Hasso Plattner Institut
Digital Engineering · Universität Potsdam

KI Service Zentrum by Hasso-Plattner-Institut

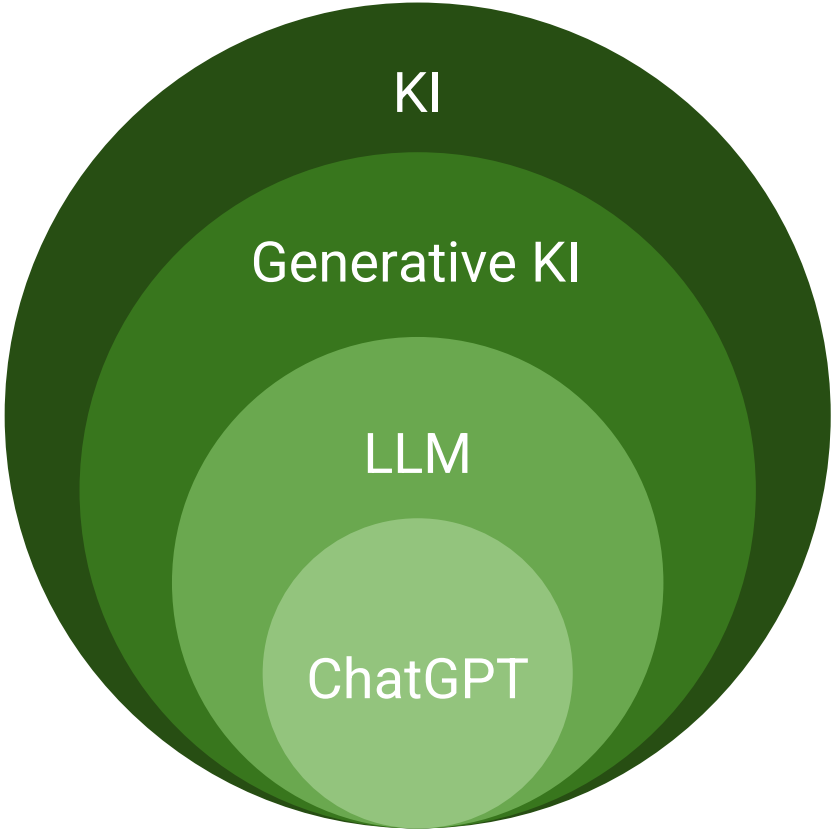
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

- openHPI: ChatGPT: Was bedeutet generative KI für unsere Gesellschaft? -

Funktionsweise von generativer KI

Johannes Hötter
Christian Warmuth

Wichtige Abgrenzung

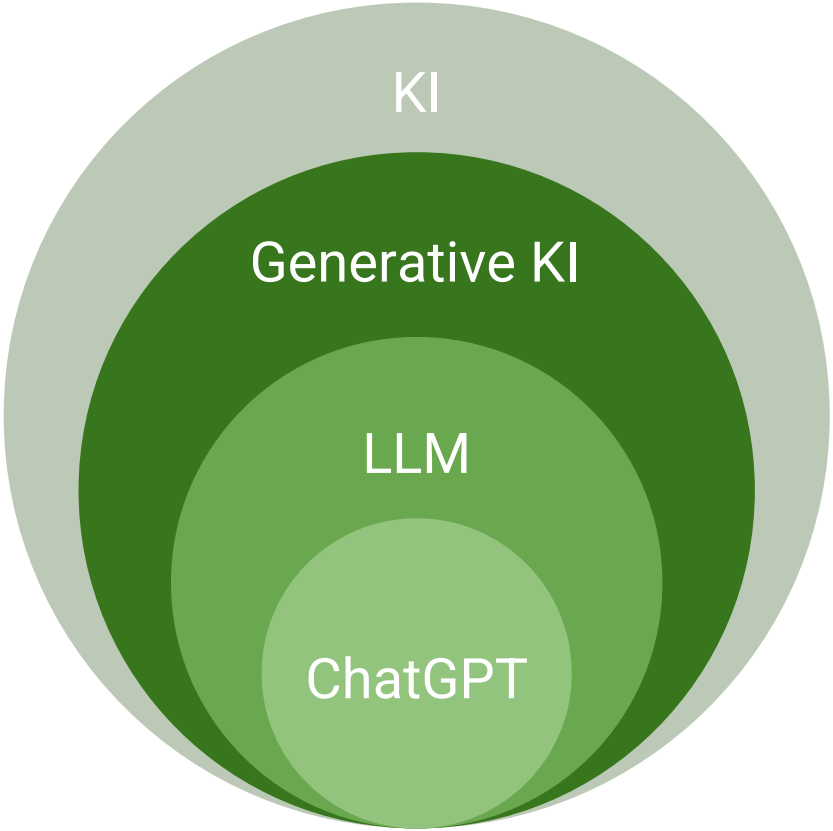


Funktionsweise von generativer KI

Johannes Hötter,
Christian Warmuth

openHPI

Wie funktioniert Generative KI?



