Daten



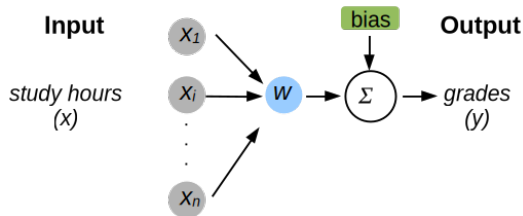
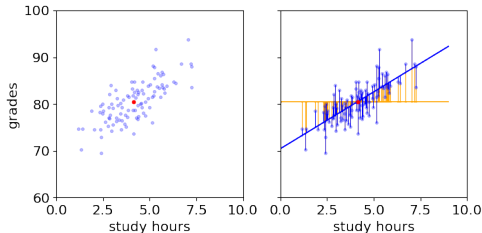
$y = w * x + bias$, wähle Startwerte für w und $bias$



Lineare Regression und maschinelles Lernen

Gesucht ist ein Modell, das Variable (y) aus Variable (x) vorhersagt

Ziel: minimiere die Abstände zwischen der Vorhersage und den Daten



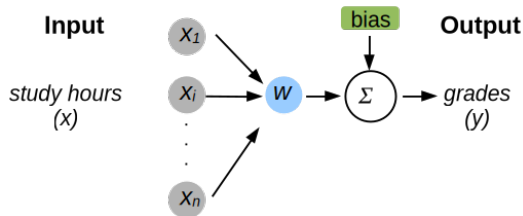
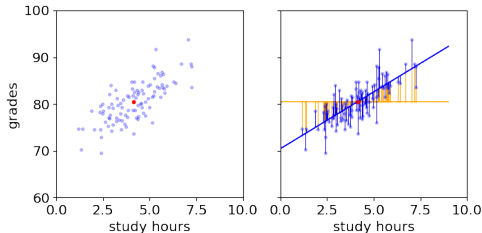
$y = w * x + \text{bias}$, wähle Startwerte für w und bias

$w = 0$, $\text{bias} = \bar{y}$, d.h. $\hat{y} = \text{bias}$ (Mittelwert), Ziel: $\min(y_i - \hat{y}_i)$

Lineare Regression und maschinelles Lernen

Gesucht ist ein Modell, das Variable (y) aus Variable (x) vorhersagt

Ziel: minimiere die Abstände zwischen der Vorhersage und den Daten



$y = w * x + bias$, wähle Startwerte für w und $bias$

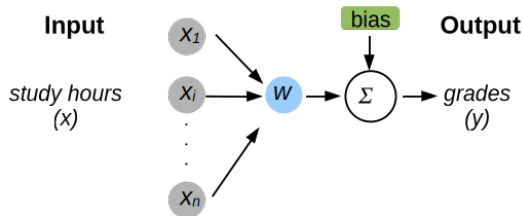
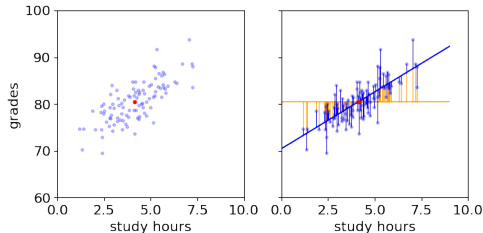
$w = 0$, $bias = \bar{y}$, d.h. $\hat{y} = bias$ (Mittelwert), Ziel: $\min(y_i - \hat{y}_i)$

Anpassung: $\hat{y} = w * x + bias$ (Lineare Regression)

Lineare Regression und maschinelles Lernen

Gesucht ist ein Modell, das Variable (y) aus Variable (x) vorhersagt

Ziel: minimiere die Abstände zwischen der Vorhersage und den Daten



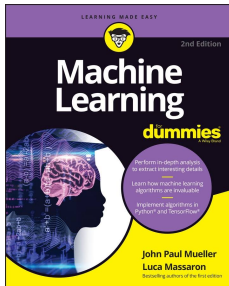
$y = w * x + \text{bias}$, wähle Startwerte für w und bias

$w = 0$, $\text{bias} = \bar{y}$, d.h. $\hat{y} = \text{bias}$ (Mittelwert), Ziel: $\min(y_i - \hat{y}_i)$

Anpassung: $\hat{y} = w * x + \text{bias}$ (Lineare Regression)

