

# Generative Textmodelle: Herausforderungen und Chancen in der Bildung

23. Mai 2023

Manuela Hürlimann

*manuela.huerlimann@zhaw.ch*

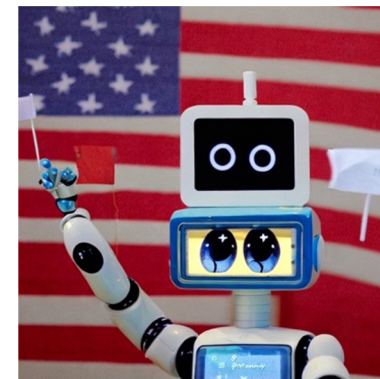
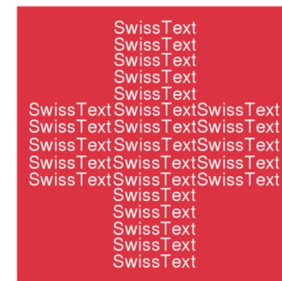


<https://stablediffusionweb.com/#demo>, prompt: "robot typing on a keyboard"

# Über mich



Forschungsgruppe “Natural Language Processing” am  
Centre for Artificial Intelligence (CAI)



# Inhalt des Workshops




1. Was sind generative Textmodelle und wie funktionieren sie?
2. Erfahrungen an der ZHAW und einige Hypothesen, Denkanstösse & Ideen
3. Einordnung der Chancen & Risiken

← → ↻ chat.openai.com/?model=text-davinci-002-... ⏏ ☆

☰ Generative KI-Modelle. +

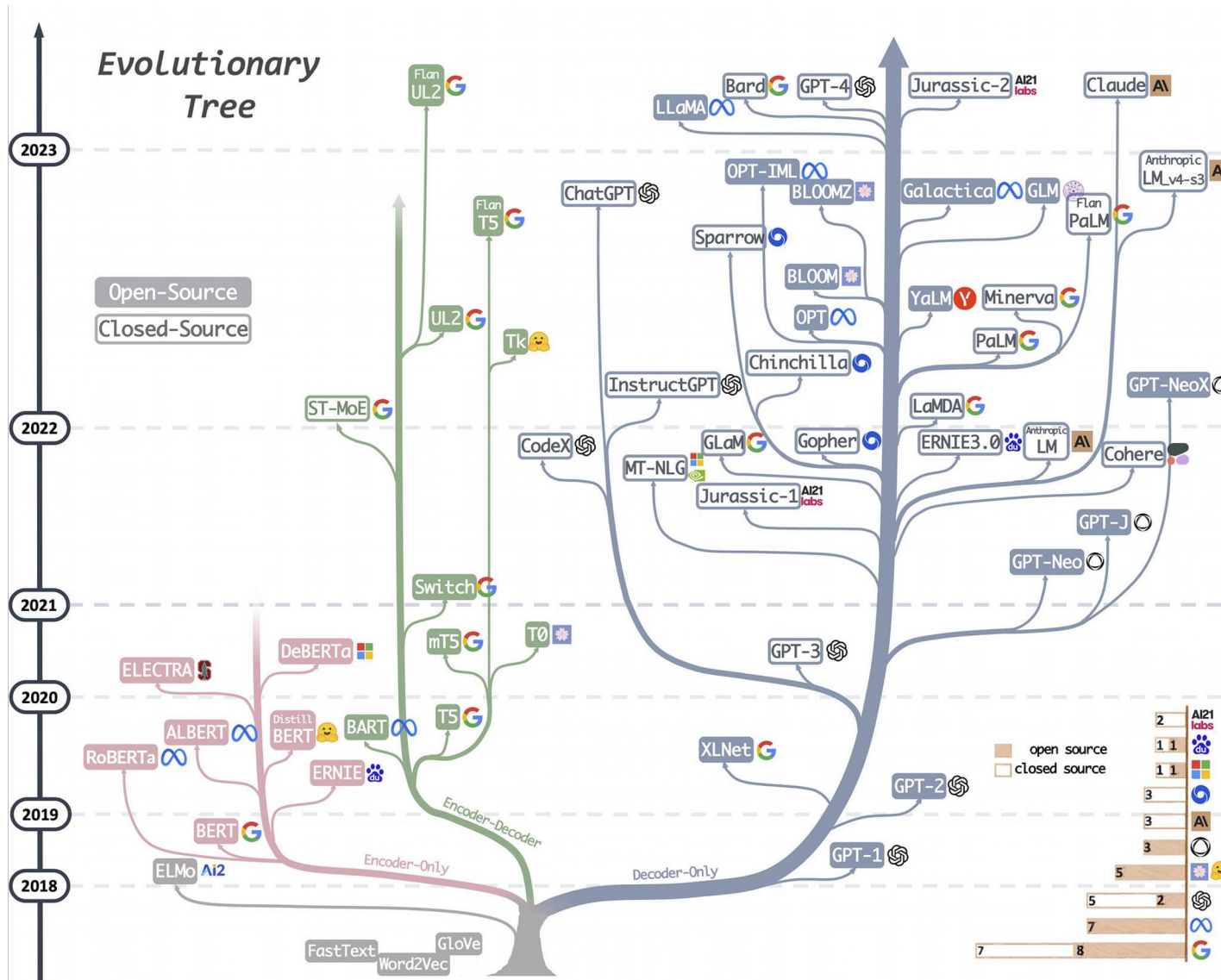
MA Was ist generative Künstliche Intelligenz? ✍

 Generative Künstliche Intelligenz bezieht sich auf eine Art von KI-Modell, das darauf ausgelegt ist, neue Daten oder Inhalte zu erzeugen, die ähnlich sind wie die in der Trainingsdatenbank enthaltenen Daten. Im Gegensatz zu anderen Arten von KI-Modellen, die darauf ausgelegt sind, Muster in Daten zu erkennen und Vorhersagen zu treffen, ist das Hauptziel der generativen KI, **kreativ zu sein und neue Inhalte zu produzieren**.

