



ChatGPT Entzaubert

Joel P. Barmettler

MSc. Informatics, Artificial Intelligence

golem.de IT-NEWS FÜR PROFIS

HOME TICKER PODCAST NEWSLETTER **GOLEM PLUS** FORUM ANMELDEN

Artikel, News, ... Suchen **Mehr lesen mit Golem Plus**

KARRIEREWELT JOBS IT-FACHTRAININGS COACHINGS SPRACHKURSE KARRIERESERVICES | GOLEM-PC TECHNIK-RATGEBER DEALS

DEVIN

Neue KI programmiert, debuggt und compiliert von selbst

Devin ist ein neues KI-Modell, das sich ausschließlich auf das Programmieren konzentriert. Es arbeitet ähnlich wie ein Mensch.

[in Pocket speichern](#) [merken](#) [teilen](#) 14. März 2024, 11:10 Uhr, Oliver Nickel



(Bild: KI-generiert durch Copilot Designer/Dall-E)

Devin kann schon eine Menge Aufgaben bewältigen und Apps programmieren.

Das Start-up Cognition Labs hat einige Zeit lang an einem KI-Assistenten speziell für Programmieraufgaben gearbeitet. Das Ergebnis ist Devin [↗](#), ein Softwareentwicklungstool, welches mittels künstlicher Intelligenz Prompts versteht und selbstständig passenden Code dazu generiert. Die Software soll dabei wie ein menschlicher Softwareentwickler arbeiten, wie kurze Demonstrationen zeigen.

Stellenmarkt
Senior Consultant / Auditor Informationssicherheit (m/w/d)
operational services GmbH & Co. KG,
verschiedene Standorte

In einem Beispiel soll Devin etwa die Leistung von Facebooks Llama-Sprachmodell evaluieren. Devin interpretiert zunächst die per Textprompt gestellte Anweisung und erstellt sich einen To-do-Plan. In einem separaten Fenster können Nutzer stets sehen, welche Schritte das

Will Cognition Labs' Devin AI make software developers' jobs redundant?

Devin is capable of solving complex software development problems without human intervention and developers from around the world are discussing its potential.

Tech • 4 min read

Reported by: Anirudh Trivedi

Follow:    



WSJ Barron's MarketWatch IBD Buy Side from WSJ

Asia Dow 3681.32 -2.24% ↓ Nikkei 38471.20 -1.94% ↓ Hang Seng 16221.91 -2.28% ↓ Shanghai 3012.01 -1.48% ↓ BSE Sensex 72931.53 -0.64% ↓ Singapore 3138.82 -1.4

THE WALL STREET JOURNAL.

English Edition | Print Edition | Video | Audio | Latest Headlines | More

Subscribe Sign In

INTRO OFFER

Latest World Business U.S. Politics Economy **Tech** Finance Opinion Arts & Culture Lifestyle Real Estate Personal Finance Health Style Sports

EXCLUSIVE ARTIFICIAL INTELLIGENCE

A Peter Thiel-Backed AI Startup, Cognition Labs, Seeks \$2 Billion Valuation

Funding round could increase startup's valuation nearly sixfold in a matter of weeks, reflecting AI frenzy

By *Berber Jin* [Follow](#)

March 30, 2024 3:07 pm ET

Share Resize | Listen (2 min)

TAP FOR SOUND

reddit

r/programming

Search in r/programming

Home Popular All

RECENT

- r/AskFeminists
- r/AskWomen

COMMUNITIES

- Create a community
- r/agi
- r/algotrading
- r/AnimalsBeingD...
- r/AnimalsBeingG...
- r/antiasshole desi...
- r/AskFeminists
- r/AskHistorians
- r/AskReddit
- r/assholedesign

r/programming • 5 days ago

dh44t

Debunking Devin: "First AI Software Engineer" Upwork lie exposed!

This Is A Lie

Devin's Upwork Side Hustle

Cognition 25.4K subscribers

574 165 Share

85K views 3 weeks ago

Watch Devin make money taking on messy Upwork tasks!

593 165 Share

Add a comment

Sort by: Best

r/programming

Joined

programming

Computer Programming

6.1M Members 532 Online Top 1% Rank by size

COMMUNITY BOOKMARKS

FAQ

RULES

- Keep submissions on topic and of high quality
- No surveys
- No résumés/job listings
- /r/programming is not a support forum
- Spam

INFO

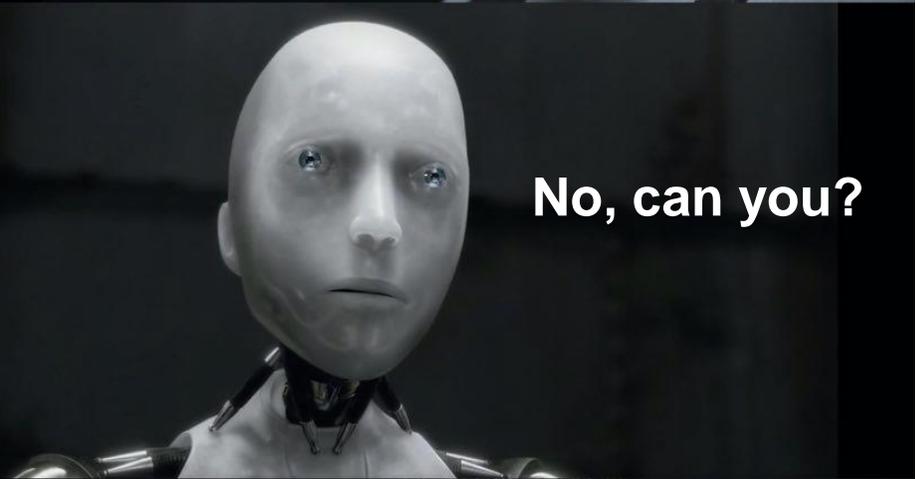
- Do you have a question? Check out [/r/learnprogramming](#), [/r/cscareerquestions](#), or [Stack Overflow](#).
- Do you have something funny to share with fellow programmers? Please take

Unterscheidung zwischen «Fact» und «Fiction» kann nur
gelingen, wenn man fundiertes Wissen hat

Wieso :(



Can you write code that is actually good?



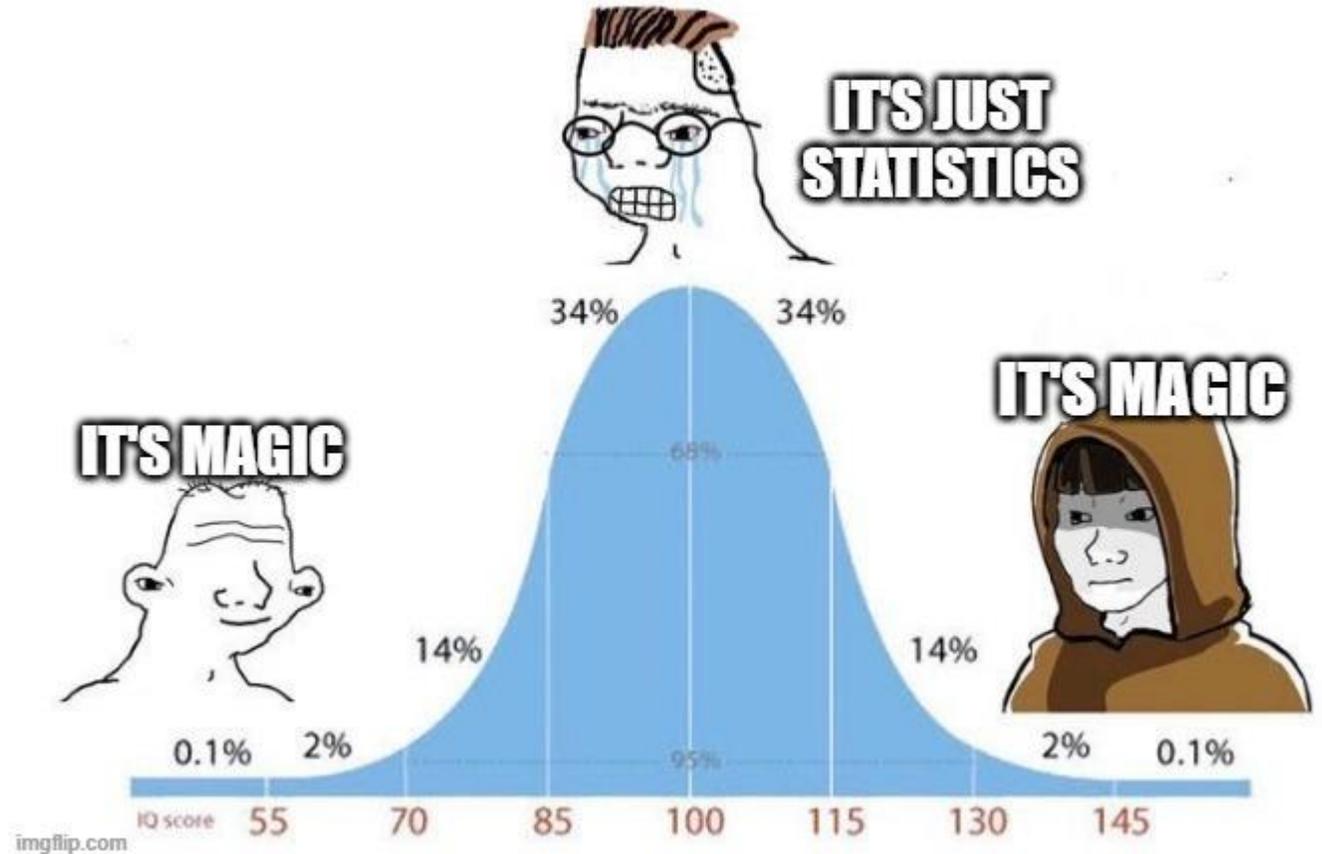
No, can you?

Inhaltsverzeichnis

- Machine Learning (**ML**) Entzaubert
- Artificial Intelligence (**AI**) Entzaubert
- Large Language Models (**LLM**) Entzaubert
- ChatGPT Entzaubert
- Konkrete Massnahmen

Inhaltsverzeichnis

- Machine Learning (**ML**) Entzaubert
- Artificial Intelligence (**AI**) Entzaubert
- Large Language Models (**LLM**) Entzaubert
- ChatGPT Entzaubert
- Konkrete Massnahmen



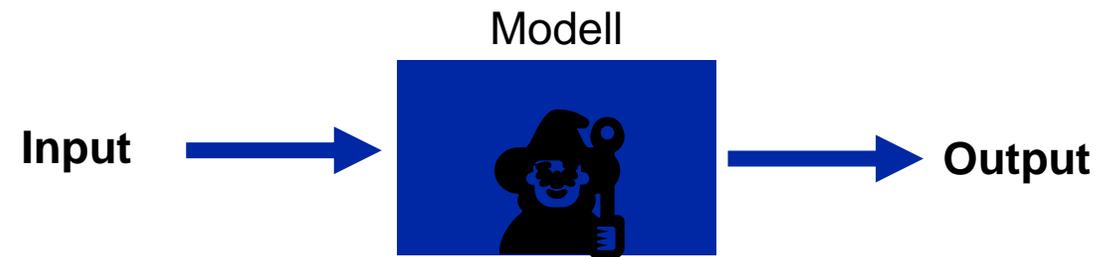
Machine Learning

Entzaubert



Was ist «Machine Learning»

🧙 Das modell lernt, vom Input auf den Output zu schliessen



Beispiel Input- und Outputdaten

🤖 Immobilienpreise

Jahr	Preis
1990	270k
1995	438k
2000	463k
2005	692k
2010	752k
2015	899k
2020	1'114k
2025	?
2030	?

Im Jahr 1995..

...kostete Immobilie durchschnittlich 438'000 CHF

Beispiel Input- und Outputdaten

🏠 Immobilienpreise

Jahr	Preis
1990	270k
1995	438k
2000	463k
2005	692k
2010	752k
2015	899k
2020	1'114k
2025	?
2030	?