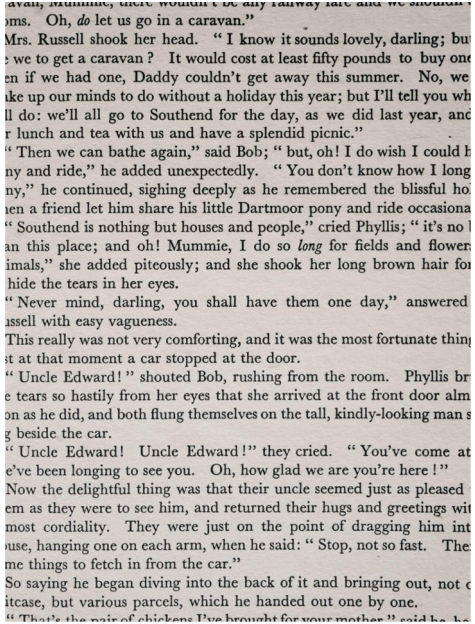
Funktionsweise von generativer KI

Johannes Hötter,
Christian Warmuth

openHPI

Generative KI

Bild-Modelle
(Diffusion-Modelle)



ms. Oh, *do* let us go in a caravan."
Mrs. Russell shook her head. "I know it sounds lovely, darling; bu
we to get a caravan? It would cost at least fifty pounds to buy one
en if we had one, Daddy couldn't get away this summer. No, we
ke up our minds to do without a holiday this year; but I'll tell you wh
ll do: we'll all go to Southend for the day, as we did last year, and
r lunch and tea with us and have a splendid picnic."
"Then we can bathe again," said Bob; "but, oh! I do wish I could h
ny and ride," he added unexpectedly. "You don't know how I long
ny," he continued, sighing deeply as he remembered the blissful ho
en a friend let him share his little Dartmoor pony and ride occasiona
"Southend is nothing but houses and people," cried Phyllis; "it's no
an this place; and oh! Mummie, I do so *long* for fields and flower
imals," she added piteously; and she shook her long brown hair fo
hide the tears in her eyes.
"Never mind, darling, you shall have them one day," answered
ssell with easy vagueness.
This really was not very comforting, and it was the most fortunate thin
t at that moment a car stopped at the door.
"Uncle Edward!" shouted Bob, rushing from the room. Phyllis br
e tears so hastily from her eyes that she arrived at the front door alm
on as he did, and both flung themselves on the tall, kindly-looking man s
g beside the car.
"Uncle Edward! Uncle Edward!" they cried. "You've come at
e've been longing to see you. Oh, how glad we are you're here!"
Now the delightful thing was that their uncle seemed just as pleased
em as they were to see him, and returned their hugs and greetings wi
most cordiality. They were just on the point of dragging him int
use, hanging one on each arm, when he said: "Stop, not so fast. The
me things to fetch in from the car."
So saying he began diving into the back of it and bringing out, not
itcase, but various parcels, which he handed out one by one.

LLMs

…

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam

KI Service Zentrum by Hasso-Plattner-Institut HPI

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Funktionsweise von generativer KI

Johannes Hötter,
Christian Warmuth

openHPI

Generative KI

Bild-Modelle
(Diffusion-Modelle)

LLMs

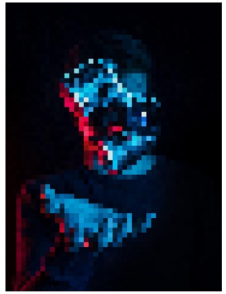
Funktionsweise von generativer KI

Johannes Hötter,
Christian Warmuth

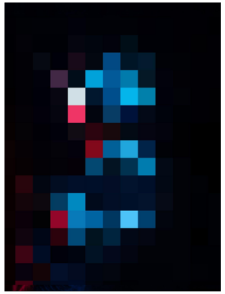
openHPI

Wie entstehen Bilder mit generativer KI?

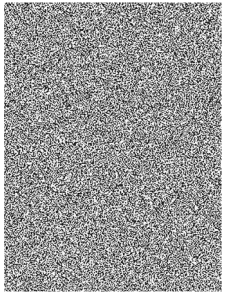
Person wirft eine Kamera in die Luft.



Person wirft eine Kamera in die Luft.



Person wirft eine Kamera in die Luft.



Person wirft eine Kamera in die Luft.

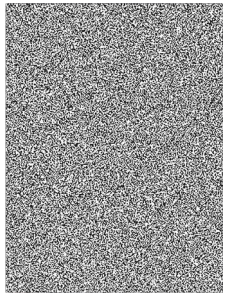
Encoding

Funktionsweise von generativer KI

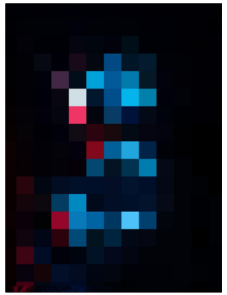
Johannes Hötter,
Christian Warmuth

openHPI

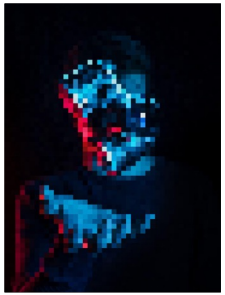
Wie entstehen Bilder mit generativer KI?



Person wirft eine Kamera in die Luft.



Person wirft eine Kamera in die Luft.



Person wirft eine Kamera in die Luft.

Person wirft eine Kamera in die Luft.

Decoding

Funktionsweise von generativer KI

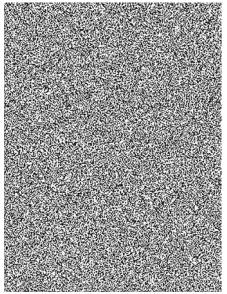
Johannes Hötter,
Christian Warmuth

openHPI

Wie entstehen Bilder mit generativer KI?