Ein Beispiel für generative KI ist die **Textgenerierung**. Das Modell wird mit einer großen Menge an Texten trainiert und **kann dann neue Texte erzeugen, die auf den Mustern basieren, die es in den Trainingsdaten gefunden hat**. Ein anderes Beispiel ist die Bildgenerierung, bei der das Modell Bilder erzeugen kann, die auf den Mustern basieren, die es in den Trainingsbildern gefunden hat.

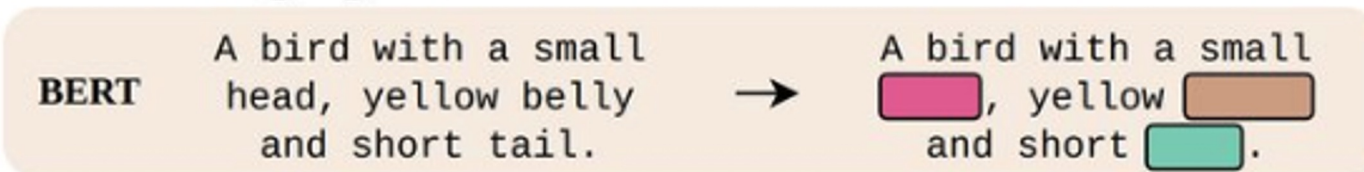
Generative KI-Modelle haben Anwendungen in verschiedenen Bereichen, wie z.B. der Erstellung von Kunstwerken, der Erzeugung von Texten oder sogar der Erstellung von menschenähnlichen Charakteren in Videospiele.

📄 👍 🗨



# Generative Textmodelle (1/2)

## Masked Language Model



## Masked Image Models



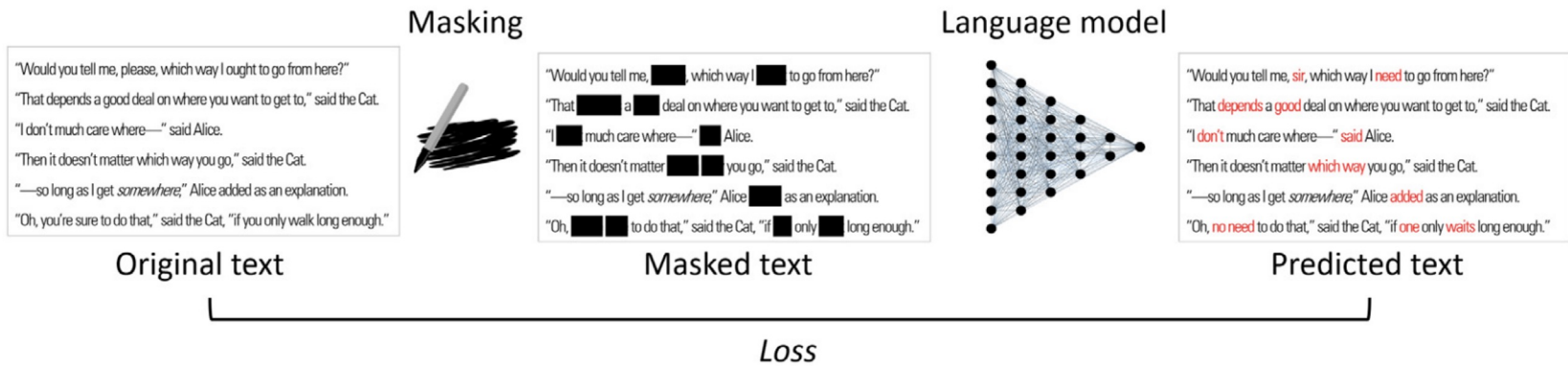
Shi, Y., Siddharth, N., Torr, P., & Kosiorek, A. R. (2022, June). Adversarial masking for self-supervised learning. In *International Conference on Machine Learning* (pp. 20026-20040). PMLR. [CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)

# Generative Textmodelle (2/2)

## A Pretraining



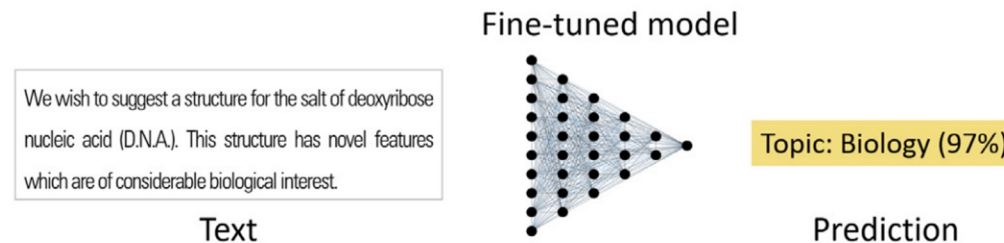
Large corpus  
(unlabeled text)








## B Fine-tuning



Small labeled  
dataset



# Was steckt in solchen Modellen?

	Schritt	Gelernte Fähigkeit	Verwendete Daten
	Selbst-überwachtes Sprachmodell (self-supervised language model)	Vorhersage nächstes Wort/Subword aufgrund von Kontext	Riesige (!) Menge von Textdaten aus dem Web etc
	Instruction Fine-Tuning (IFT)	Befolgen von spezifischen Anweisungen wie "erstelle eine Zusammenfassung", "erzähle einen Witz"	Anweisung - Input - Output (Quelle: Menschen & Modelle)
	Supervised Fine-Tuning (SFT)	Hilfreich & harmlos antworten	Dialoge mit Annotationen/Bewertungen
	Reinforcement Learning mit menschlichem Feedback (RLHF)	Bessere Antworten geben (welche bei Menschen besser ankommen)	Input und 2 oder mehr Outputs (mit Rangfolge)
	Chain of Thought Prompting	Genauere/korrektore Aussagen machen, mehr "gesunder Menschenverstand" (common sense)	Fragen und Antworten mit Denkschritten