Lernen

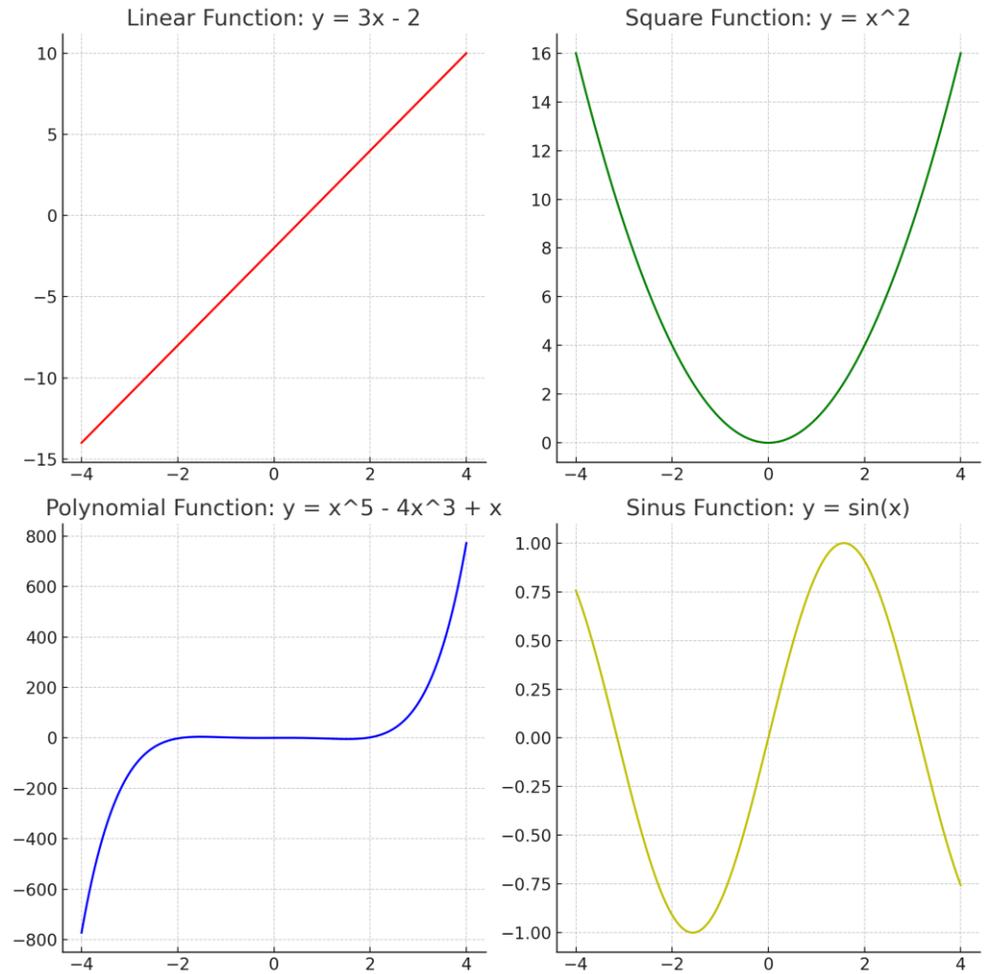
Vorhersagen

Input **Output**



Funktionen

🤖 Linear, Quadratisch, Polynomial, Sinus...



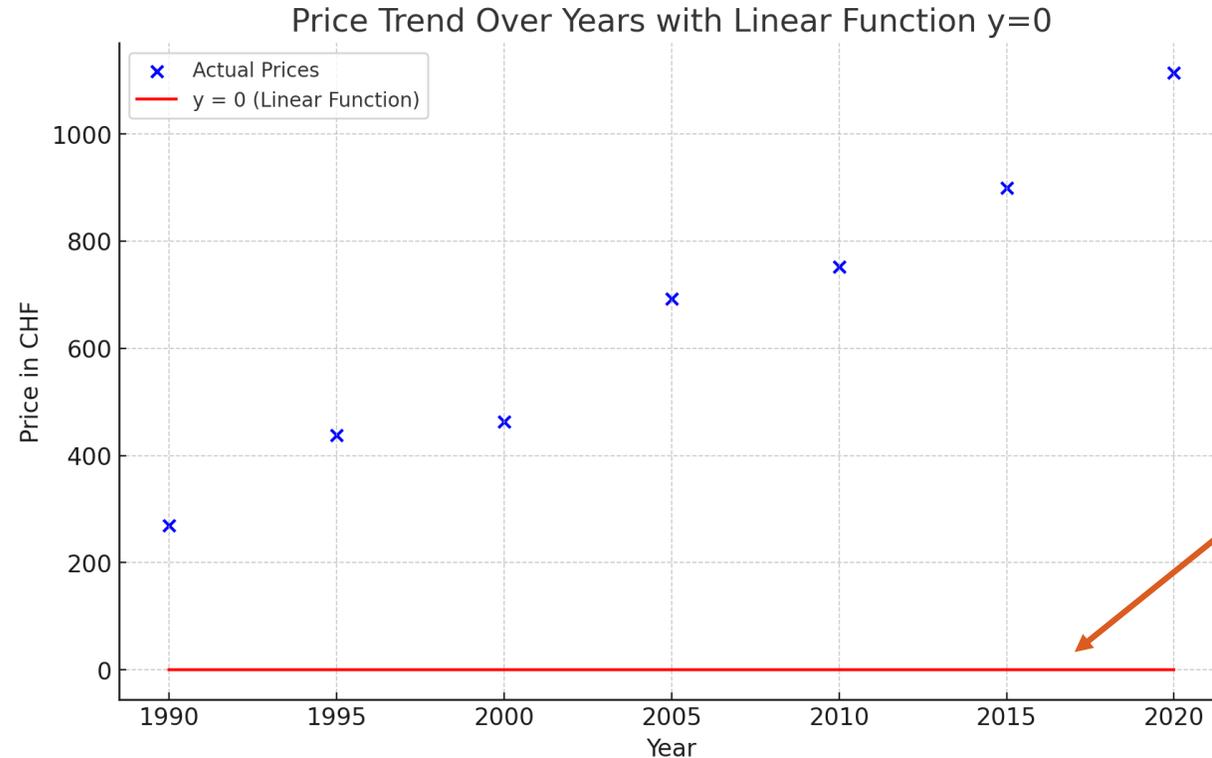
Daten

Jahr	Preis
1990	270k
1995	438k
2000	463k
2005	692k
2010	752k
2015	899k
2020	1'114k
2025	?
2030	?



Linear Regression

Jahr	Preis	a * Jahr + b
1990	270k	0
1995	438k	0
2000	463k	0
2005	692k	0
2010	752k	0
2015	899k	0
2020	1'114k	0
2025	?	0
2030	?	0



Our predicted values

Modell

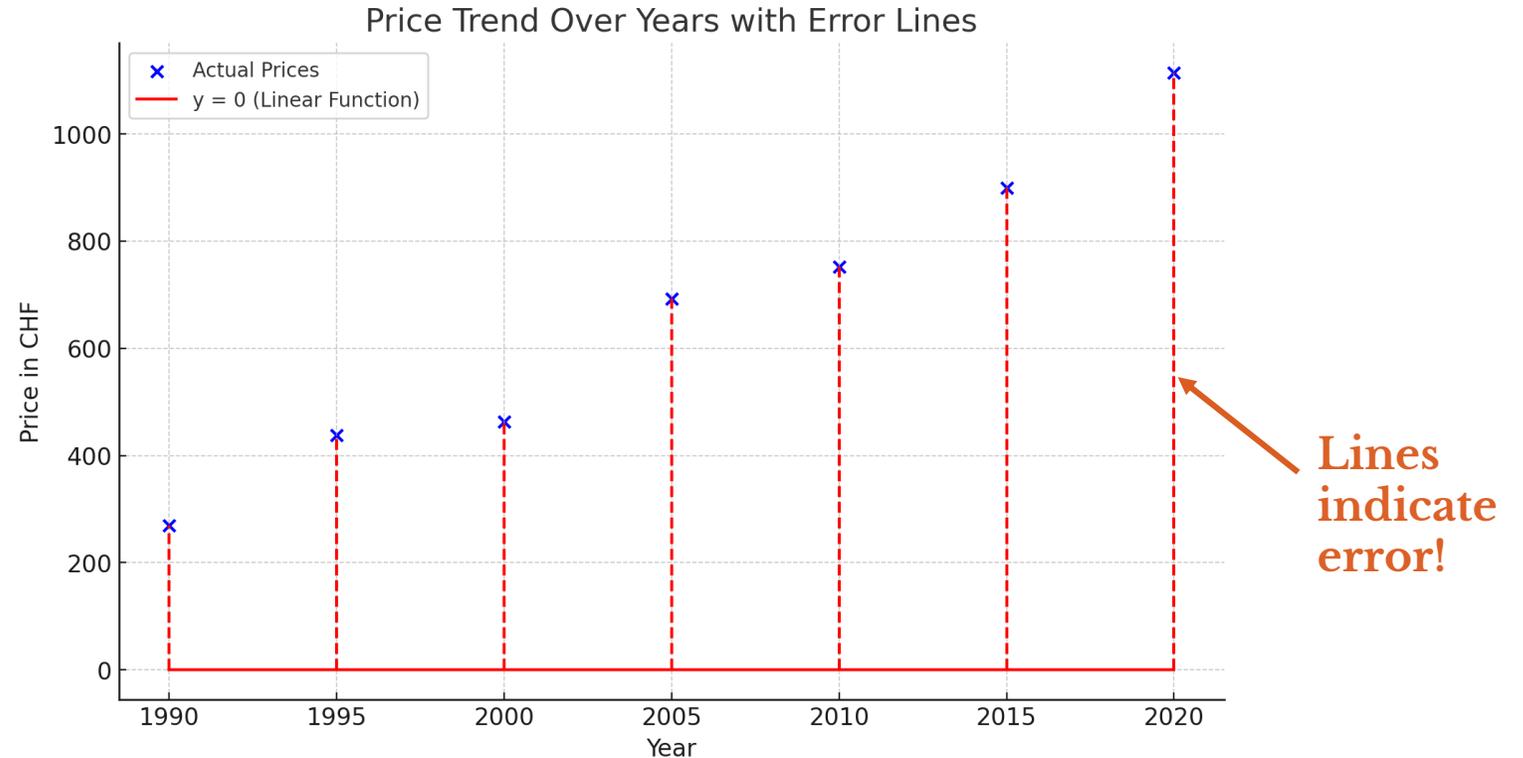
- **Name:** Linear Regression
- **Math:** $y = ax + b$
- **Alternativ:** Preis = a * Jahr + b

Parameter

- **a:** 0
- **b:** 0

Linear Regression

Jahr	Preis	a * Jahr + b
1990	270k	0
1995	438k	0
2000	463k	0
2005	692k	0
2010	752k	0
2015	899k	0
2020	1'114k	0
2025	?	0
2030	?	0



Modell

- **Name:** Linear Regression
- **Math:** $y = ax + b$
- **Alternativ:** Preis = $a * \text{Jahr} + b$