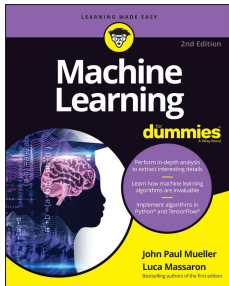
- Modell steckt in den Gewichten und der Architektur

Lineare Regression und Maschinelles Lernen



- Modell für lineare Regression ist bekannt, d.h. Fehler kann als geschlossene Form berechnet werden
 - man kann den Fehler aber auch schätzen:
 - zufällige Startwerte für die Parameter wählen
 - Vorhersage berechnen und Differenz zu den tatsächlichen Werten
- back-propagation des Fehlers und Anpassen der Parameter
- so lange bis der Fehler nicht mehr kleiner wird

Lineare Regression und Maschinelles Lernen



- Modell für lineare Regression ist bekannt, d.h. Fehler kann als geschlossene Form berechnet werden
 - man kann den Fehler aber auch schätzen:
 - zufällige Startwerte für die Parameter wählen
 - Vorhersage berechnen und Differenz zu den tatsächlichen Werten
- back-propagation des Fehlers und Anpassen der Parameter
- so lange bis der Fehler nicht mehr kleiner wird

Einfache lineare Regression

$$y = \beta * x + \alpha$$

x, y - Abszisse, Ordinate

β, α - Anstieg, Achsenabschnitt

Modell ist Linie

Multiple lineare Regression

$$Y = \beta_1 * X_1 + \dots + \beta_i * X_i + \beta_n * X_n + \alpha$$

X_i, Y - Dimensionen im Vektorraum

β, α - Gewichte, Fehler

Modell ist Hyperebene (hyperplane)

Mensch und Maschine: Künstliche Intelligenz

1996 Garri Kasparov, damals amtierender Schachweltmeister, verliert gegen Schachcomputer Deep Blue



Mensch und Maschine: (Generative) Künstliche Intelligenz

10. Februar 2016, 10:18 Uhr Künstliche Intelligenz

**Dieser Computer malt fast so schön
wie ein Mensch**



(Quelle:

<https://www.sueddeutsche.de/digital/kuenstliche-intelligenz-dieser-computer-malt-fast-so-schoen-wie-ein-mensch-1.2856388>)

Mensch und Maschine: (Generative) Künstliche Intelligenz

10. Februar 2016, 10:18 Uhr Künstliche Intelligenz

**Dieser Computer malt fast so schön
wie ein Mensch**



(Quelle:

<https://www.sueddeutsche.de/digital/kuenstliche-intelligenz-dieser-computer-malt-fast-so-schoen-wie-ein-mensch-1.2856388>)

Links: Claude Monets "Antibes, Rechts: ein von einem Algorithmus berechnetes Bild

ELIZA

```
Welcome to

      EEEEE LL      IIII 2222222  AAAAA
      EE   LL      II   ZZ   AA  AA
      EEEE  LL      II   222  AAAAAA
      EE   LL      II   ZZ   AA  AA
      EEEEE LLLLLL IIII 2222222  AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:   █
```

(Quelle: <https://de.wikipedia.org/wiki/ELIZA>)

- 1966 Joseph Weizenbaum
- "Großmutter" heutiger Chatbots
- Unterhaltung in natürlicher Sprache
- Ein- und Ausgabe in Form von Sätzen in der Konsole
- festgelegte Themen, die vorher programmiert wurden

- Cover story: ELIZA ist eine Psychiaterin, die weit weg ist und mit der man sich nur über Telegramme unterhalten kann

Smalltalk mit ELIZA

Patientin: Männer sind alle gleich.

ELIZA: Inwiefern?

Patientin: Immer wollen sie irgendwas von uns.

ELIZA: Können Sie ein konkretes Beispiel nennen?

Patientin: Nun ja, mein Freund wollte, dass ich zu Ihnen komme.

ELIZA: Ihr Freund wollte, dass Sie zu mir kommen?

Patientin: Er sagt, dass ich so oft deprimiert bin.

ELIZA: Es tut mir Leid zu hören, dass Sie so oft deprimiert sind.

Patientin: Es ist leider wahr. Ich bin unglücklich.

ELIZA: Denken Sie, dass ich Ihnen helfen kann, nicht unglücklich zu sein?

Patientin: Ich brauche Hilfe, so viel steht fest.

ELIZA: Was würde es für Sie bedeuten, Hilfe zu bekommen?

Patientin: Vielleicht könnte ich lernen, besser mit meiner Mutter auszukommen.