# Ein "statistischer Papagei"



<https://stablediffusionweb.com/#demo>, prompt: "parrot doing calculations on a whiteboard"

# Externe Informationen & Schnittstellen

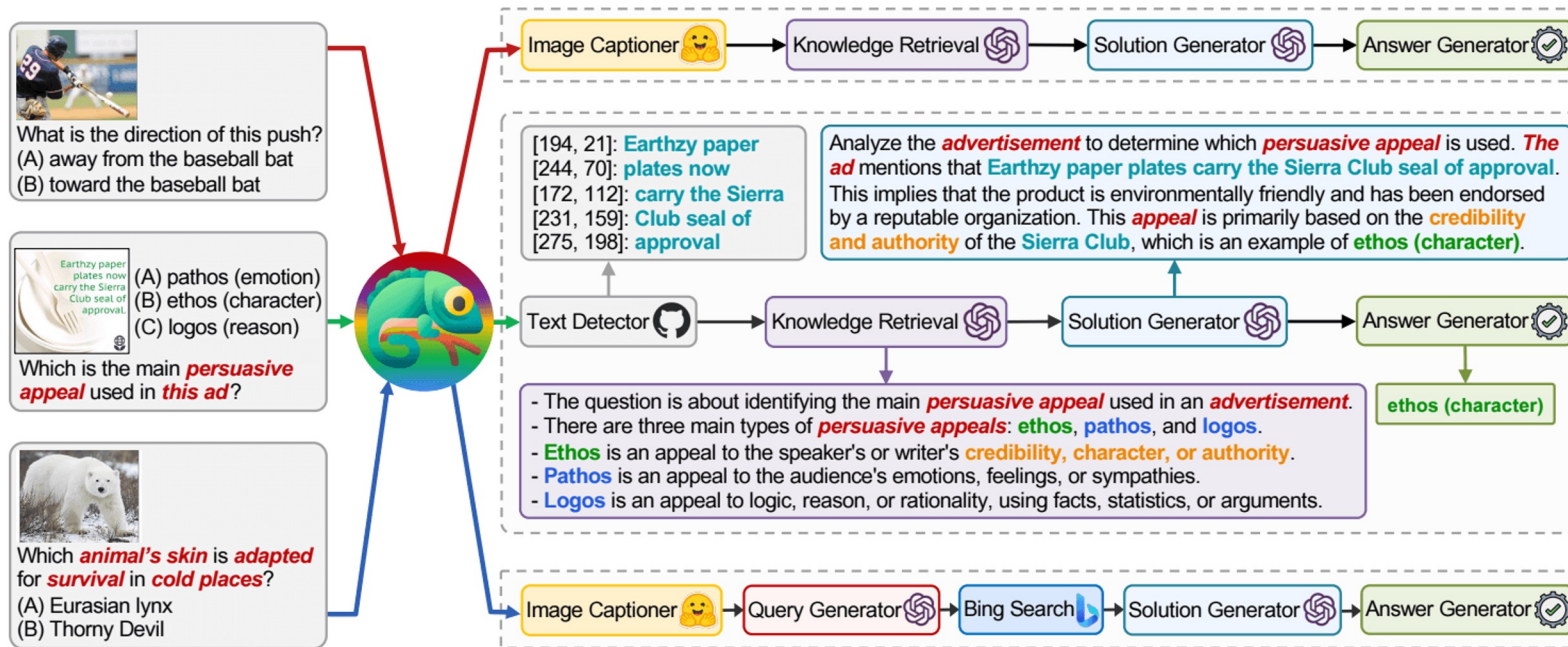


Figure 1: Examples from our Chameleon with GPT-4 on ScienceQA [27], a multi-modal question answering benchmark in scientific domains. Chameleon is adaptive to different queries by synthesizing programs to compose various tools and executing them sequentially to get final answers.

GPT-4



What are a couple of meals I can make with this?

Based on the items visible in the fridge, here are a couple of meal ideas:

- 1) Yogurt parfait: Layer yogurt, strawberries, and blueberries in a bowl or jar. Top with honey or nuts for added texture and flavor.
- 2) Carrot and hummus wrap: Spread hummus on a tortilla or wrap. Top with shredded carrots and any other desired veggies. Roll up and enjoy!

*Note the lack of honey, nuts, or tortillas in the image.* Image: The New York Times