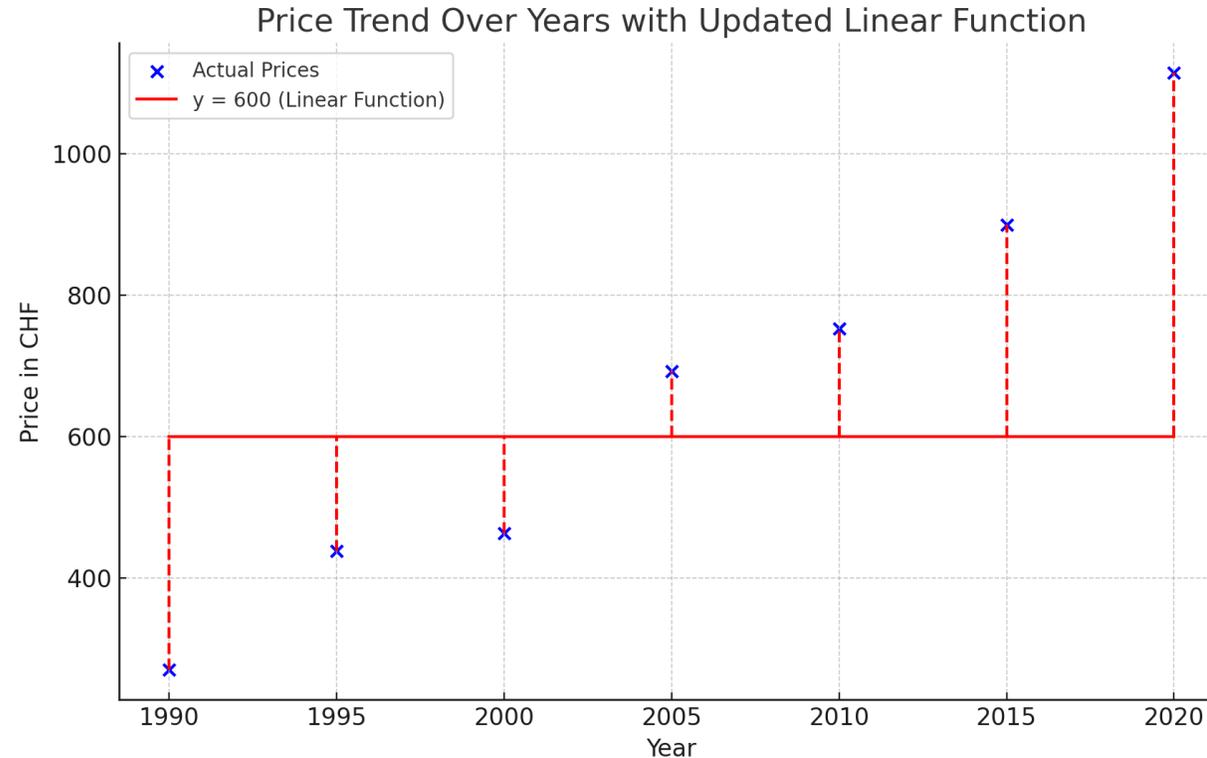
Parameter

- **a:** 0
- **b:** 0

Linear Regression

Jahr	Preis	a * Jahr + b
1990	270k	600k
1995	438k	600k
2000	463k	600k
2005	692k	600k
2010	752k	600k
2015	899k	600k
2020	1'114k	600k
2025	?	600k
2030	?	600k

Better predictions!



Updated value!

Modell

- **Name:** Linear Regression
- **Math:** $y = ax + b$
- **Alternativ:** Preis = $a * \text{Jahr} + b$

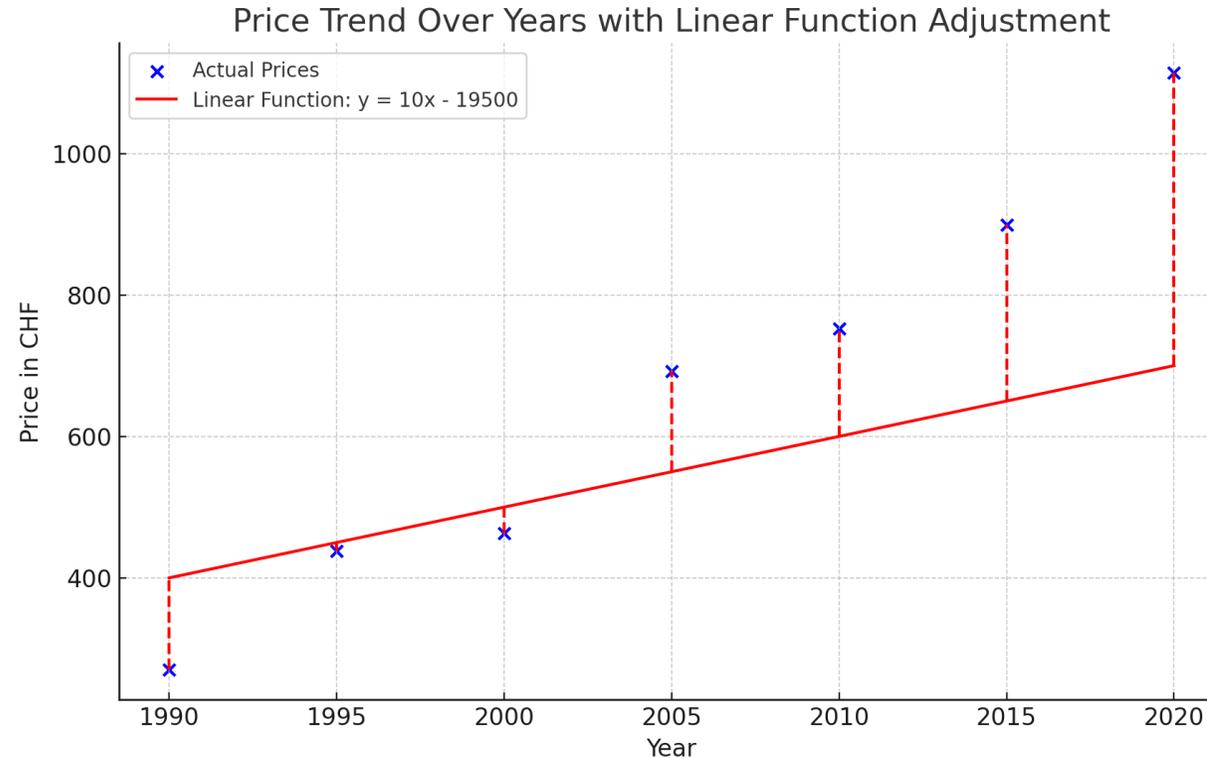
Parameter

- **a:** 0
- **b:** 600

Linear Regression

Jahr	Preis	a * Jahr + b
1990	270k	400k
1995	438k	450k
2000	463k	500k
2005	692k	550k
2010	752k	600k
2015	899k	650k
2020	1'114k	700k
2025	?	750k
2030	?	800k

Better predictions!



Updated values!

Modell

- **Name:** Linear Regression
- **Math:** $y = ax + b$
- **Alternativ:** Preis = $a * \text{Jahr} + b$

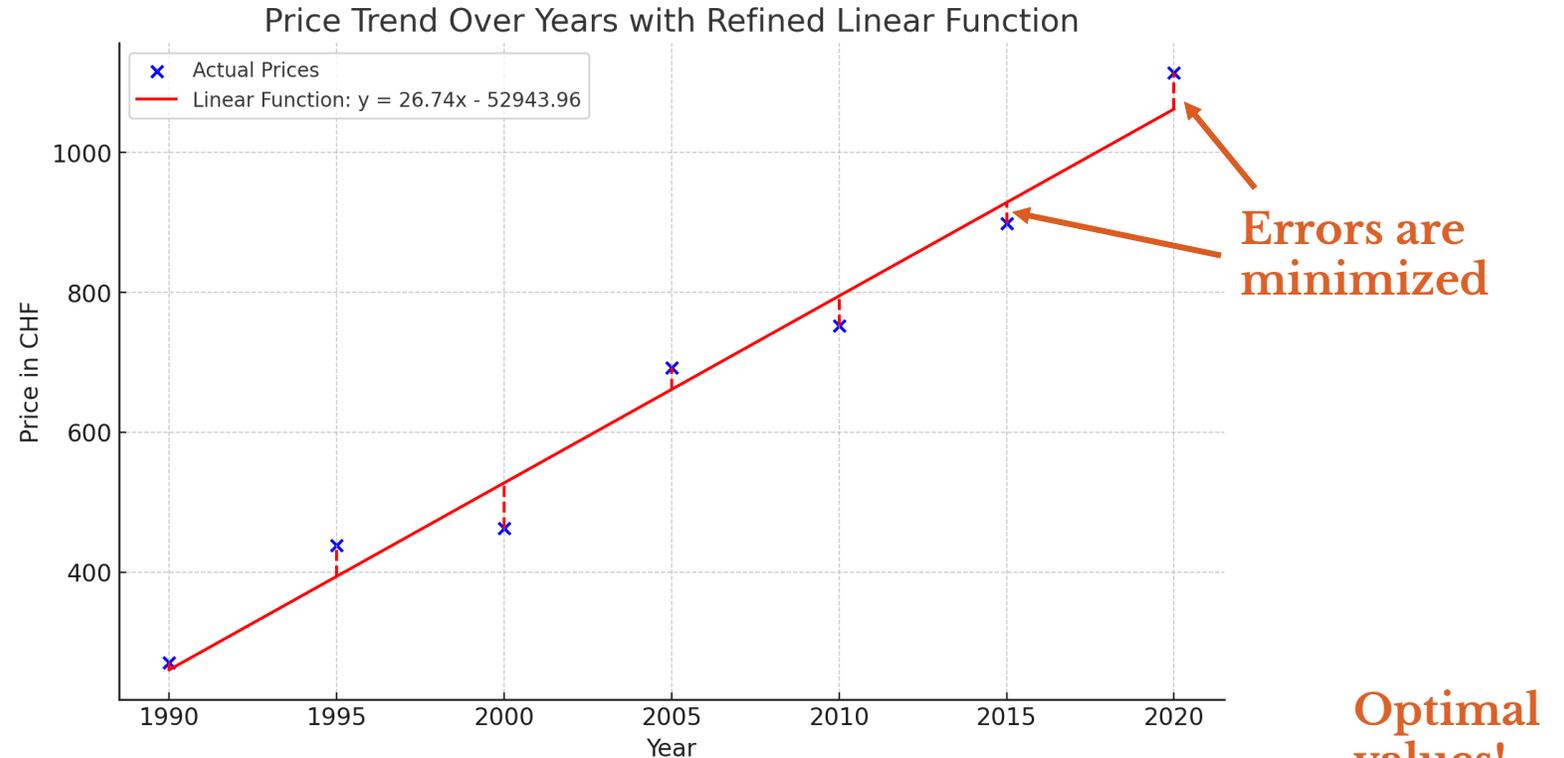
Parameter

- **a:** 10
- **b:** -19'500

Linear Regression

Jahr	Preis	a * Jahr + b
1990	270k	260k
1995	438k	393k
2000	463k	527k
2005	692k	661k
2010	752k	794k
2015	899k	928k
2020	1'114k	1'062k
2025	?	1'195k
2030	?	1'329k

Best predictions!



Modell

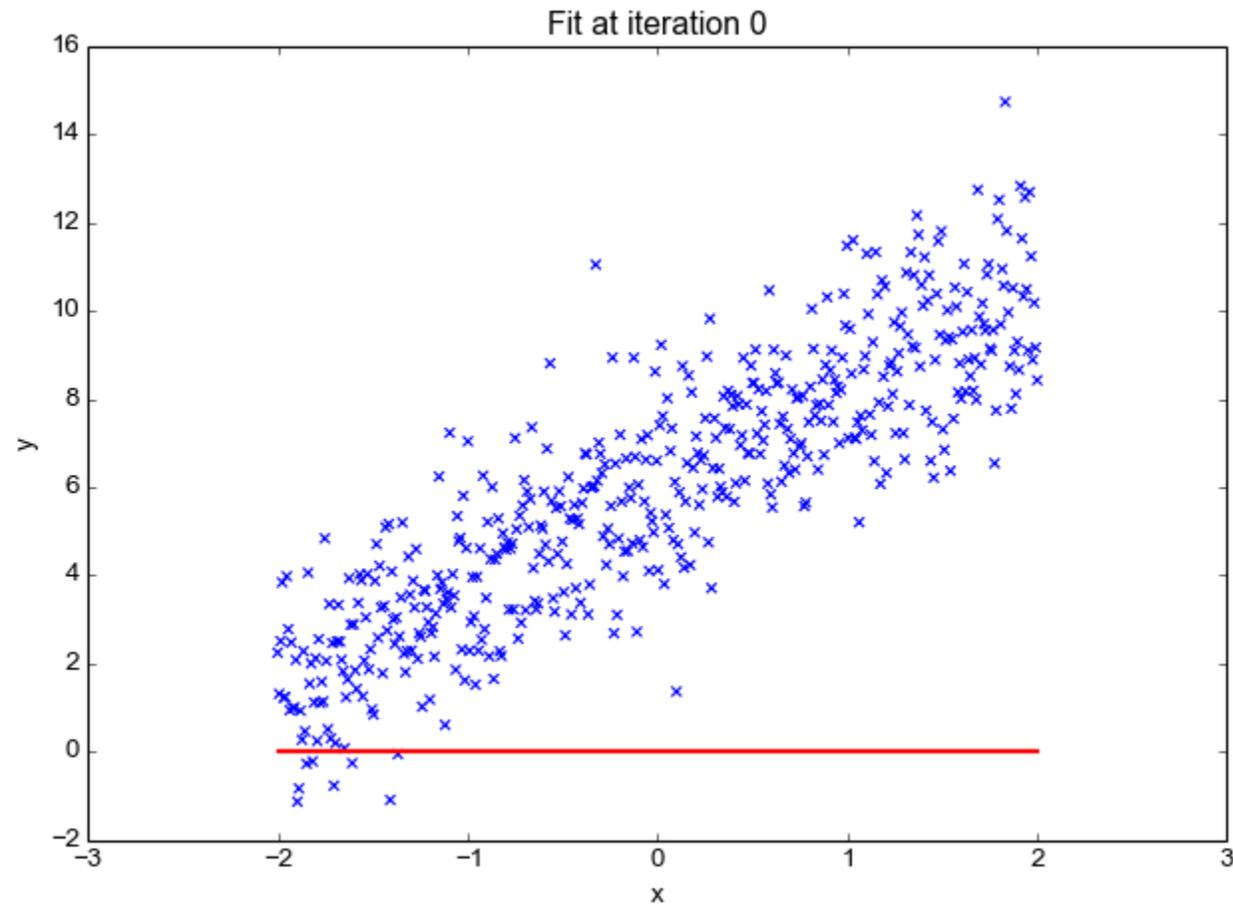
- **Name:** Linear Regression
- **Math:** $y = ax + b$
- **Alternativ:** Preis = **a** * Jahr + **b**