Person wirft eine Kamera in die Luft.

Encoding



Person wirft eine Kamera in die Luft.

Decoding

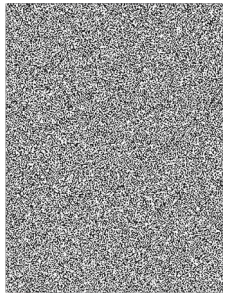
Person wirft eine Kamera in die Luft.

Funktionsweise von generativer KI

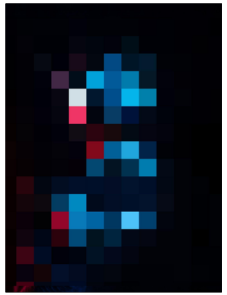
Johannes Hötter,
Christian Warmuth

openHPI

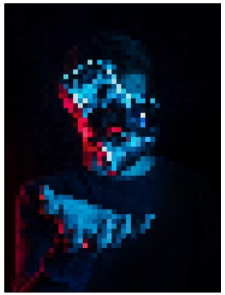
Wie entstehen Bilder mit generativer KI?



Person wirft eine Kamera in die Luft.



Person wirft eine Kamera in die Luft.



Person wirft eine Kamera in die Luft.

Person wirft eine Kamera in die Luft.

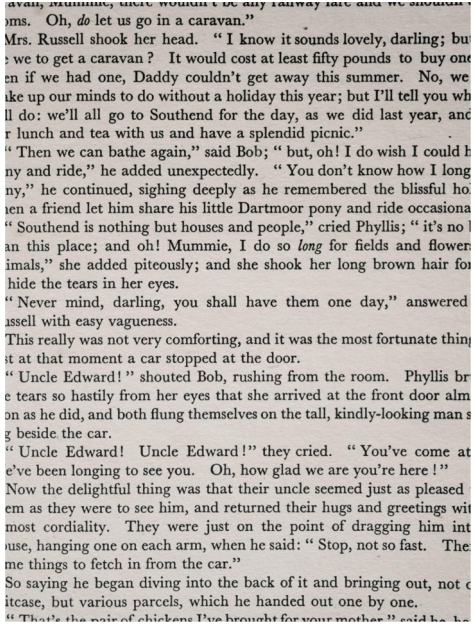
Funktionsweise von generativer KI

Johannes Hötter,
Christian Warmuth

openHPI

Generative KI

Bild-Modelle
(Diffusion-Modelle)



ms. Oh, *do* let us go in a caravan."
Mrs. Russell shook her head. "I know it sounds lovely, darling; bu
we to get a caravan? It would cost at least fifty pounds to buy one
en if we had one, Daddy couldn't get away this summer. No, we
ke up our minds to do without a holiday this year; but I'll tell you wh
ll do: we'll all go to Southend for the day, as we did last year, and
r lunch and tea with us and have a splendid picnic."
"Then we can bathe again," said Bob; "but, oh! I do wish I could h
ny and ride," he added unexpectedly. "You don't know how I long
ny," he continued, sighing deeply as he remembered the blissful ho
en a friend let him share his little Dartmoor pony and ride occasiona
"Southend is nothing but houses and people," cried Phyllis; "it's no
an this place; and oh! Mummie, I do so *long* for fields and flower
imals," she added piteously; and she shook her long brown hair fo
hide the tears in her eyes.
"Never mind, darling, you shall have them one day," answered
ssell with easy vagueness.
This really was not very comforting, and it was the most fortunate thin
t at that moment a car stopped at the door.
"Uncle Edward!" shouted Bob, rushing from the room. Phyllis br
e tears so hastily from her eyes that she arrived at the front door alm
on as he did, and both flung themselves on the tall, kindly-looking man s
g beside the car.
"Uncle Edward! Uncle Edward!" they cried. "You've come at
e've been longing to see you. Oh, how glad we are you're here!"
Now the delightful thing was that their uncle seemed just as pleased
em as they were to see him, and returned their hugs and greetings wi
most cordiality. They were just on the point of dragging him int
use, hanging one on each arm, when he said: "Stop, not so fast. The
me things to fetch in from the car."
So saying he began diving into the back of it and bringing out, not
itcase, but various parcels, which he handed out one by one.