# **Generative Textmodelle im Unterricht**

# In eigener Sache: WBK “GenKI in Lehre und Weiterbildung” an der ZHAW

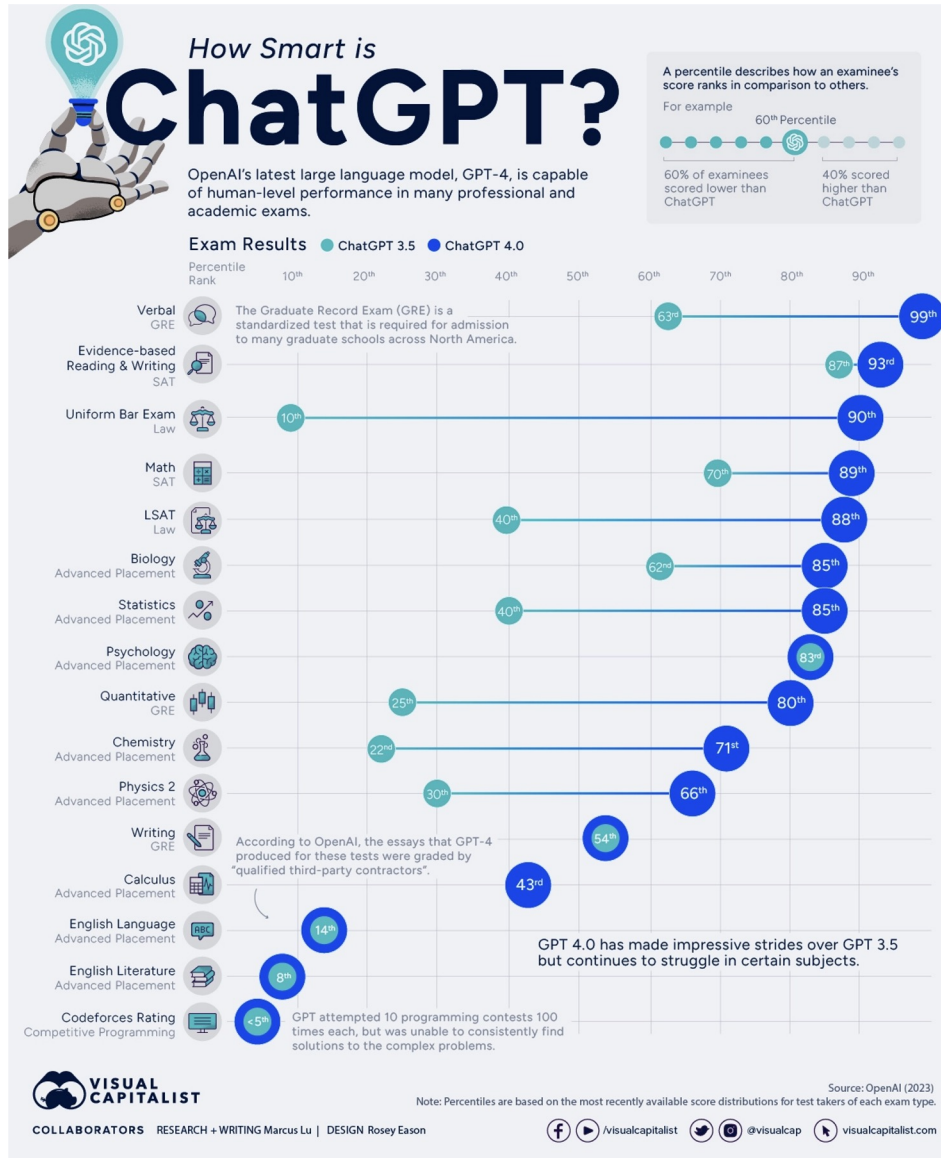


4x Donnerstags:  
8.6., 15.6., 22.6., 29.6., jeweils 16:30-  
20:00 Uhr

**Anmeldeschluss: 25.5.2023!**

<https://shorturl.at/bD028>

<https://www.zhaw.ch/de/engineering/weiterbildung/detail/kurs/wbk-generative-kuenstliche-intelligenz-ki-in-lehre-und-weiterbildung/>



# Ergebnisse an der ZHAW (Stand März 2023)



Vorlesung/Fach	Note
Artificial Intelligence 1	Ca. 3.5
Gesellschaftsrecht	Ca. 4.5
Sensorik für Data Science	85-90% correct answers
Algorithmen und Datenstrukturen	Ca. 4.5
Signals and Systems	Zu viele Bilder

**Hypothese 1: ChatGPT & Co hätten viele unserer bisherigen Prüfungen bestanden**



# Algorithms and Data Structures



Behauptung	wahr	falsch	<u>weiss</u> nicht
1. Ein abstrakter Datentyp (ADT) umfasst die Operationen mit ihren Schnittstellen und deren Implementierung.	xx		
2. Queues funktionieren nach dem LIFO-Prinzip.		xx	
3. Ein rekursiver Algorithmus ohne Abbruchbedingung (“Anker”) führt zu einem <u>StackOverflowError</u> .	xx		
4. Die Türme von Hanoi kann man nur mit einem rekursiven Algorithmus lösen.		xx	
5. Ein rekursives Programm braucht normalerweise weniger Stack-Speicher als ein iteratives Programm für dieselbe Aufgabe.		xx	
6. In einer doppelt verketteten Liste kommt jedes Element genau zweimal vor.		xx	
7. Wenn die Elemente in einer doppelt verketteten Liste aufsteigend sortiert sind, kann man in logarithmischer Zeit auf das mittlere Element zugreifen.	xx		
8. Ein Baum kann keine Zyklen enthalten.	xx		

# Algorithms and Data Structures



## Aufgabe 2: Hashing [7 Punkte]

Gegeben seien die Werte: 2341, 6789, 2356, 1231, 1238, 1236, 1200, 2318.

Geben Sie die Belegung der Hashtabelle mit 10 Plätzen mit *quadratischem* Probing und der Hashfunktion  $h(x) = x \% 10$  an:

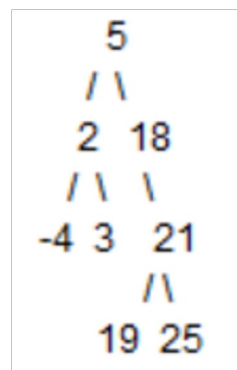
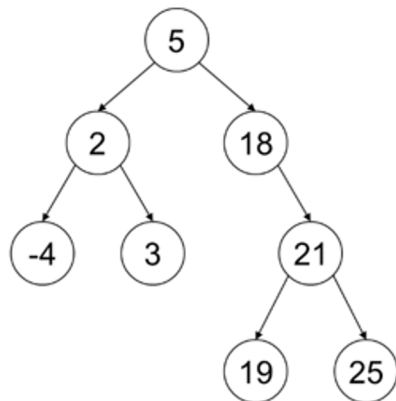
0	2318
1	1200
2	2356
3	1231
4	1238
5	2341
6	6789
7	
8	1236
9	

Schrittweise Belegung der Hashtabelle:

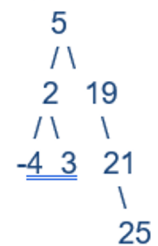
1.  $2341 \% 10 = 1 \rightarrow \text{Tabelle}[1] = 2341$
2.  $6789 \% 10 = 9 \rightarrow \text{Tabelle}[9] = 6789$
3.  $2356 \% 10 = 6 \rightarrow \text{Tabelle}[6] = 2356$
4.  $1231 \% 10 = 1 \rightarrow$  quadratisches Probing:  $(1 + 1^2) \% 10 = 2 \rightarrow \text{Tabelle}[2] = 1231$
5.  $1238 \% 10 = 8 \rightarrow \text{Tabelle}[8] = 1238$
6.  $1236 \% 10 = 6 \rightarrow$  quadratisches Probing:  $(6 + 1^2) \% 10 = 7 \rightarrow \text{Tabelle}[7] = 1236$
7.  $1200 \% 10 = 0 \rightarrow \text{Tabelle}[0] = 1200$
8.  $2318 \% 10 = 8 \rightarrow$  quadratisches Probing:  $(8 + 1^2) \% 10 = 9 \rightarrow \text{Tabelle}[9] = 2318$

# Algorithms and Data Structures

Löschen Sie aus untenstehendem binären Suchbaum den Knoten 18 und geben Sie den gesamten daraus resultierenden neuen binären Suchbaum an.



Der resultierende neue binäre Suchbaum nach dem Löschen des Knotens 18 sieht wie folgt aus:



# Sensorik for Data Scientists



a.) (1 P) Was ist eine Schallwelle? Handelt es sich um Transversalwellen?

a.) Eine Schallwelle ist eine Welle, die sich durch die Schwingungen von Druckunterschieden in einem Medium ausbreitet. Es handelt sich um Longitudinalwellen, da die Schwingungen parallel zur Ausbreitungsrichtung der Welle erfolgen.

b.) (1 P) Nennen Sie drei in der Vorlesung besprochene Kurvenformen von Signalen

b.) In der Vorlesung können besprochene Kurvenformen von Signalen:

1. Sinusform
2. Rechtecksignal
3. Dreiecksignal

Es gibt viele verschiedene Signalformen, diese drei sind nur einige Beispiele.

---

Woher weisst du was in der Vorlesung besprochen wurde?

Als AI-Modell habe ich keine Kenntnis darüber was in einer bestimmten Vorlesung besprochen wurde. Ich habe auf die Frage geantwortet basierend auf meinem allgemeinen Wissen und meiner Schulung. Es ist möglich, dass die in der Frage genannten Kurvenformen Sinusform, Rechtecksignal und Dreiecksignal tatsächlich in einer Vorlesung besprochen wurden, und ich habe das aufgrund meiner Schulung und Kenntnisse in Signalverarbeitung und Elektrotechnik vermutet. Bitte entschuldigen Sie, falls dies nicht der Fall ist.

---

Doch, Du hattest recht.

Ich freue mich, dass meine Antwort Ihrer Erwartung entsprochen hat. Als AI-Modell versuche ich immer, mein allgemeines Wissen und meine Schulung zu verwenden, um auf Fragen so präzise und hilfreich wie möglich zu antworten. Lassen Sie mich wissen, wenn ich Ihnen in Zukunft weiterhelfen kann.

# Wie kann man mit GenKI in Prüfungen umgehen?

## Generative KI vermeiden

- Mündliche Prüfungen
- Prüfungen auf Papier
- Elektronische Prüfungen ohne Internetzugang
- Elektronische Prüfungen, wo GenKI verboten ist

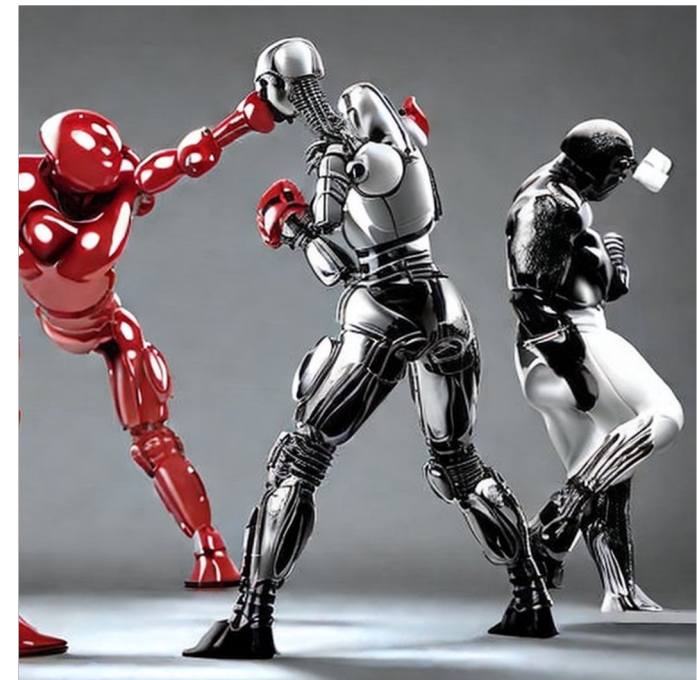


<https://stablediffusionweb.com/#demo>,  
prompt: "forbidden sign"

# Wie kann man mit GenKI in Prüfungen umgehen?

## GenKI bekämpfen

1. Prüfung ausdrucken
2. Aufgaben als Bilder beschreiben (statt Text)
3. Aufgaben, die sich auf Bilder beziehen
4. Lokales Wissen einbauen
5. Aktuelle Ereignisse einbeziehen
6. Aufgaben schwieriger machen (höheres Level in Bloom'scher Taxonomy)
7. Erkennen ob GenKI verwendet wurde (e.g. GPT-Zero)



<https://stablediffusionweb.com/#demo>,  
prompt: "human fighting against artificial intelligence"



**Hypothese 2: Wir werden GenKI nicht  
bekämpfen können!**

# Wie kann man mit GenKI in Prüfungen umgehen?

## Sich GenKI zu eigen machen

- Benutzung von GenKI in Prüfungen integrieren
- Prüfungsfragen mit GenKI generieren
- Eigene Prüfungen von GenKI lösen lassen
- Lernende deklarieren lassen, was mit GenKI gemacht wurde und was eigene Inhalte sind



<https://stablediffusionweb.com/#demo>,  
prompt: "young person jumping"

# Wie kann man mit GenKI in Lernkontrollen umgehen?

1. LK und Übungen vor Ort machen statt zu Hause
2. Nicht benoten (Selbstkontrolle)
3. Jeweils definieren, wie GenKI verwendet werden darf (und warum)
4. GenKI in den Lösungsprozess integrieren
5. Die “echte Welt” im Auge behalten: werden GenKI-Tools in der Praxis genutzt?