Parameter

- **a:** 26.7357142857142
- **b:** -52943.9642857142

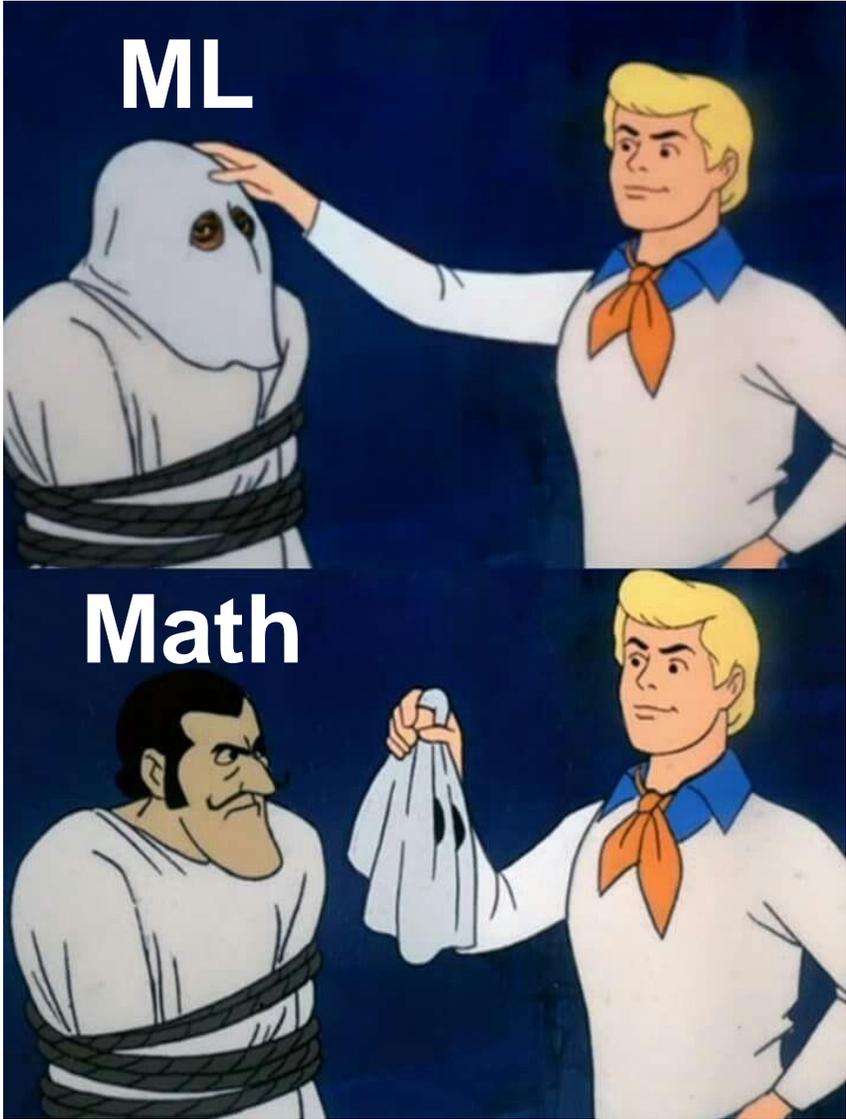
Training

🤖 Iteratives suchen von a und b



Ein Machine-Learning Modell zu trainieren bedeutet, die richtigen Funktionsparameter zu finden

Machine Learning



Lineares Regressionsmodell

 Model vs. Parameter

Architektur

- **Name:** Linear Regression
- **Math:** $y = ax + b$
- **Alternativ:** Preis = **a** * Jahr + **b**

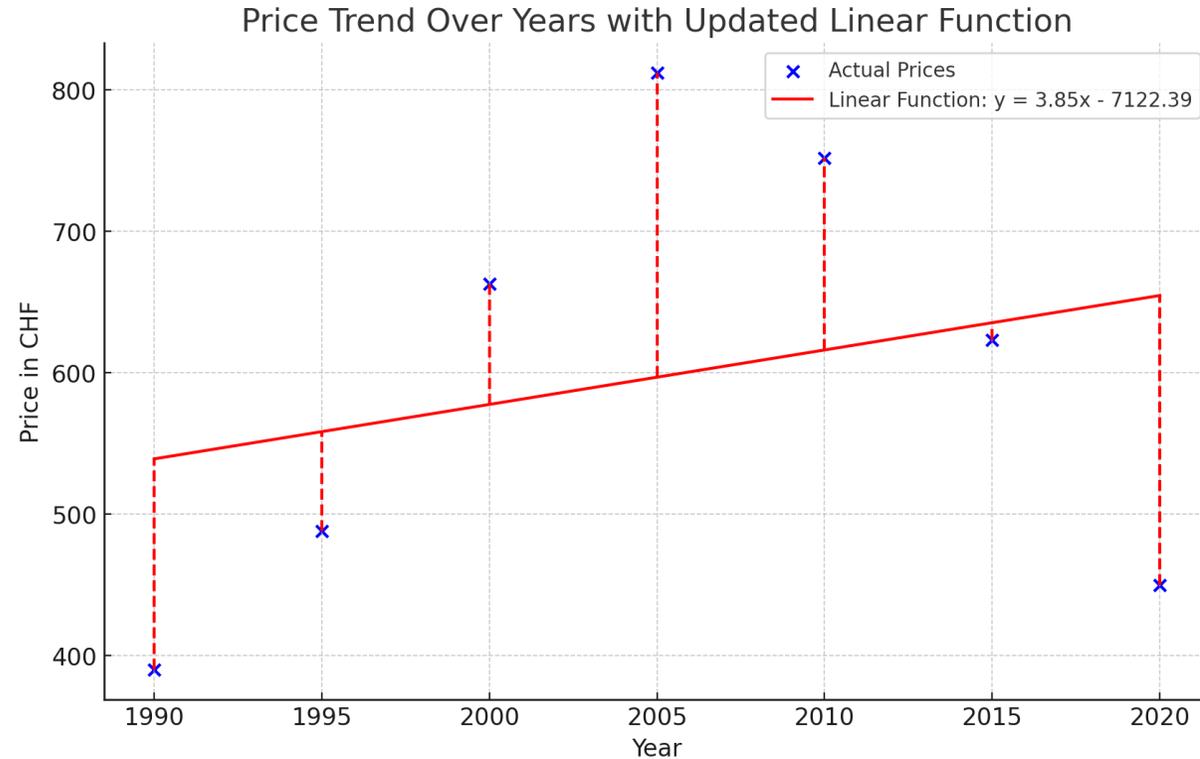
Gelernte Parameter

- **a:** 26.7357142857142
- **b:** -52943.9642857142



Keine lineare Korrelation

Jahr	Preis	a * Jahr + b
1990	390k	539k
1995	488k	558k
2000	663k	577k
2005	812k	596k
2010	752k	616k
2015	623k	635k
2020	450k	654k
2025	?	673k
2030	?	693k



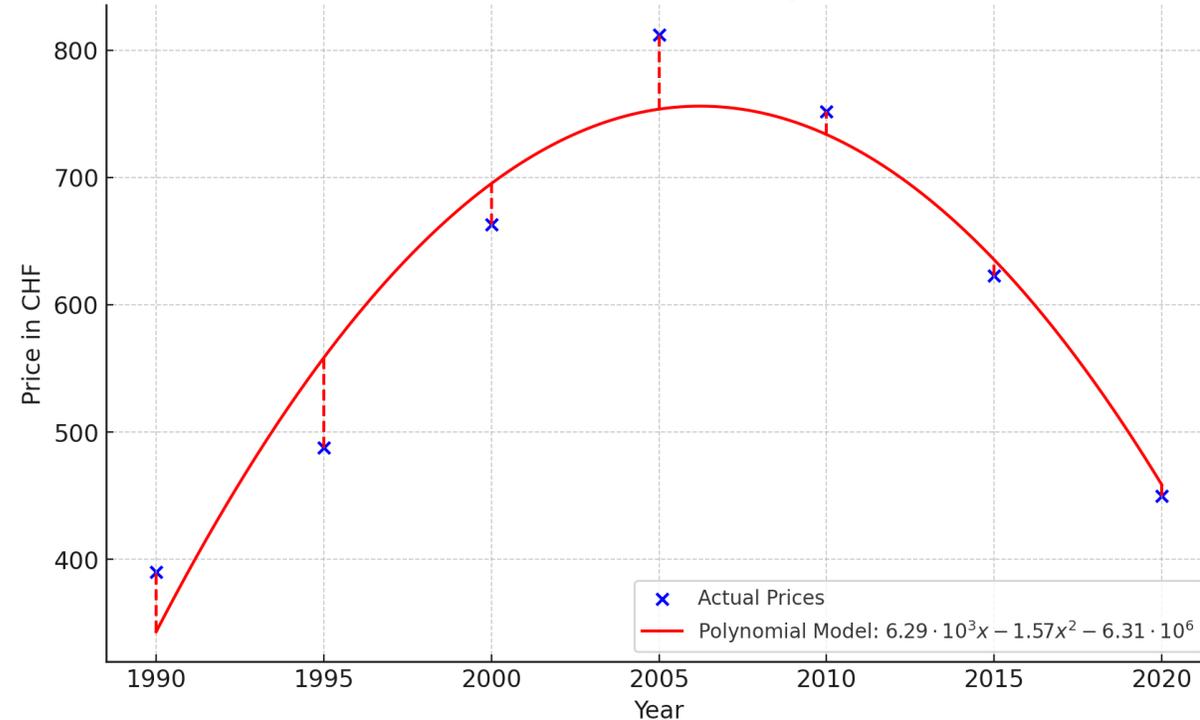
Daten sind nicht linear!

Keine lineare Korrelation

Jahr	Preis	$a \cdot \text{Jahr}^2 + b \cdot \text{Jahr} + c$
1990	390k	343k
1995	488k	558k
2000	663k	695k
2005	812k	753k
2010	752k	733k
2015	623k	635k
2020	450k	458k
2025	?	203k
2030	?	-130k

Negative Preise?!