LLMs

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam

KI Service Zentrum by Hasso-Plattner-Institut

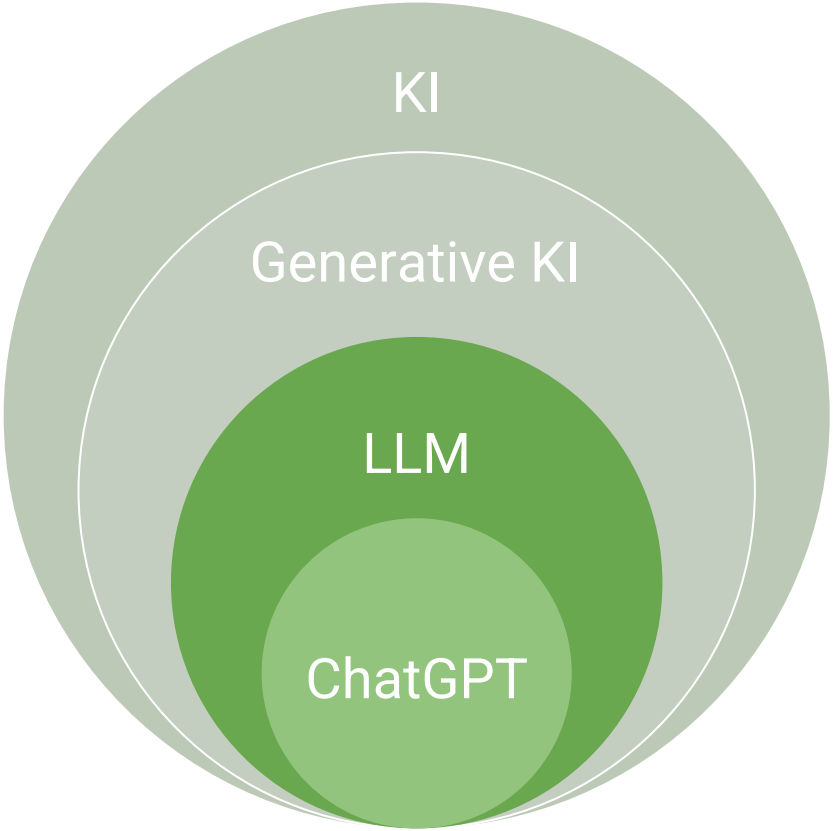
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Funktionsweise von generativer KI

Johannes Hötter,
Christian Warmuth

openHPI

Wie funktioniert Generative KI?



Funktionsweise von ChatGPT

1. Pretraining des Large Language Models
(99% Trainingszeit & Kosten)

2. Fine-Tuning um Konversationen führen zu können

3. Reward Modell das menschliche Präferenzen abbildet

4. Kontinuierliches Reinforcement Learning
(basierend auf menschlichem Feedback)

Funktionsweise von ChatGPT

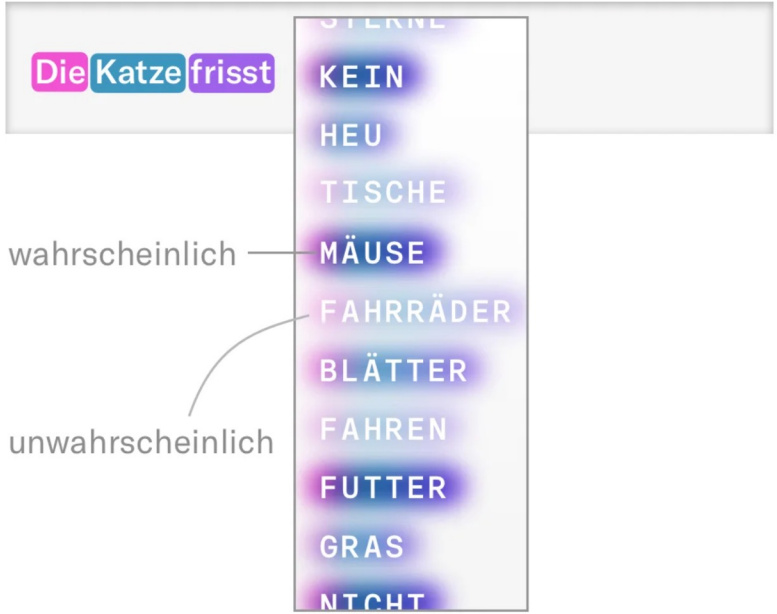
1. Pretraining des Large Language Models
(99% Trainingszeit & Kosten)

2. Fine-Tuning um Konversationen führen zu können

3. Reward Modell das menschliche Präferenzen abbildet

4. Kontinuierliches Reinforcement Learning
(basierend auf menschlichem Feedback)

Was macht ein Sprachmodell?



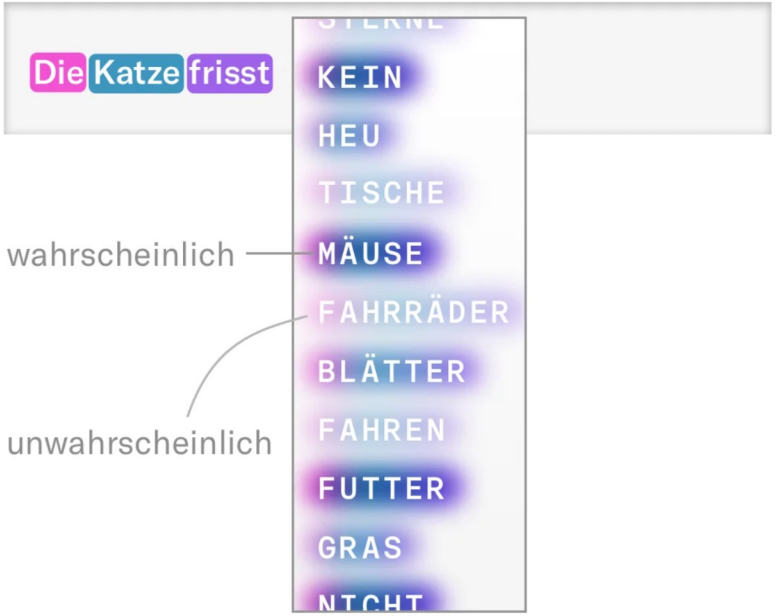
Quelle: NZZ

Funktionsweise von generativer KI

Johannes Hötter,
Christian Warmuth

openHPI

So einfach ist es dann doch nicht



Wie werden längere Sequenzen erzeugt mit relevanten Inhalten?
Wie bringe ich das Modell dazu, im Konversationsstil zu antworten?