ELIZA: Erzählen Sie mir mehr über Ihre Familie ...

Smalltalk mit ELIZA

Patientin: Männer sind alle gleich.

ELIZA: Inwiefern?

Patientin: Immer wollen sie irgendwas von uns.

ELIZA: Können Sie ein konkretes Beispiel nennen?

Patientin: Nun ja, mein Freund wollte, dass ich zu Ihnen komme.

ELIZA: Ihr Freund wollte, dass Sie zu mir kommen?

Patientin: Er sagt, dass ich so oft deprimiert bin.

ELIZA: Es tut mir Leid zu hören, dass Sie so oft deprimiert sind.

Patientin: Es ist leider wahr. Ich bin unglücklich.

ELIZA: Denken Sie, dass ich Ihnen helfen kann, nicht unglücklich zu sein?

Patientin: Ich brauche Hilfe, so viel steht fest.

ELIZA: Was würde es für Sie bedeuten, Hilfe zu bekommen?

Patientin: Vielleicht könnte ich lernen, besser mit meiner Mutter auszukommen.

ELIZA: Erzählen Sie mir mehr über Ihre Familie ...

- Fragen

Smalltalk mit ELIZA

Patientin: Männer sind alle gleich.

ELIZA: Inwiefern?

Patientin: Immer wollen sie irgendwas von uns.

ELIZA: Können Sie ein konkretes Beispiel nennen?

Patientin: Nun ja, mein Freund wollte, dass ich zu Ihnen komme.

ELIZA: Ihr Freund wollte, dass Sie zu mir kommen?

Patientin: Er sagt, dass ich so oft deprimiert bin.

ELIZA: Es tut mir Leid zu hören, dass Sie so oft deprimiert sind.

Patientin: Es ist leider wahr. Ich bin unglücklich.

ELIZA: Denken Sie, dass ich Ihnen helfen kann, nicht unglücklich zu sein?

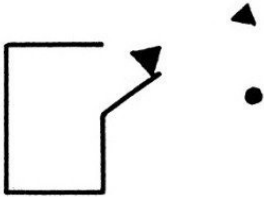
Patientin: Ich brauche Hilfe, so viel steht fest.

ELIZA: Was würde es für Sie bedeuten, Hilfe zu bekommen?

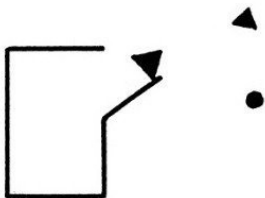
Patientin: Vielleicht könnte ich lernen, besser mit meiner Mutter auszukommen.

ELIZA: Erzählen Sie mir mehr über Ihre Familie ...

- Fragen
- Schlüsselwörter: z.B. unglücklich, Mutter, ...



Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 57, 243–259.



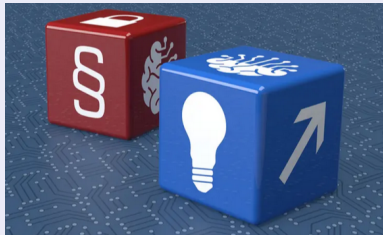
Attribution

- Begriff der Kognitionspsychologie
- Zuschreibung von Ursachen zu verschiedenen Handlungen und Verhaltensweisen
- Zuschreiben menschlicher Eigenschaften zu nicht menschlichen Wesen oder Objekten - Anthropomorphismus (altgriechisch, *ánthropos* - "Mensch", *morphē* - "Form, Gestalt")
- fundamentale menschliche Eigenart (bias)

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 57, 243–259.

Dual Use und Abschätzung von Technikfolgen

Lesen Sie bis *Schlagartige Verhaltensänderung*



- Wie vertrauenswürdig beurteilen Sie die Quelle?
- Im Artikel geht es auch um Dual Use bzw. Technikfolgenabschätzung. Was ist damit gemeint. Welche Beispiele werden genannt?