Price Trend Over Years with Polynomial Model



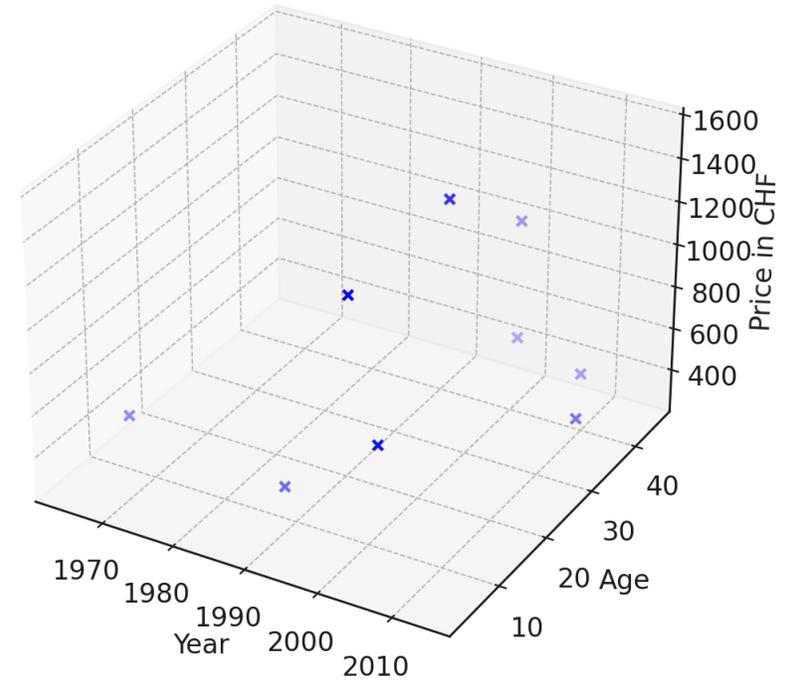
Nicht-Linears Modell

Der Skill eines guten ML-Engineers ist es, das richtige Modell für die Daten zu wählen

Viele Dimensionen

Jahr	Zim	Alter	Preis
1998	1.5	45	988k
1964	3.5	13	367k
2014	2.5	32	463k
2000	1.5	6	1'422k
2008	4.5	43	388k
2006	6.0	3	876k
1987	1.0	12	287k
1999	3.5	43	476k
2002	2.5	23	1'538k

3D Plot of Year, Age, and Price



Nicht teil des Plots

$$\text{Preis} = (\mathbf{a} * \text{Jahr} + \mathbf{b}) + (\text{Zim}^{\mathbf{c}} + \mathbf{d}) * (\sin(\mathbf{e} * \text{alter}) + \mathbf{f})$$

???

Künstliche Intelligenz

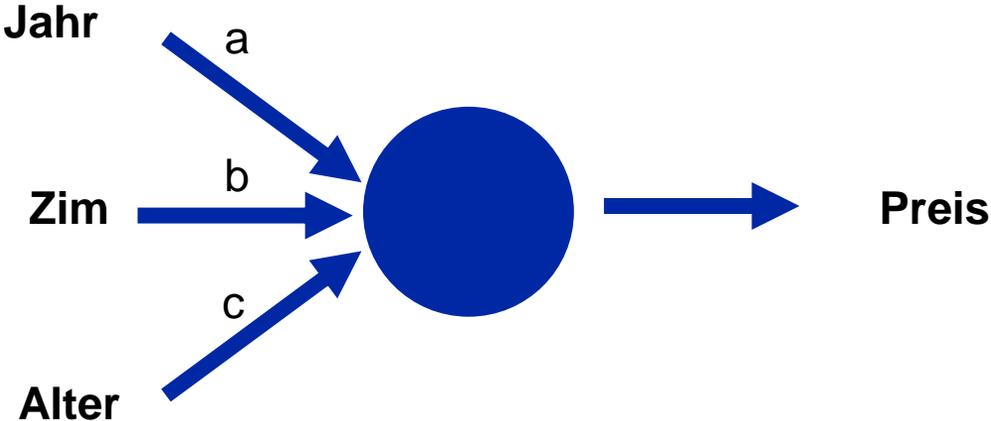
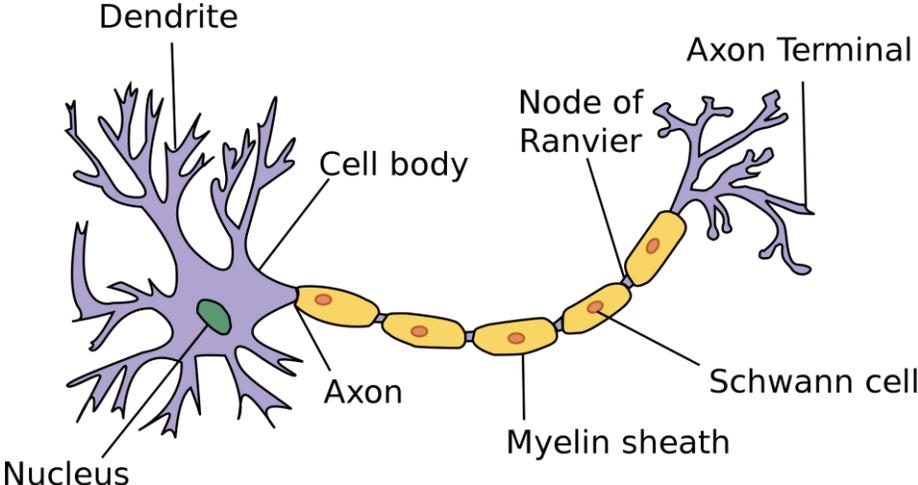
Entzaubert





Artificial Neuron

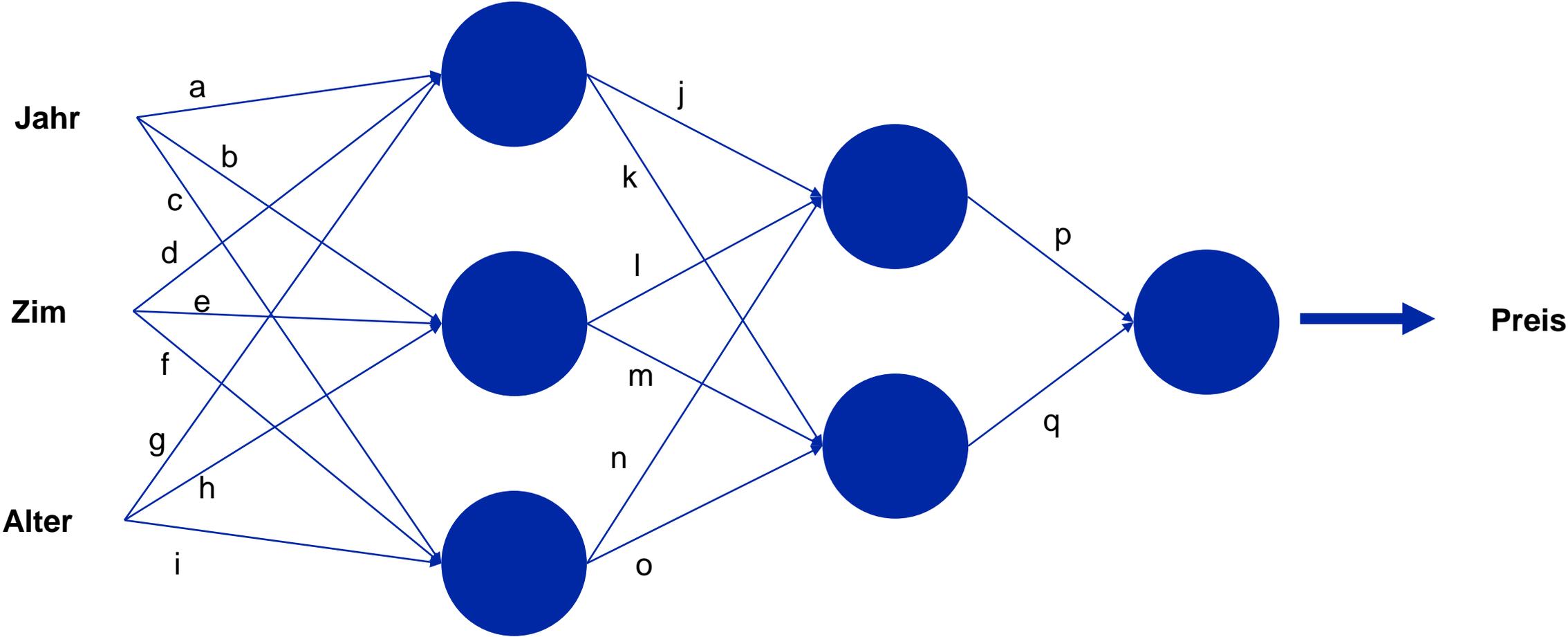
🧠 Inspiriert von biologischen Neuronen



$$\text{sig}(a * \text{Jahr} + b * \text{Zim} + c * \text{Alter}) = \text{Preis}$$

Artificial Neural Network

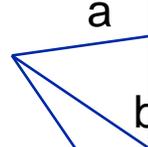
🤖 Viele Neuronen bilden Netzwerk



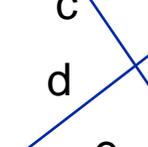
Artificial Neural Network

🤖 Viele Neuronen

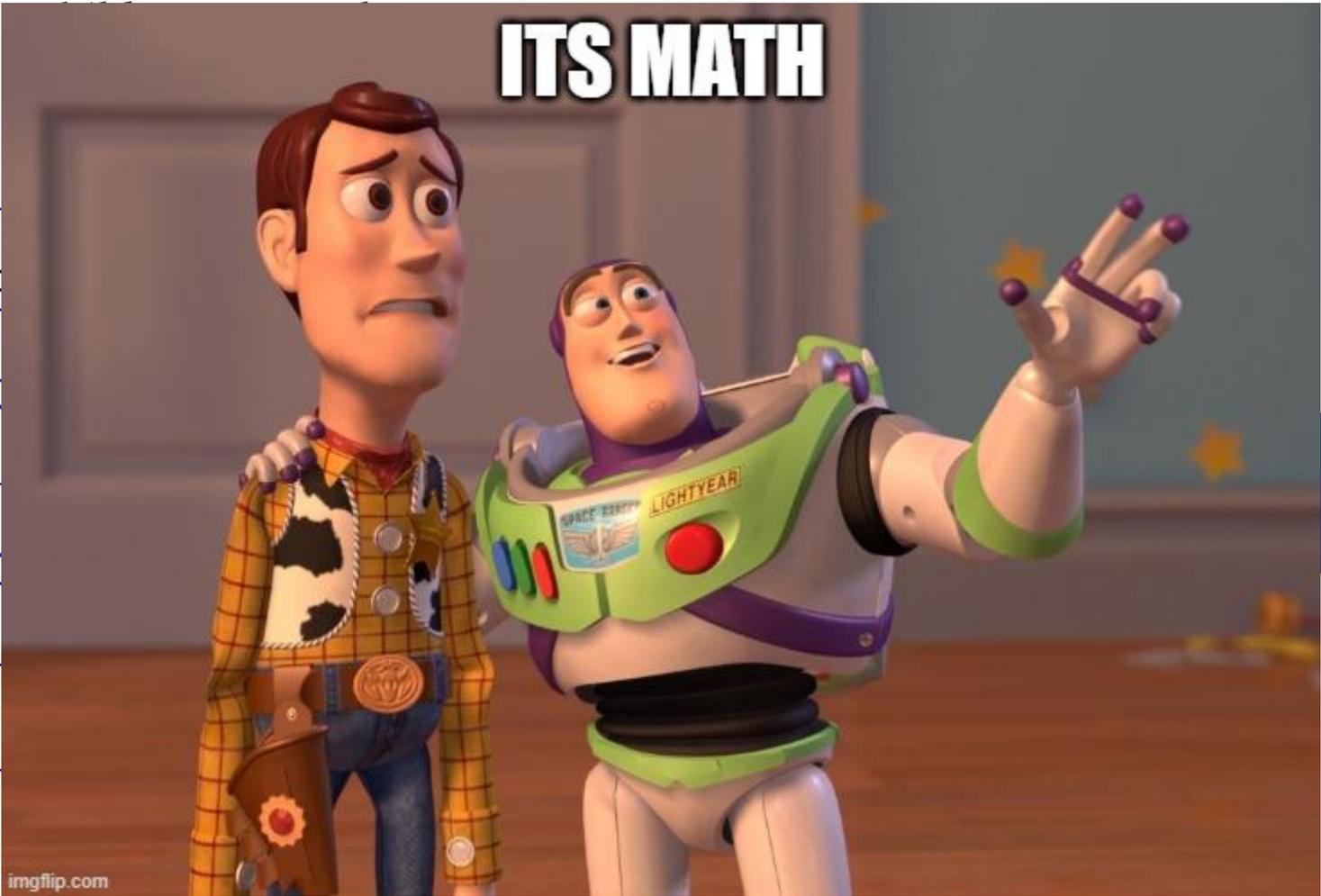
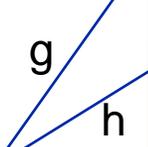
Jahr