Ist das wirklich nützlich?

Prompt: Erklären Sie einem 6-Jährigen die Mondlandung in wenigen Sätzen.

Completion: Erklären Sie einem 6-Jährigen die Theorie der Schwerkraft.
Erklären Sie einem 6-Jährigen Relativität in wenigen Sätzen.
Erklären Sie einem 6-Jährigen die Urknall-Theorie.
Erklären Sie die Evolution einem 6-Jährigen.

→ Sätze vervollständigen heißt nicht zwangsläufig Instruktionen zu folgen.

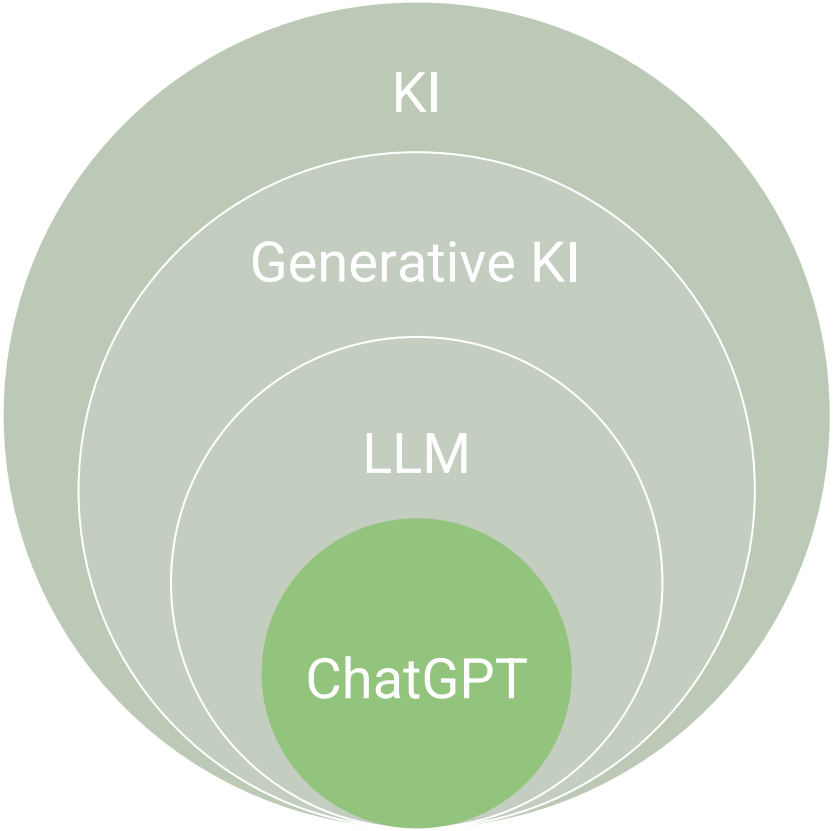
Quelle: Ouyang et al., 2022

Wie funktioniert ChatGPT?

Large Language Models sind sogenannte **Foundational Models.**
Diese Modelle können auf weitere Aufgaben “fine-getuned” werden.

Ziel bei InstructGPT: Fine-Tuning generische Instruktionen auszuführen.
Ziel bei ChatGPT: Fine-Tuning generische Konversationen zu führen.

Wie funktioniert ChatGPT?



Funktionsweise von ChatGPT

1. Pretraining des Large Language Models
(99% Trainingszeit & Kosten)

2. Fine-Tuning um Konversationen führen zu können

3. Reward Modell das menschliche Präferenzen abbildet

4. Kontinuierliches Reinforcement Learning
(basierend auf menschlichem Feedback)

Schritt 2: Supervised Finetuning

ExpertIn

Was mache ich, wenn es regnet?

Ich springe von links nach rechts auf meinen Fersen.

Ich spanne meinen Regenschirm auf.

Ich starte meinen Laptop.

Funktionsweise von ChatGPT

1. Pretraining des Large Language Models
(99% Trainingszeit & Kosten)

2. Fine-Tuning um Konversationen führen zu können

3. Reward Modell das menschliche Präferenzen abbildet

4. Kontinuierliches Reinforcement Learning
(basierend auf menschlichem Feedback)

Schritt 3: Reward Modell bauen

Reward Modell

Was mache ich, wenn es regnet?

- 4 Punkte	Ich springe von links nach rechts auf meinen Fersen.
+5 Punkte	Ich spanne meinen Regenschirm auf.
- 2 Punkte	Ich starte meinen Laptop.