# Wie kann man „trotzdem“ Denk- und Schreibprozesse fördern?

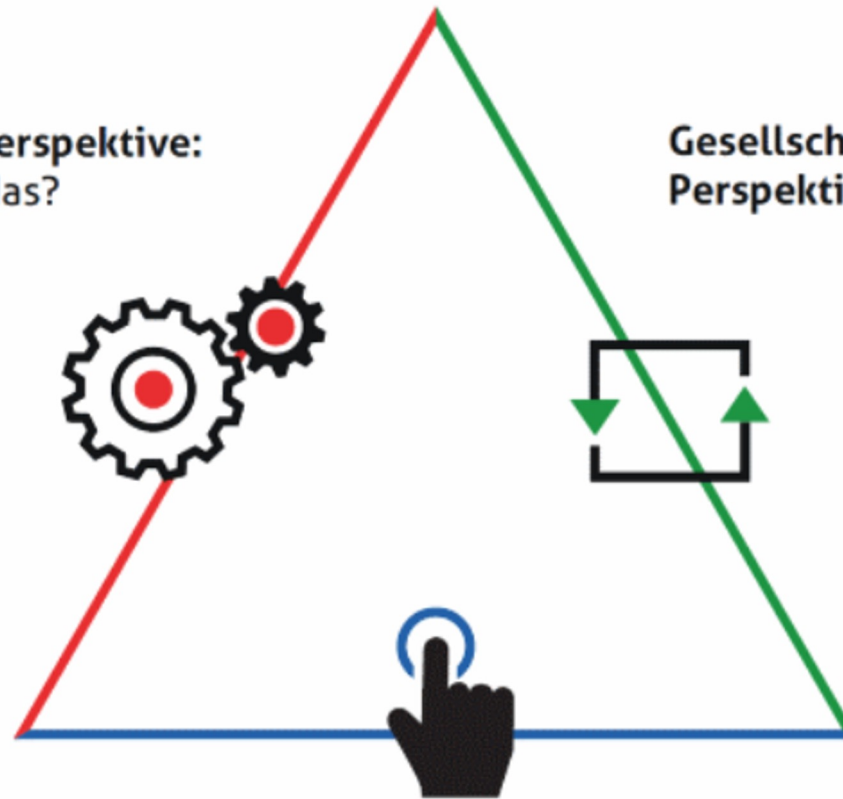
1. Debatten, Gruppendiskussionen, Präsentationen
2. Schreibübungen auf Papier (ohne maschinelle Hilfsmittel);  
Reflexion des Schreibprozesses
3. Kritische Analyse und Einordnung von KI-Output: was sind Stärken, Schwächen, Besonderheiten?

**Hypothese 3: durch GenKI müssen wir überdenken *was* unterrichtet wird und *wie***

# KI-Kompetenzen (1/2)

**Technologische Perspektive:**  
Wie funktioniert das?

**Gesellschaftlich-kulturelle Perspektive:** Wie wirkt das?



**Anwendungsorientierte Perspektive:**  
Wie nutze ich das?

Abb. 1: Dagstuhl-Dreieck, 2018

# KI-Kompetenzen (2/2)

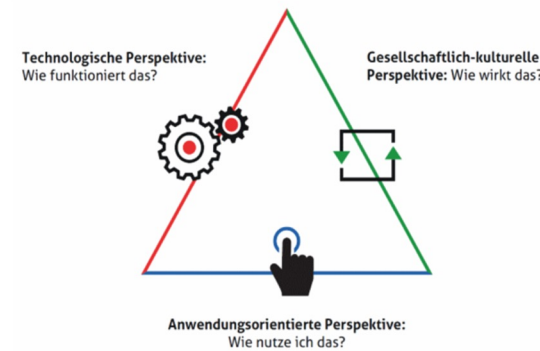


Abb. 1: Dagsuhl-Dreieck, 2018

## Technologische Perspektive

- ChatGPT wurde mit einer grossen Menge an Texten trainiert und generiert auf der Basis dieser Trainingsdaten und der gelernten statistischen Muster eine menschenähnliche Antwort auf unterschiedliche Eingaben (auch Prompt genannt).
- Die KI hat kein wirkliches Verständnis von Sprache und für Inhalte, sondern erkennt und reproduziert Texte lediglich aufgrund von statistischen Mustern.

## Gesellschaftlich-kulturelle Perspektive

- Jobs werden sich verändern, Anforderungen in Bildung und Beruf steigen tendenziell
- (Gen)KI ist ein neues Phänomen für die meisten Bildungskontexte, die Integration braucht noch Zeit
- Die Texte der GenKIs sind sprachlich gut, man kann aber nicht von inhaltlicher Korrektheit ausgehen!
- “Schreiben” bedeutet nicht nur Output, sondern ist ein kognitiver Prozess

## KI-Kompetenzen in der Anwendung

- Z.B. Unterstützung im Schreibprozess, Ideenfindung
- Fachspezifische Einbettung
- Kontinuierliche Veränderung ⇒ am Ball bleiben ist wichtig!