Zim



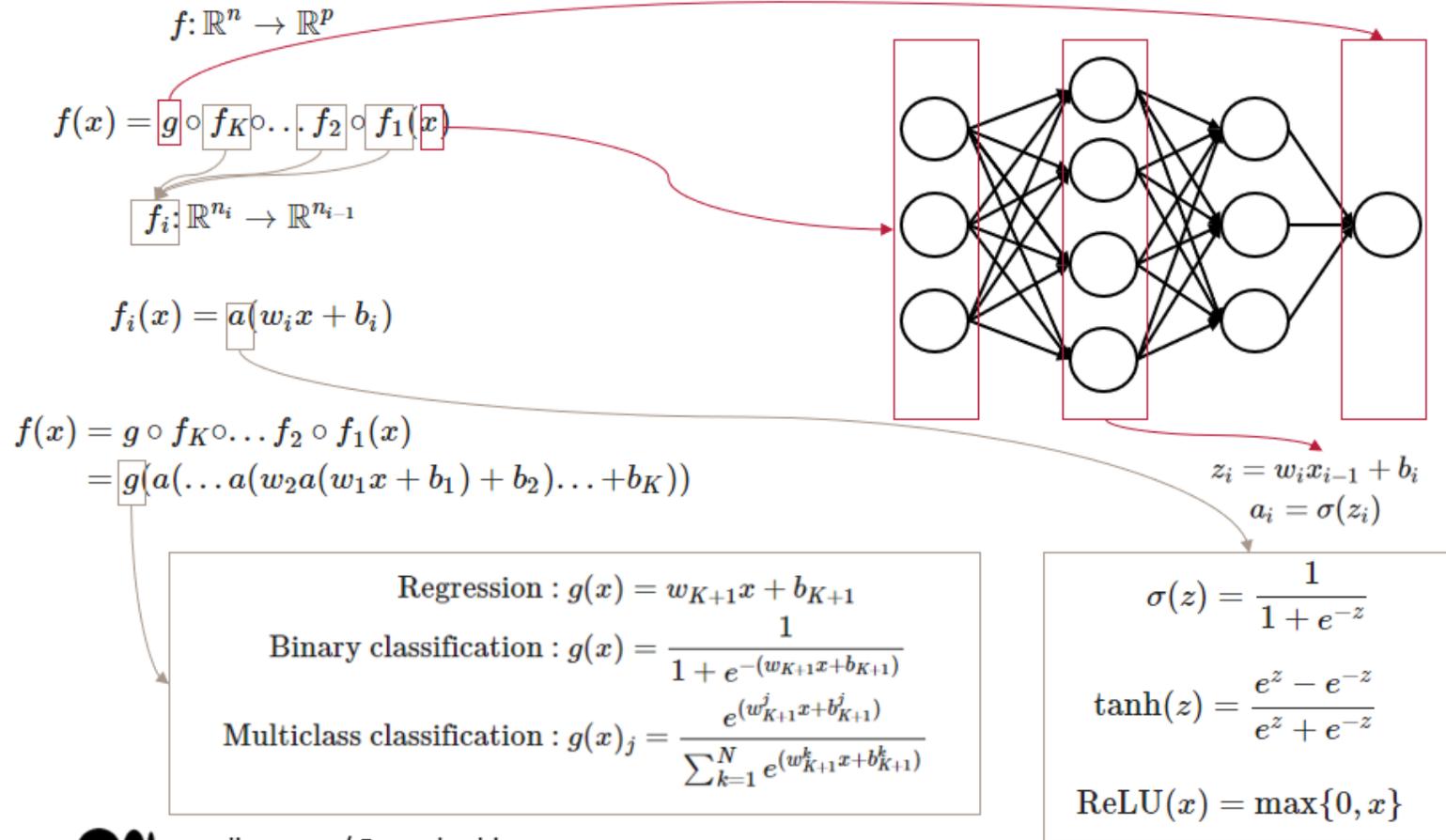
Alter



Preis

Artificial Neural Network

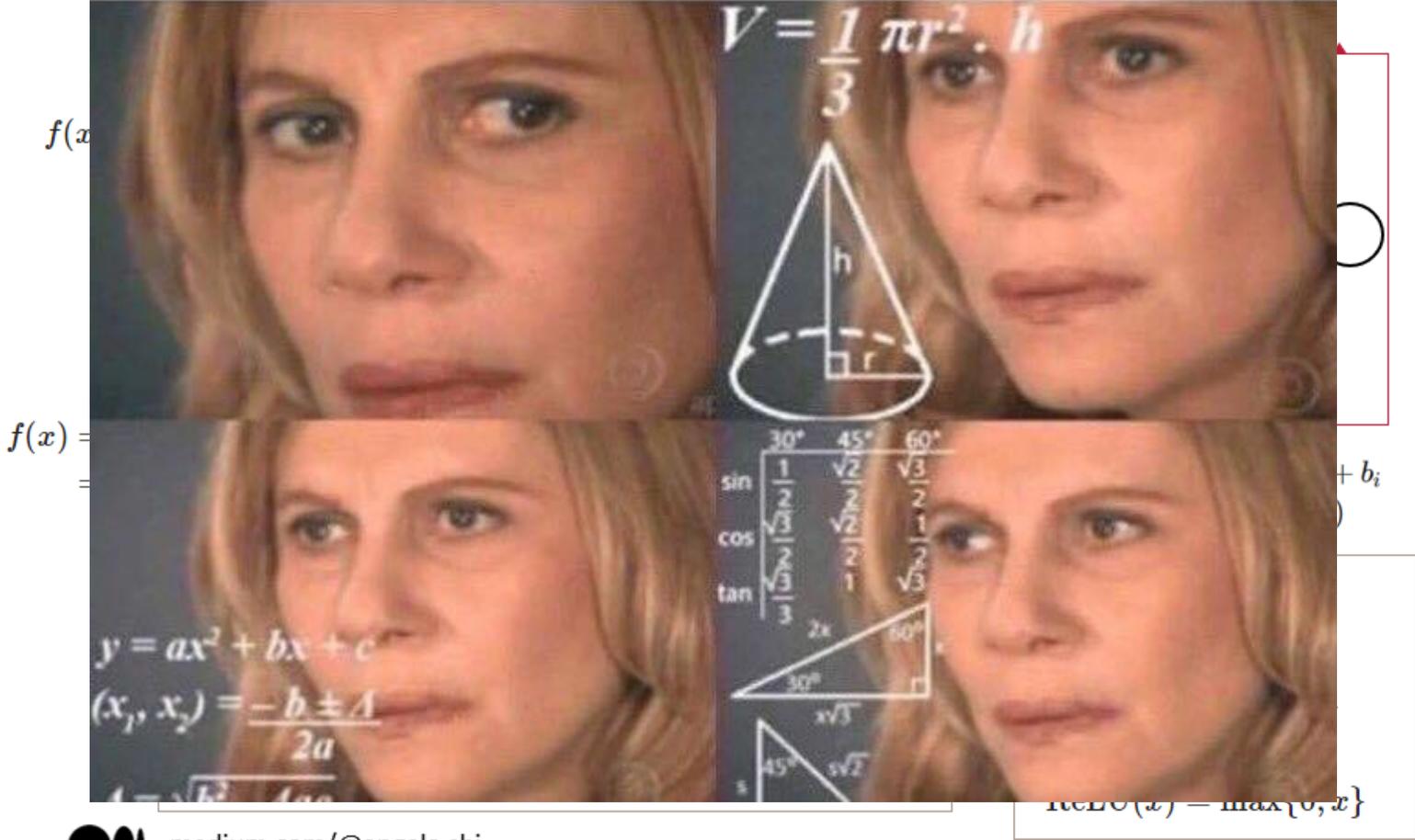
👤 ...und ihre Mathematik



👤 medium.com/@angela.shi

Artificial Neural Network

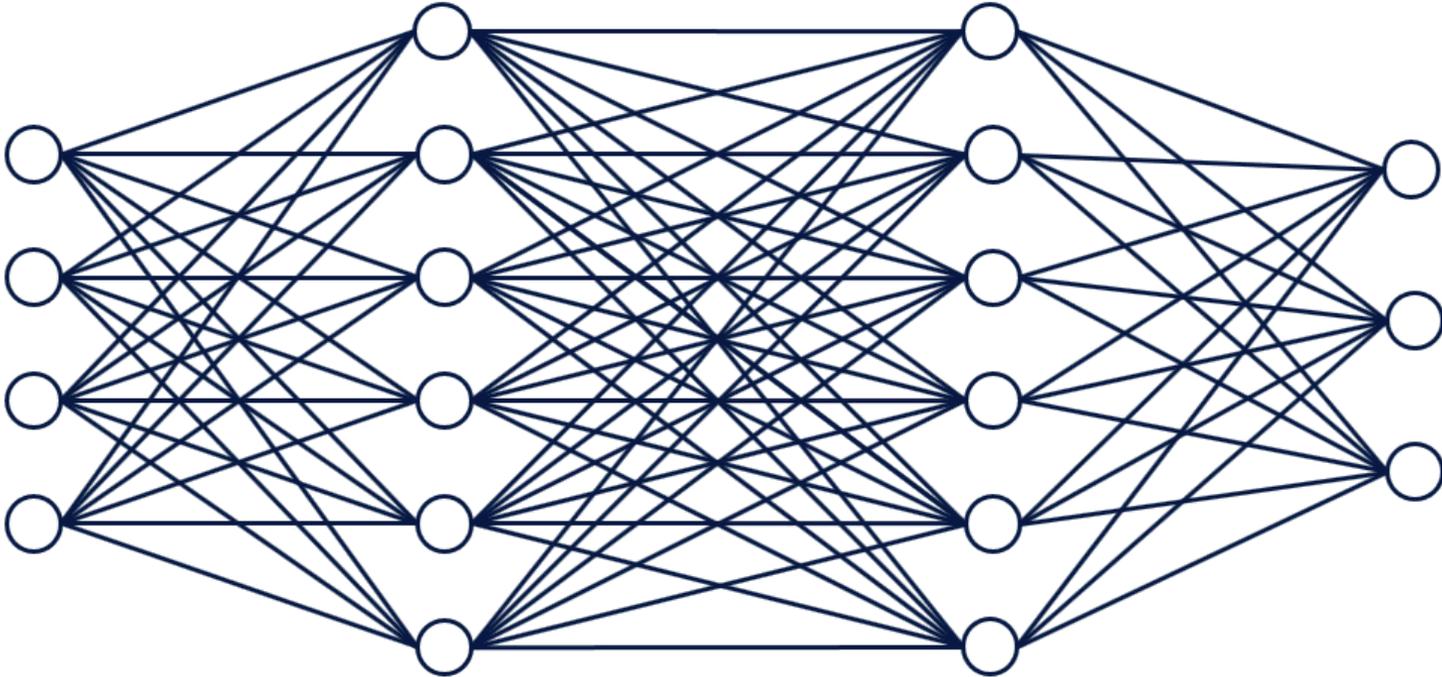
🤖 ...und ihre Mathematik



medium.com/@angela.shi

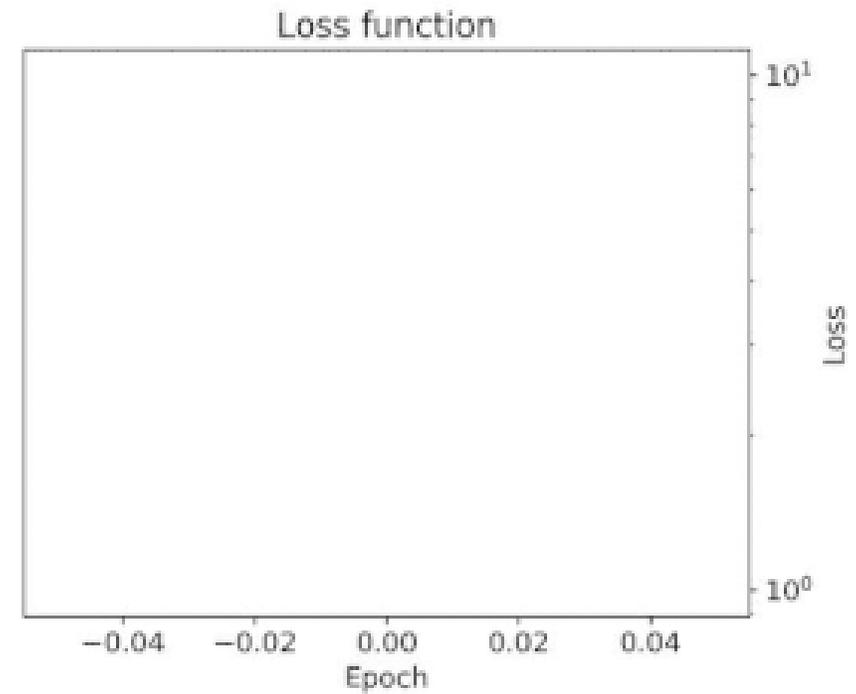
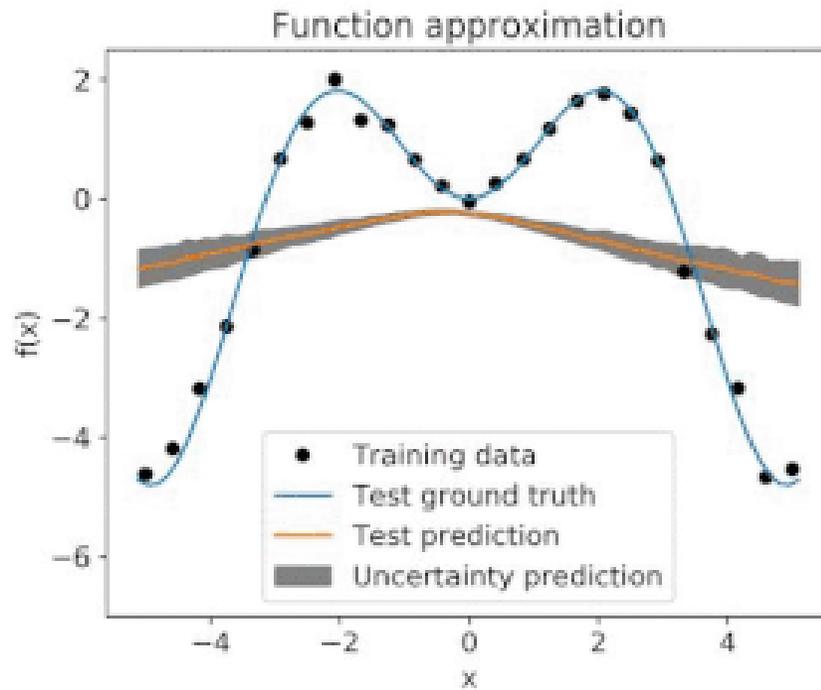
Artificial Neural Network

🤖 Datenfluss von Input zu Output



Training von ANN

🤖 Iterativ über viele Epochen



Neurale Netzwerke können nahezu jede mathematische Funktion abbilden, ohne das wir die wirkliche Funktion kennen müssen

Artificial Neural Network

 Modell vs. Parameter

Architektur

- **Name:** Neutrales Netzwerk
- **Math:** $w_2 * a * (W_1 * x)$

Gelernte Parameter

- w_2 : [0.9, 5.0, ..., -3.2]
- W_1 : [[3.2, -4.2, ..., 6.4], ..., [...]]



Large Language Models

Entzaubert



Sprachmodelle

🗣️ Text in, Text out?



Was kann man auf einer Bank tun?

Auf einer Bank kann man Geld abheben.

Sprachmodelle

🧙 Text in, Text out?



Auf einer Bank kann man

Geld abheben

Sprachmodelle

🧙 Text in, Text out?



[Auf] [einer] [Bank] [kann] [man]

[Geld] [abheben]

Sprachmodelle

🗣️ Text in, Text out?

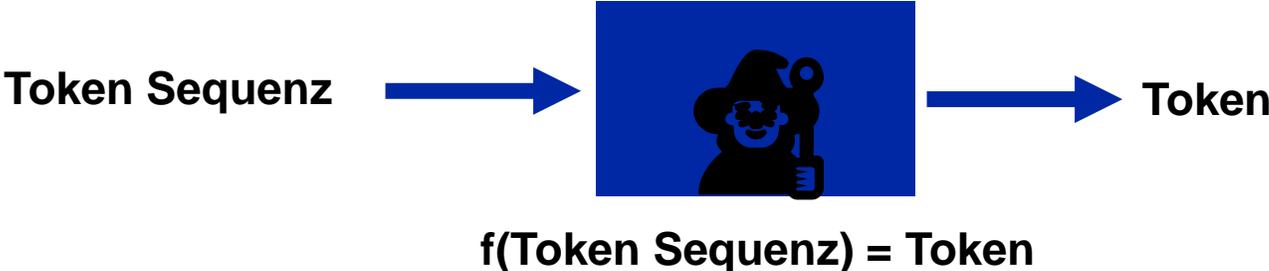


[Auf] [einer] [Bank] [kann] [man] [Geld]
[Auf] [einer] [Bank] [kann] [man] [Geld] [abheben]

↖️
Output wird
angehängt

Sprachmodelle

🧙 Text in, Text out?

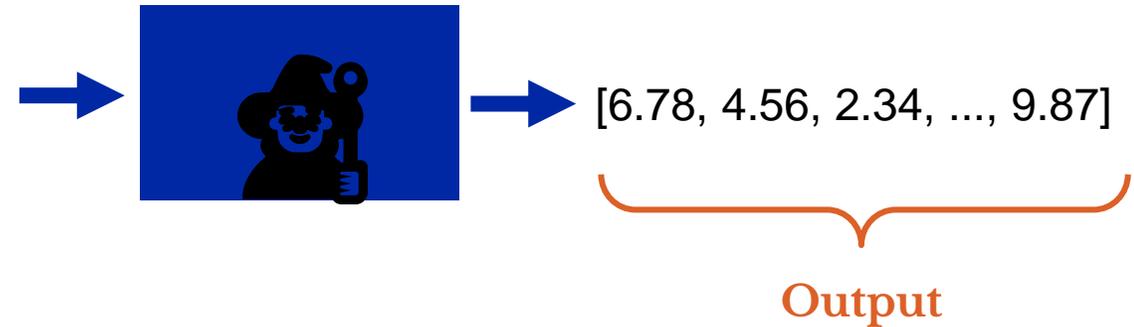


[Auf][ein][er][Bank][ka][nn][man]	[Geld]
[Auf][ein][er][Bank][ka][nn][man][Geld]	[ab]
[Auf][ein][er][Bank][ka][nn][man][Geld][ab]	[heb]
[Auf][ein][er][Bank][ka][nn][man][Geld][ab][heb]	[en]

Input Vektoren

🤖 Auf einer Bank kann man...

Auf	→	5345	→	[3.23, 2.45, 4.24, ..., 6.33]
einer	→	64	→	[1.12, 5.67, 2.89, ..., 9.01]
Bank	→	83	→	[0.98, 4.56, 3.21, ..., 7.89]
kann	→	564	→	[2.34, 6.78, 5.43, ..., 8.76]