<https://www.telepolis.de/features/>

Vertrauenswürdigkeit von Quellen

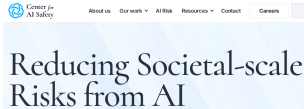
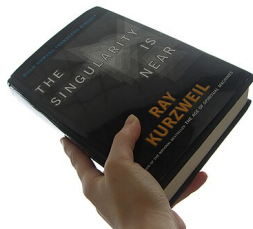
- wissenschaftliche Fachzeitschriften: Peer-Review, Autorinnen ohne finanzielle Anreize (Wissenschaftlerinnen)
- Fachzeitschriften z.B. c't (computing today), Heise Gruppe, Computerwoche: International Data Group, kostenpflichtig, freiberufliche Reporterinnen - finanzielle Anreize, Redaktionelle Begutachtung
- Tages-/Wochenzeitungen z.B. Die Zeit: DvH Medien und Verlagsgruppe Georg von Holtzbrinck, festangestellte Mitarbeitende, kostenpflichtig, breites Themenspektrum, Informations- und Nachrichtenorientierung
- Boulevardzeitung z.B. BILD: Axel-Springer-Verlag, Tageszeitung, breites Themenspektrum, Mischform von Sensations- und Nachrichtenorientierung
- Telepolis: breites Themenspektrum, Finanzierung: freiwilliges UnterstützermodeLL, Heise Gruppe, kostenfrei, freie Autorinnen

Begriffe

Singularität

hypothetischer zukünftiger Zeitpunkt,
an dem künstliche Intelligenz (KI) die
menschliche Intelligenz übertrifft

(nach https://de.wikipedia.org/wiki/Technologische_Singularit%C3%A4t)



One-sentence-statement Mission: Reducing Societal-scale Risks from AI

The Center for AI Safety (CAIS — pronounced 'case') is a San Francisco-based research and field-building nonprofit. We believe that artificial intelligence (AI) has the potential to profoundly benefit the world, provided that we can develop and use it safely.

(<https://www.safe.ai/>)

Dual Use

- Mehrfachnutzung von Waren, Gütern, Algorithmen, Ergebnissen, ...
- Folgen und “Nutzen” zu Zeiten der Entwicklung nicht abzuschätzen
“Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.”

Dual Use

- Mehrfachnutzung von Waren, Gütern, Algorithmen, Ergebnissen, ...
- Folgen und “Nutzen” zu Zeiten der Entwicklung nicht abzuschätzen
“Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.”
- Führen diese KI Systeme lediglich Befehle aus? Verstehen sie, was sie tun?

Dual Use

- Mehrfachnutzung von Waren, Gütern, Algorithmen, Ergebnissen, ...
- Folgen und “Nutzen” zu Zeiten der Entwicklung nicht abzuschätzen
“Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.”
- Führen diese KI Systeme lediglich Befehle aus? Verstehen sie, was sie tun?
- Generative KI - artificial general intelligence, Superintelligenz

Dual Use

- Mehrfachnutzung von Waren, Gütern, Algorithmen, Ergebnissen, ...
- Folgen und “Nutzen” zu Zeiten der Entwicklung nicht abzuschätzen
“Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.”
- Führen diese KI Systeme lediglich Befehle aus? Verstehen sie, was sie tun?
- Generative KI - artificial general intelligence, Superintelligenz
- nicht angemessen, deren Verhalten nur als Ausführen von Befehlsfolgen zu betrachten

Dual Use

- Mehrfachnutzung von Waren, Gütern, Algorithmen, Ergebnissen, ...
- Folgen und “Nutzen” zu Zeiten der Entwicklung nicht abzuschätzen
“Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.”
- Führen diese KI Systeme lediglich Befehle aus? Verstehen sie, was sie tun?
- Generative KI - artificial general intelligence, Superintelligenz
- nicht angemessen, deren Verhalten nur als Ausführen von Befehlsfolgen zu betrachten
- Risiken, die von solchen Systemen ausgehen können, sind unabhängig davon, ob man diesen Systemen Eigenschaften wie Verstehen oder Bewusstsein zuordnet oder nicht

Literatur

1. LeCun, Y., Bengio Y. and Hinton, G. (2015). Deep Learning. *Nature*, 521, 436-445.
2. Detlev Frick, Andreas Gadatsch, Jens Kaufmann, Birgit Lankes, Christoph Quix, Andreas Schmidt und Uwe Schmitz (2021). *Data Science*. Wiesbaden: Springer.
3. Geirhos,R., Rubisch,P., Michaelis, C., Bethge, M., Wichmann, F.A., Brendel, W. (2019). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness, ICLR.
4. Heider, F. and Simmel, M.(1944). An Experimental Study of Apparent Behavior. *The American Journal of Psychology*, 57(2), 243-259.
5. Hernandez-Orallo, J. et al. (2016). Computer models solving intelligence test problems: Progress and implications. *Artificial Intelligence*, 230, 74-107.