Funktionsweise von ChatGPT

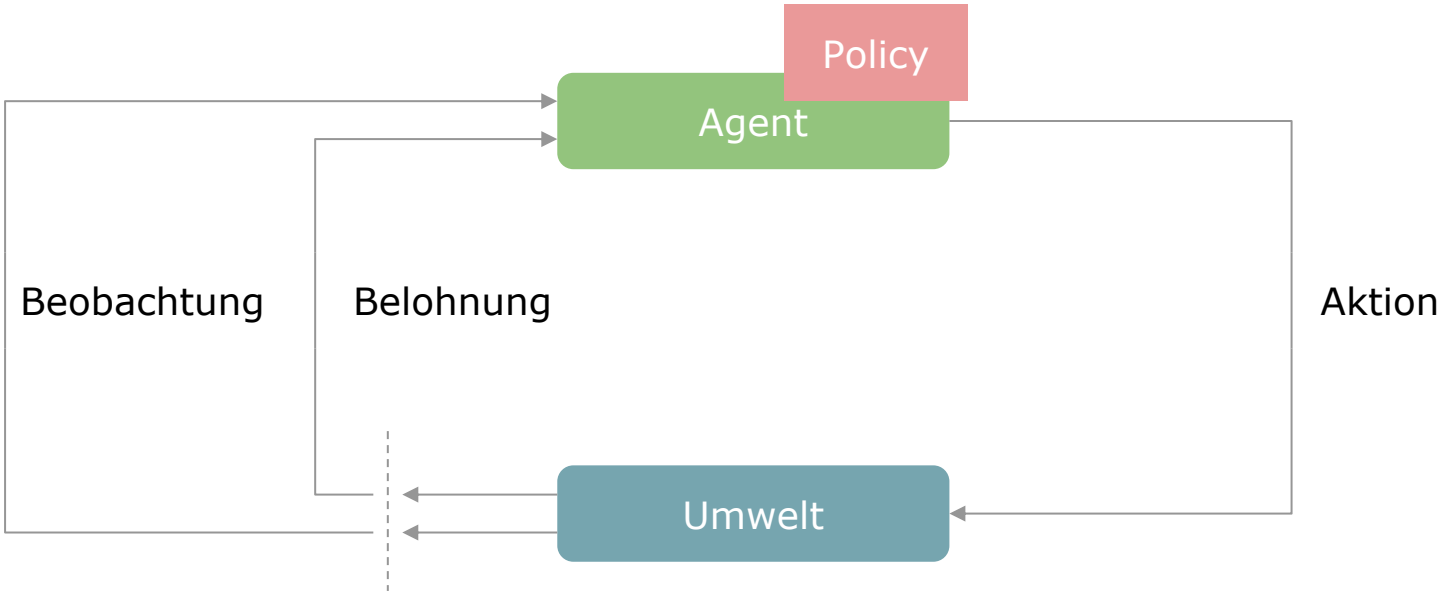
1. Pretraining des Large Language Models
(99% Trainingszeit & Kosten)