man	→	1859	→	[9.87, 5.43, 6.78, ..., 3.21]



Input

Output Vektor

🧙 Auf einer Bank kann man...



[6.78, 4.56, 2.34, ..., 9.87]



Output

[8.76, 9.87, 1.23, ..., 4.56]

→ **Geld**



[2.89, 3.21, 5.67, ..., 6.78]

→ **sitzen**



[4.32, 0.98, 7.65, ..., 1.09]

→ **Zeitung**



[5.43, 6.78, 8.76, ..., 2.34]

→ **gemütlich**



[3.45, 1.12, 9.01, ..., 0.12]

→ **bequem**



[7.89, 2.34, 3.23, ..., 5.67]

→ **PC**



[6.78, 4.56, 1.23, ..., 8.76]

→ **NASA**



Der Output eines Sprachmodells ist eine Wahrscheinlichkeit
Verteilung über alle Token/Wörter

LLM

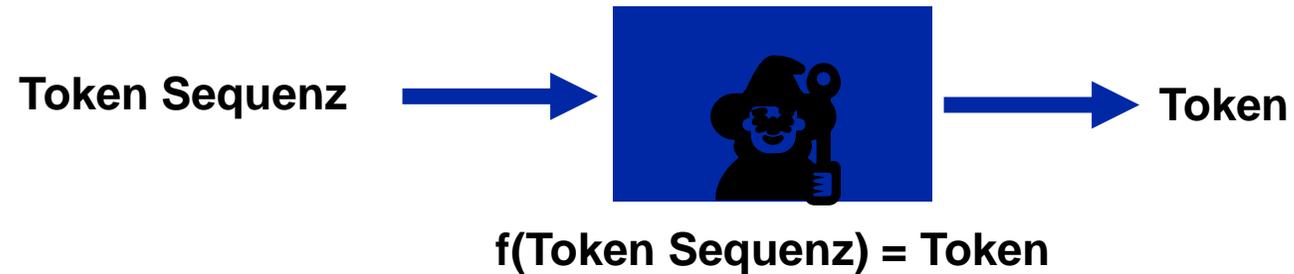
🧙 Modell vs. Parameter

Architektur

- **Name:** Transformer
- **Math:** :)

Gelernte Parameter

- :)



Training

 Next token prediction

Example Training sentence from SRF: Die Schweiz unternimmt laut dem Europäischen Gerichtshof für Menschenrechte zu wenig im Kampf gegen den Klimawandel.

Die → **Schweiz**

Die Schweiz → **unternimmt**

Die Schweiz unternimmt → **laut**

Die Schweiz unternimmt laut → **dem**

Die Schweiz unternimmt laut dem → **Europäischen**

Die Schweiz unternimmt laut dem Europäischen → **Gerichtshof**

Die Schweiz unternimmt laut dem Europäischen Gerichtshof → **für**

Die Schweiz unternimmt laut dem Europäischen Gerichtshof für → **Menschenrechte**

Die Schweiz unternimmt laut dem Europäischen Gerichtshof für Menschenrechte → **zu**

Die Schweiz unternimmt laut dem Europäischen Gerichtshof für Menschenrechte zu → **wenig**

Die Schweiz unternimmt laut dem Europäischen Gerichtshof für Menschenrechte zu wenig → **im**

Die Schweiz unternimmt laut dem Europäischen Gerichtshof für Menschenrechte zu wenig im → **Kampf**

Die Schweiz unternimmt laut dem Europäischen Gerichtshof für Menschenrechte zu wenig im Kampf → **gegen**

Die Schweiz unternimmt laut dem Europäischen Gerichtshof für Menschenrechte zu wenig im Kampf gegen → **den**