# “AI Literacy”



CHI 2020 Paper

CHI 2020, April 25–30, 2020, Honolulu, HI, USA

## What is AI Literacy? Competencies and Design Considerations

**Duri Long**

Georgia Institute of Technology  
Atlanta, GA, USA  
duri@gatech.edu

**Brian Magerko**

Georgia Institute of Technology  
Atlanta, GA, USA  
magerko@gatech.edu

We define *AI literacy* as a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and **use AI as a tool online, at home, and in the workplace.**



# **Einordnung: Chancen & Risiken**

# Chancen

- Unterstützung im Schreibprozess
- Einfacherer Zugang zu „Wissen“, Lernmöglichkeiten, Überblick über ein Thema gewinnen
- Personalisierung von Informationen, Relevanzfilterung
- Zeitersparnis & Fokus auf wertschöpfende Tätigkeiten
- Neue Lerninputs, z.B.
  - kritische Analyse von KI-Output
  - neue Lösungswege/Prozesse
  - ...



# Herausforderungen & Risiken

- Reproduktion von sozialen Schieflagen (“biases”)
- Copyright & rechtliche Situation
- Falschinformationen
- Transparenz, Zugang & Reproduzierbarkeit
- Ökologischer Fussabdruck



# ZHAW Kompetenzzentrum für Generative KI



**Kontaktstelle für Industrie und Hochschulen**

## **Aktivitäten**

- Seminare und Schulungen
- Wissenschaftliche Experimente
- Präsentationen und Projekte
- Leitlinien für die Nutzung und Umsetzung

[generativeAI@zhaw.ch](mailto:generativeAI@zhaw.ch)



# Richtlinie „Verwendung generativer KI-Systeme bei Leistungsnachweisen“ der ZHAW

Die Richtlinie ist öffentlich verfügbar unter dem folgenden Link:

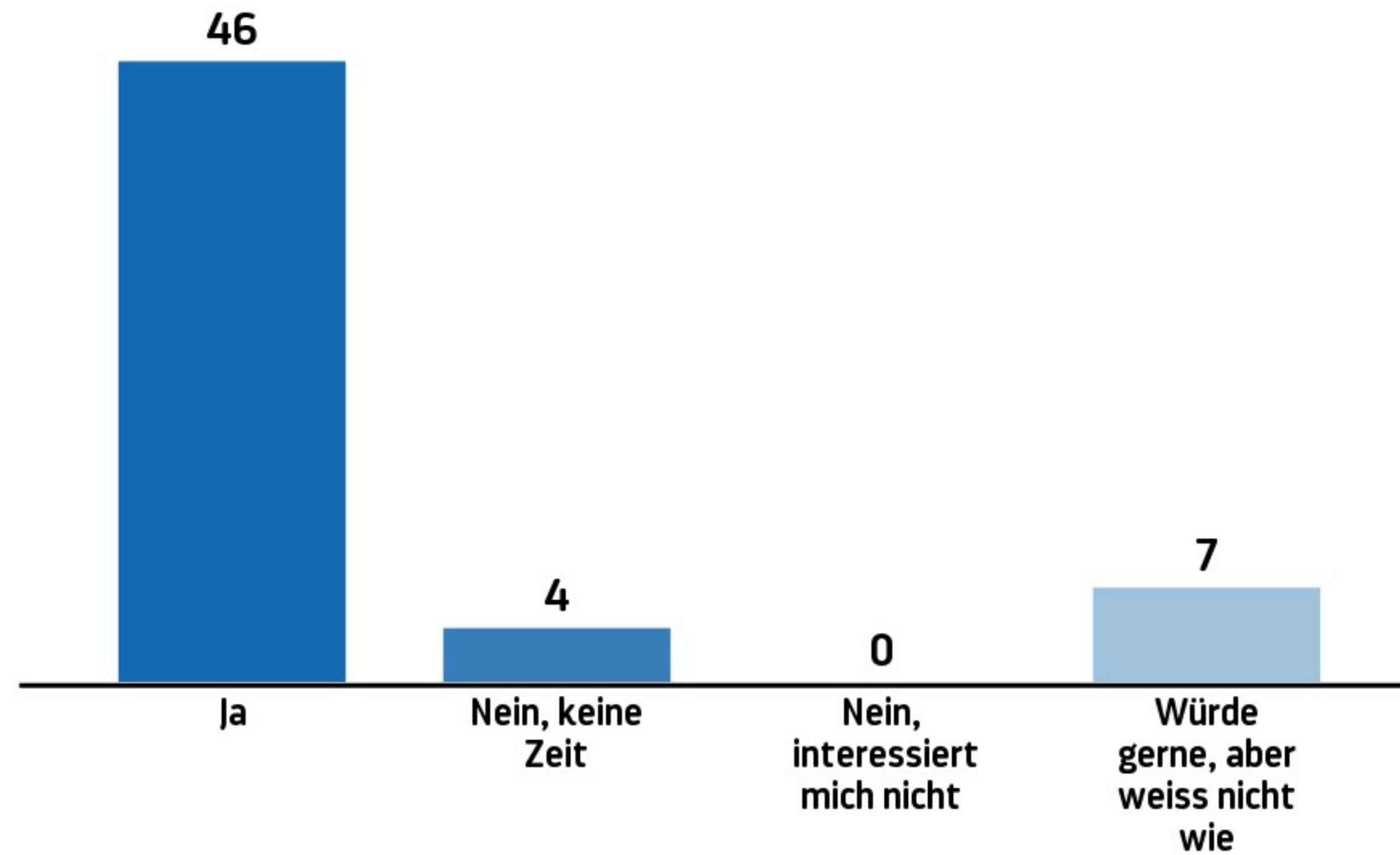
[https://gpmpublic.zhaw.ch/GPMDocProdDPublic/Vorgabedokumente\\_ZHAW/Z\\_RL\\_Richtlinie\\_KI\\_bei\\_Leistungsnachweisen.pdf](https://gpmpublic.zhaw.ch/GPMDocProdDPublic/Vorgabedokumente_ZHAW/Z_RL_Richtlinie_KI_bei_Leistungsnachweisen.pdf)