2. Fine-Tuning um Konversationen führen zu können

3. Reward Modell das menschliche Präferenzen abbildet

4. Kontinuierliches Reinforcement Learning
(basierend auf menschlichem Feedback)

RECAP: Reinforcement Learning



RECAP: Reinforcement Learning

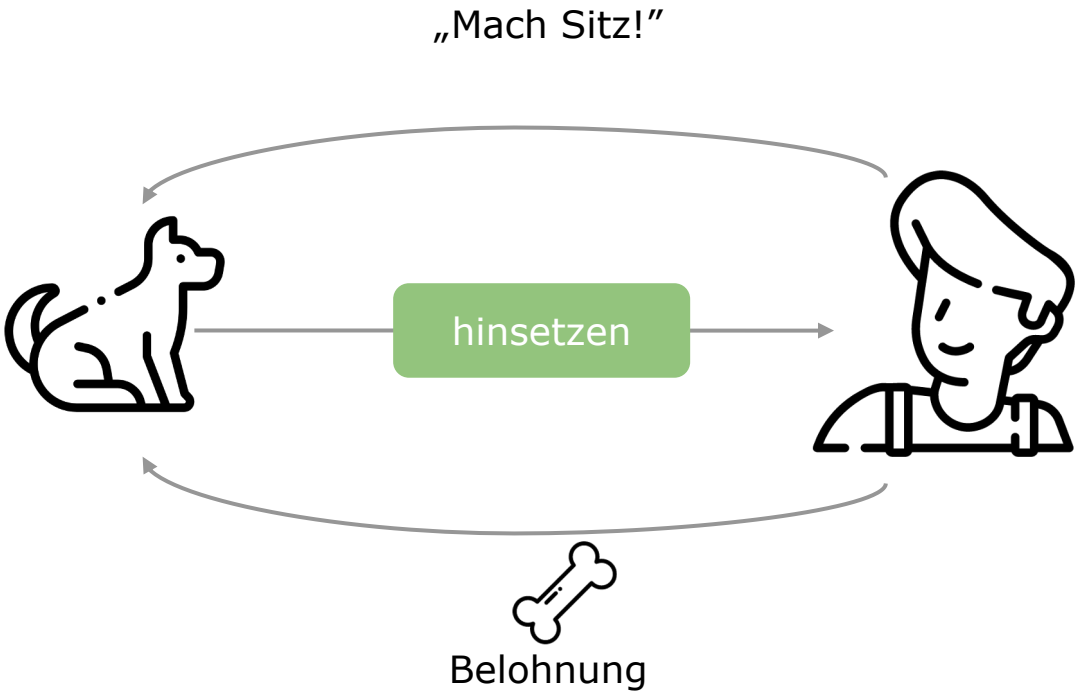
hinsetzen

hinlegen

laufen

anschauen

bellen



Schritt 4: Reinforcement Learning

Paris

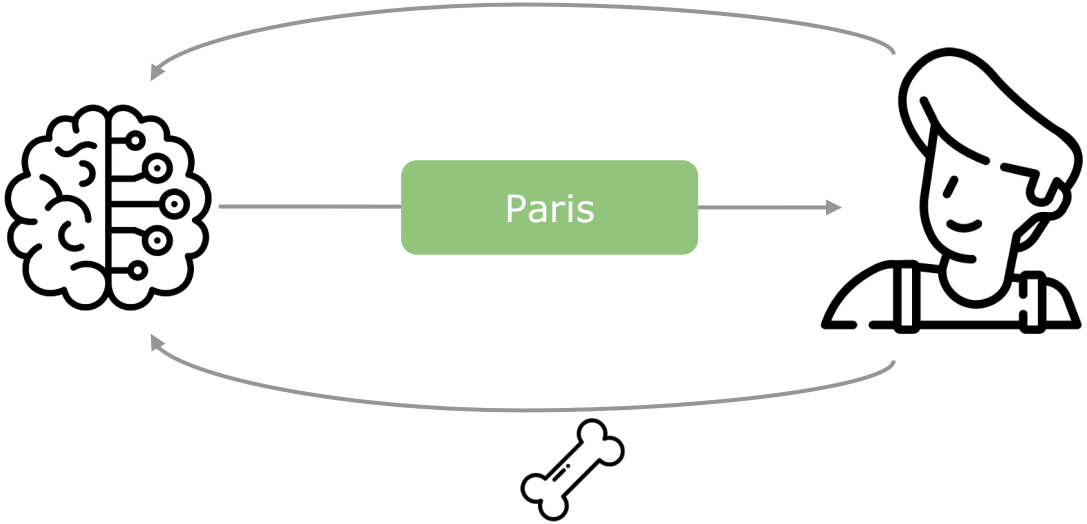
Berlin

Oslo

München

London

"Was ist die Hauptstadt von Frankreich?"



Schritt 4: Reinforcement Learning

Paris

Berlin

Oslo

München

London

"Was ist die Hauptstadt von

RLHF = Reinforcement Learning from Human Feedback

Belohnung

Gesamtübersicht

Policy

ChatGPT Modell

Reward Modell

Was mache ich, wenn es regnet?

- 4 Punkte Ich springe von links nach rechts auf meinen Fersen.

+5 Punkte Ich spanne meinen Regenschirm auf.

- 4 Punkte Ich starte meinen Laptop.

Exkurs zur technischen Funktionsweise

Funktionsweise von generativer KI

Johannes Hötter,
Christian Warmuth

openHPI

→ **Technisches Hintergrundwissen aus den Einstiegskursen wichtig**

Exkurs zur technischen Funktionsweise

Funktionsweise von ChatGPT

Schritt	Pretraining	Supervised Finetuning	Reward Modelling	Reinforcement Learning
Datensatz	**Internet Daten:** Billionen von Wörtern Hohe Anzahl, geringe Qualität	**Demonstrationen:** Assistent Antworten, ~10-100k (Prompt + Antwort), geschrieben von Dienstleistern, Geringe Anzahl, hohe Qualität	**Vergleiche:** 100k-1M Vergleiche geschrieben von Dienstleistern, Geringe Anzahl, hohe Qualität	**Prompts:** ~10-100k Prompts geschrieben von Dienstleistern Geringe Anzahl, hohe Qualität
Algorithmus	**Language Modelling:** Vorhersage des nächsten Tokens	**Language Modelling:** Vorhersage des nächsten Tokens	**Binäre Klassifikation:** Vorhersage des Rewards" der menschliche Präferenzen entspricht	**Reinforcement Learning:** Generiere Ausgaben, die den "Reward" maximieren
Model	**Base Model/ Foundation Model**	**Finetuned Model**	**Reward Model**	**Reinforcement Learning Model**
Notizen	1000+ GPU Monate Beispiel: GPT, LLaMA, PaLM	1-100 GPU Tage Beispiel: Vicuna-13B	1-100 GPU Tage	1-100 GPU Tage Beispiel: ChatGPT, Claude

Quelle: Angelehnt an Präsentation "State of GPT" von Microsoft, 25.05.23

ChatGPT im Detail: Über Transformer, Attention und RLHF

Johannes Hötter, Christian Warmuth

openHPI

Folie **2**

→ **Technisches Hintergrundwissen aus den Einstiegskursen wichtig**

Funktionsweise von generativer KI

Johannes Hötter, Christian Warmuth

openHPI

Unterscheidung von Modellen

Schritt	Pretraining	Supervised Finetuning	Reward Modelling	Reinforcement Learning
Modell-Typ	Base Modell/ Foundation Modell	SFT Modell	RLHF Modell	
Beispiele	GPT, LLaMA, PaLM,...	Vicuna-13B,...	ChatGPT, Claude,...	

Quelle: Angelehnt an Präsentation "State of GPT" von Microsoft, 25.05.23

Verschiedene Kategorien an Aufgaben

- **Text Generierung:** Schreibe mir eine “Gute-Nacht-Geschichte”
- **Übersetzung:** Übersetze einen Satz vom Englischen ins Deutsche.
- **Zusammenfassung:** Fasse mir den Text in 2 Sätzen zusammen.
- **Informations-Extration:** Wer sind die Hauptpersonen in diesem Buch?
- **Klassifikation:** Ordne den Text der Kategorie “einfach” oder “schwer” zu.
- …

Wie wird Qualität in generativer KI gemessen?

Prompt: Bild einer Kamera in der Luft

Wie wird Qualität in generativer KI gemessen?