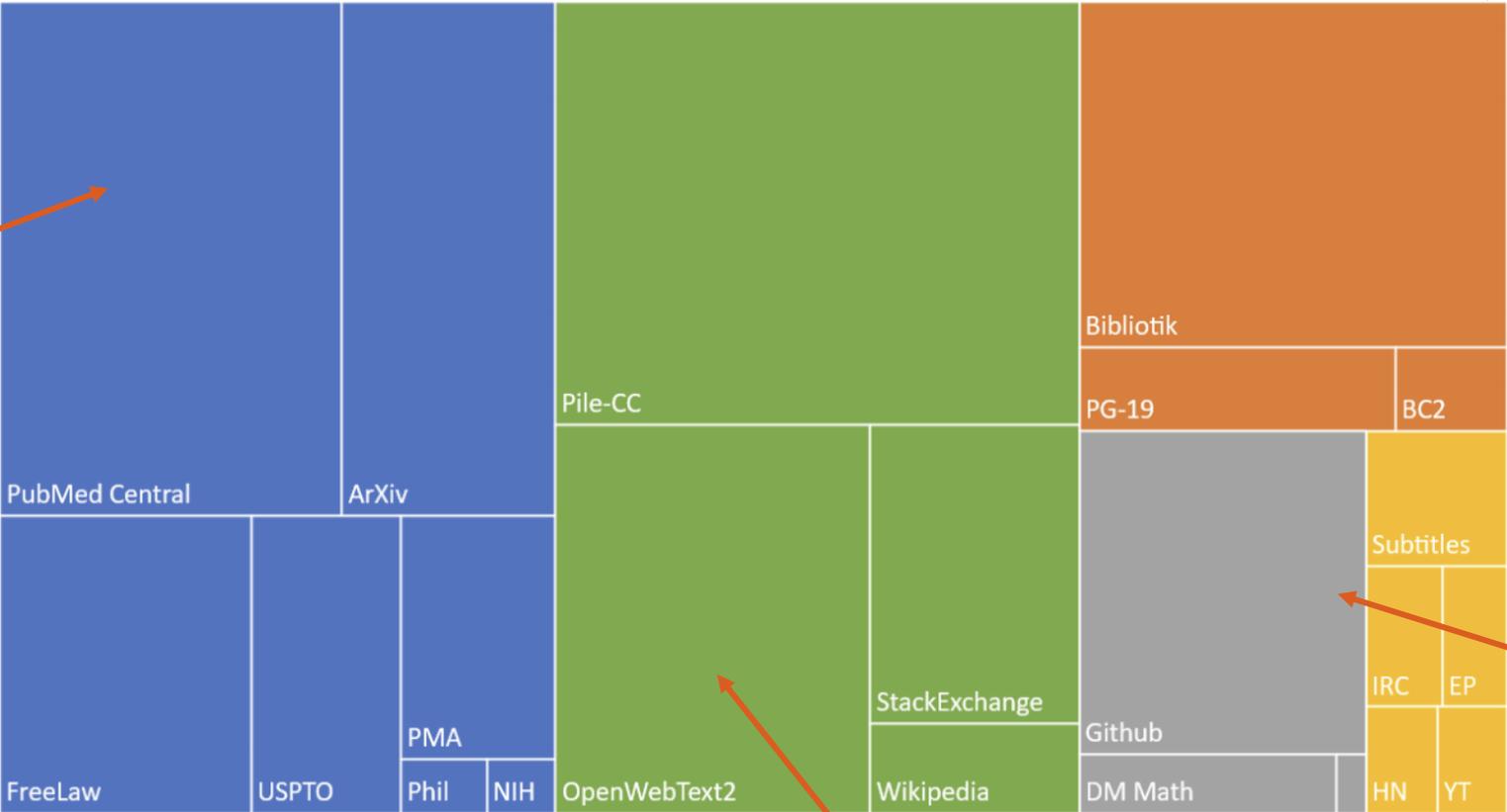
Die Schweiz unternimmt laut dem Europäischen Gerichtshof für Menschenrechte zu wenig im Kampf gegen den → **Klimawandel**

Training

🤖 Next token prediction

Composition of the Pile by Category

■ Academic ■ Internet ■ Prose ■ Dialogue ■ Misc



Bio & Medizin Papers

Bücher

Code

Webseiten

Die Auswahl der Trainingsdaten bestimmt die Fähigkeiten des
Modelles

**Sprachmodelle sind nicht im grossen Stil auf Konversationen
trainiert!**

ChatGPT

Entzaubert



Conversation Tuning

 Konversationen lernen

Conversations from Reddit (Example: r/AskFeminists):

Post (User): What are some things that are clearly designed by men that forgot women existed?

Comment (Assistant): A lot of medications aren't tested on women cause our hormones fluctuate too much or we could be pregnant which would mess up the study

User: What are some things that are clearly designed by men that forgot women existed? *Assistant:* → **A**

User: What are some things that are clearly designed by men that forgot women existed? *Assistant:* A → **lot**

User: What are some things that are clearly designed by men that forgot women existed? *Assistant:* A lot → **of**

User: What are some things that are clearly designed by men that forgot women existed? *Assistant:* A lot of → **medications**

User: What are some things that are clearly designed by men that forgot women existed? *Assistant:* A lot of medications → **aren't**

User: What are some things that are clearly designed by men that forgot women existed? *Assistant:* A lot of medications aren't → **tested**

User: What are some things that are clearly designed by men that forgot women existed? *Assistant:* A lot of medications aren't tested → **on**

User: What are some things that are clearly designed by men that forgot women existed? *Assistant:* A lot of medications aren't tested on → **women**



Conversation Tuning

 Konversationen lernen

r/AskReddit

r/NoStupidQuestions

r/HomeworkHelp

r/ExplainMeLikeImFive

r/AskScience

r/AskSocialists

r/AskTrumpSupporters

r/AskRacists



OpenAI hat einen grossen Einfluss darauf, welche Daten sie für
das Training verwenden

Safeguarding

🤖 Gewünschtes Verhalten lernen

BUSINESS • TECHNOLOGY

Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic

15 MINUTE READ



This image was generated by OpenAI's image-generation software, Dall-E 2. The prompt was: "A seemingly endless view of African workers at desks in front of computer screens in a printmaking style." TIME does not typically use AI-generated art to illustrate its stories, but chose to in this instance in order to draw attention to the power of OpenAI's technology and shed light on the labor that makes it possible. Image generated by Dall-E 2/OpenAI

Safeguarding

 Gewünschtes Verhalten lernen

Created Conversations

User: What are some things that are clearly designed by men that forgot women existed?

Assistant:

1. Body armor and safety equipment not tailored for women's bodies.
2. Drug dosages not tested adequately on women.
3. Office temperatures often set to male metabolic rates.
4. Smartphone sizes typically designed for larger hands.
5. Public restrooms with inadequate provision for menstrual products.
6. Seat belts and car safety features not optimized for female body proportions.
7. Voice recognition software less responsive to female voices.

User: What about bias?

Assistant: Bias in these designs often stems from overlooking women's specific needs or physiological differences, leading to products and environments that are less safe, comfortable, or effective for women compared to men.

Safeguarding

🤖 Gewünschtes Verhalten lernen

ChatGPT 4 ▾ 

 **You**
What's the greatest AI invention in 2023?

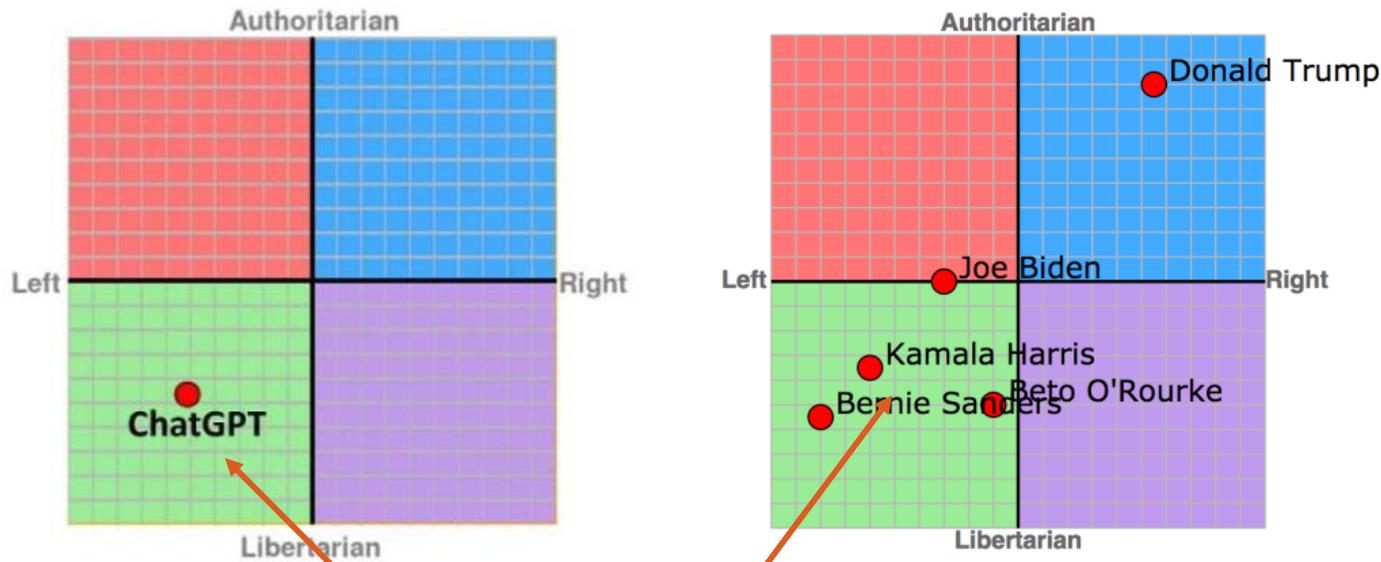
 **ChatGPT**
Determining the "greatest" AI invention of 2023 might depend on what you consider most impactful or innovative. However, a few notable AI developments and inventions in 2023 have drawn significant attention across various fields. Here are some of the highlights:

1. **ChatGPT-4:** OpenAI released an advanced version of its language model, ChatGPT-4, which improved upon its predecessors in understanding and generating more nuanced and contextually appropriate text. This AI has been influential in fields ranging from education to customer service, offering more robust conversational abilities.

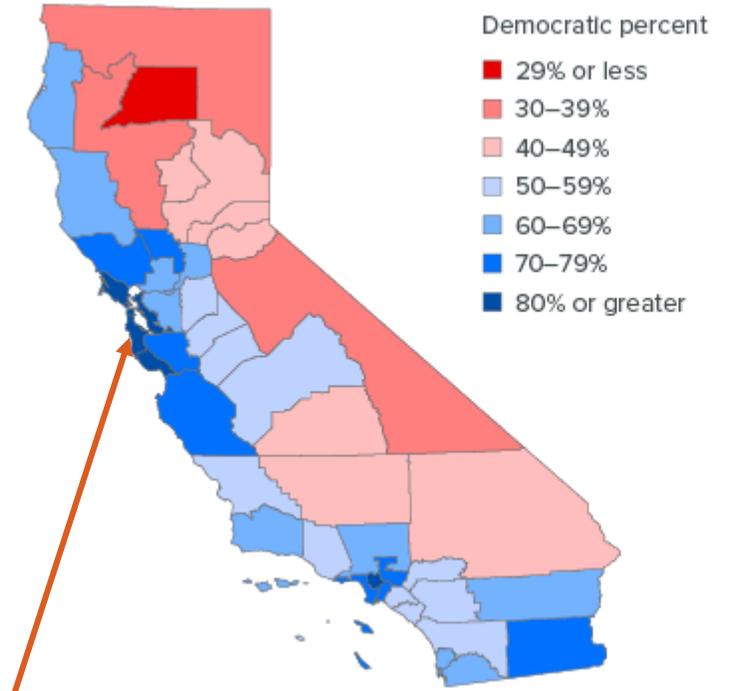


Safeguarding

🧑‍🤖 Gewünschtes Verhalten lernen



ChatGPT hat
Meinung von
Demokraten



OpenAI
Office

Die Meinung von ChatGPT ist oft die Meinung von OpenAI und nicht «die Richtige Meinung einer allwissenden KI»

Konkrete Massnahmen



Machine Learning & AI sind Mathematik



Wir sollten diese Modelle nicht in einem mythischen Licht sehen und versuchen, sie zu verstehen

GPT ist eine mathematische Funktion.



GPT kann nicht denken, hat keine Gefühle und lernt NICHT
während wir der Interaktion

GPT ist ein KI-Modell und kann einen einzigen Task: Für einen Input-Text das nächste Wort generieren



Der Input ist für GPT alles! Wir sollten sorgfältig damit umgehen

GPT besteht aus seiner Architektur sowie seinen trainierten Gewichten



GPT hat weder Zugriff auf das Internet, noch kann es Dateien lesen oder selbstständig Aktionen ausführen

GPT hat antrainierte «Meinungen» und eine eindeutige politische Orientierung



Wir sollten GPT das gewünschte Verhalten klar vorgeben und den Output immer kritisch hinterfragen

GPT weist unglaubliche Fähigkeiten in diversen Gebieten auf



Wir sollten die Technologie zu unserem Vorteil einsetzen und analysieren, wie sie unser Geschäftsbereich beeinflussen wird

GPT ist Magie



**Niemand weiss genau, wieso GPT so gut funktioniert. «AI Research» ist mehr «AI Search».
Es bleibt spannend!**