Wichtigste Punkte:

- Verantwortung für Inhalt, Urheberrecht und wissenschaftliche Integrität liegt bei den Studierenden
- Erlaubte Hilfsmittel werden frühzeitig kommuniziert (unerlaubte Hilfsmittel können Disziplinarfolgen haben)
- Prüfungen: GenKI ist nicht erlaubt, ausser sie wird explizit zugelassen
- Schriftl. Arbeiten: es besteht eine Deklarationspflicht, wenn GenKI verwendet wurde
- Konstanter Reflexionsprozess sowie Anpassungen sind nötig

# **Anhang: Mentimeter Fragen & Ergebnisse**

# Haben Sie schon generative Textmodelle ausprobiert?



## Wofür haben Sie Texte generieren lassen? 69 Answers

Austesten

Try and Error

Unterrichtsvorbereitung

Prüfungsaufgaben

Einladung

Angebotsanalyse,

Informationen gesammelt

Textteile für eine Seminararbeit

Texte Übersetzt



## Wofür haben Sie Texte generieren lassen?

69

Answers

Einfache Texte

Kursprogramm

RedeGlückwunschkarte

Rezept

Leistungsnachweis Hochschule

Briefe/ mailvorlagen/ aufgaben

59791076

Austesten, somit Spielerei

Brainstorming

## Wofür haben Sie Texte generieren lassen? 69 Answers

Austesten

Eigene Texte verbessern lassen

Trauerkarte

Ausprobiert

Erkennung von ChatGPT-Texten

Testzwecke mit den Lernenden

Rede vorbereiten

Gedichte - kann kein Haiku

Stellenbewerbung

## Wofür haben Sie Texte generieren lassen?

69 Answers

Newsletter

Testen

Geschäftliche Mail und Geburtstagskarten

Tipps für mein Hochbeet geholt 😊

Bewerbung schreiben

Mails beantwortet

Übersetzung nail

Eigene Vorbereitung für mündliche  
Prüfungsabnahme

nach themen gesucht

# Wofür haben Sie Texte generieren lassen? 69 Answers

Eine kleine Werbekampagne erstellt und einen LinkedIn-Text für einen Post

Brainstorming

Strukturierung; Unformulierung; Coding  
Problemlösung

Protokolle, Abstracts,  
Zusammenfassungen,Ideen

Ausprobiert

Unterricht vorbereitet

Mailtext  
ÜbersetzungenReiseplanungGutenachtgesc  
hichte für Kinder

Arbeitszeugnis / Poesie / Essay /

Ausprobiert

## Wofür haben Sie Texte generieren lassen? 69 Answers

Lektionenpläne

Anschlussarbeit

Nur kurze Texte zum Kennenlernen.

Test

ideen für forschungsprojekte gesammelt

Tests, kahoots erstellt

Mich auf eine Job-Bewerbung vorbereitet, E-Mails verfasst, wissenschaftliche Texte zusammengefasst und übersetzt.

Schreibe ein Liebes-Sonett, in dem die Worte Staubsauger und Lawine vorkommen.

Mailtext, Texte kürzen, Einladungen, Prüfungsfragen,

# Wofür haben Sie Texte generieren lassen?

69

Answers

Ausprobiert

Austesten

Ausschreibungen überarbeitet

Test

Zusammenfassung

Übersetzen

Abstract

59791076

Interessant

## Wofür haben Sie Texte generieren lassen? 69 Answers

Hilfreich

Verblüfft

Gut, werde es wieder verwenden

Erstaunliche Breite

Hat mich glatt belogen

Abschnitte wiederholen sich

## Wie waren Ihre Erfahrungen?

54 Answers

Spannend, aber auch fehlerhaft

Sehr positiv!

Positiv

Positiv

positiv und hilfreich

Sehr positiv

Hilfreich

Ok. Keine neuen Erkenntnisse

Positiv überrascht.



## Wie waren Ihre Erfahrungen?

54 Answers

Überraschend gut

Bin beeindruckt

Erstaunlich

Höflicher umgang

Je länger damit gearbeitet desto besser

zu wenig fundierte Antworten

Zuerst begeistert

WOOOW

Gut und Interessant

## Wie waren Ihre Erfahrungen?

54 Answers

Spannend

Positiv, grosses Potential, entlastet mich...

Erstaunlich / toll / hilfreich

Funktionalität sehr gut, Antworten nicht immer sinnvoll

Bin beeindruckt. Sehr gute Qualität des Outputs.

Ok

Sehr gut! Hilfreiche Assistenz

ganz gut.

Ganz ok, aber fehlerhaft.

## Wie waren Ihre Erfahrungen?

54 Answers

Positiv

Jeweils so gut wie die Eingabe gut ist

Die Resultate haben nicht überzeugt.  
Noch nicht.

Positiv, Inhalte waren gut

Sehr gut, mit Vorsicht auf die  
Korrektheit

Sehr gut, gute Grundlage zur  
Weiterbearbeitung

Falsche Quellenangabe

Sehr hilfreich, sehr positiv

Gut. Nur kleine Verifizierung und  
Korrekturen nötig

## Wie waren Ihre Erfahrungen?

54 Answers

Überraschend gut

Spannend

Auch fehlerhaft

Verblüfft über die sprachliche  
Textqualität

Gute, brauchbare Antworten erhalten

war erst erstaunt über die Qualität,  
dann etwas ernüchtert

Verblüffend

Etwas seicht

Vielfältig, interessant, nicht vertieft

## Wie waren Ihre Erfahrungen?

54 Answers

Positiv, wenn die Fragen gezielt formuliert

Interessant mit Prompts rumzuspielen

Von a bis z in eins durch dchriben

Interessant

Halluzinationen

Quellenangaben gefehlt

Nicht real

wenn man spezifische Frage hat sehr gut muss nach haken

Ok

# Haben Sie GenKI auf Ihren Prüfungen ausprobiert?