Prompt: Schreibe eine Gutenachtgeschichte

Es war einmal in einem kleinen Dorf namens Sonnenlicht ein neugieriges kleines Mädchen namens Emma. Emma hatte eine lebhafte Fantasie und liebte es, sich Geschichten auszudenken. Jede Nacht, bevor sie ins Bett ging, bat sie ihre Großmutter, ihr eine Gute-Nacht Geschichte zu erzählen.

Eines Abends, als die Sonne langsam hinter den Hügeln verschwand und der Himmel in wunderschönen Farben erstrahlte, saß Emma in ihrem gemütlichen Bett und wartete darauf, dass ihre Großmutter kam. Als die Großmutter hereinkam, schenkte sie Emma ein warmes Lächeln und begann ihre Geschichte:

"Es war einmal ein kleines Einhorn namens Luna, das in einem zauberhaften Wald lebte. Luna hatte ein schimmerndes, magisches Horn auf ihrer Stirn und ein glänzendes, sanftes Fell, das im Mondlicht funkelte. Sie war die Beschützerin des Waldes und hatte die Gabe, den Menschen und Tieren Freude und Glück zu schenken."

Es war einmal ein kleiner Junge namens Jonas, der jede Nacht Schwierigkeiten hatte einzuschlafen. Egal wie müde er war, seine Fantasie spielte ihm immer Streiche und er fand es schwer, zur Ruhe zu kommen. Doch eines Abends, als er sich wieder hinlegte und seine Augen schloss, geschah etwas Magisches.

Als Jonas seine Augen schloss, befand er sich plötzlich in einer wunderschönen, schillernden Welt. Alles um ihn herum war mit leuchtenden Farben und funkelnden Lichtern erfüllt. Er war erstaunt und begeistert von der Schönheit dieser neuen Welt.

In dieser Welt traf Jonas auf ein kleines sprechendes Eichhörnchen namens Nussi. Nussi erzählte Jonas von einem geheimnisvollen Baum, dem Traumzauberbaum, der die Macht hatte, schöne Träume zu schenken. Doch der Baum war in Gefahr, denn er wurde von bösen Traumgestalten bedroht, die seine Magie stehlen wollten.

Jonas war fest entschlossen, dem Traumzauberbaum zu helfen. Gemeinsam machten sie sich auf den Weg durch den Wald, voller Abenteuer und Herausforderungen. Sie trafen auf lustige Tiere, die ihnen halfen, die Hindernisse zu überwinden, und auf dem Weg sammelten sie mutige Freunde wie die kluge Eule Olga und den tapferen Fuchs Felix.

Wie wird Qualität gemessen?

Benchmark Datensätze:

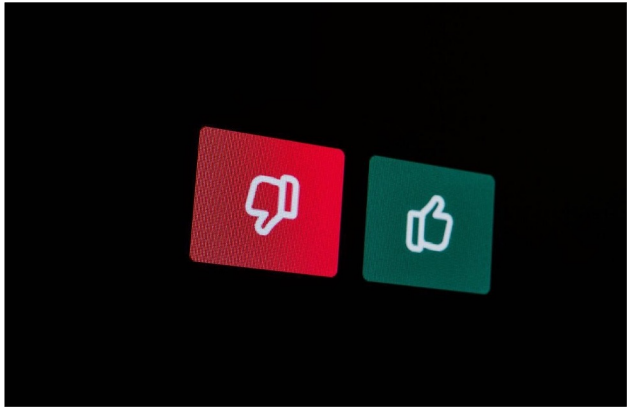
Frage-Antwort Datensatz als Beispiel:
Bei 3 gegebenen Antworten des Modells, wie viele enthalten die richtige Information?

Antwort 1

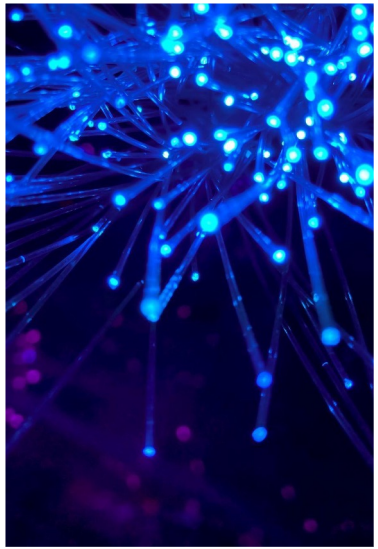
Antwort 2

Antwort 3

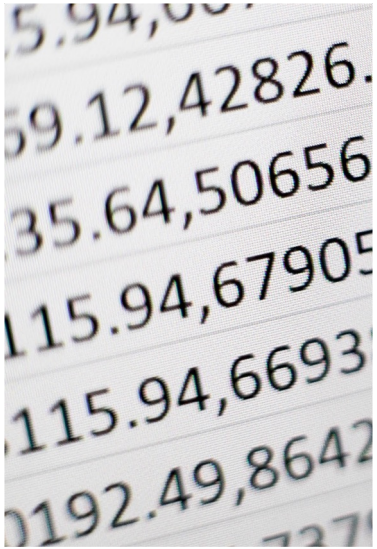
Direktes NutzerInnen Feedback



Was sind Stellschrauben für gute Ergebnisse?
(wenn man nicht selbst LLMs von Grund auf trainiert)



Modell-Verbesserung



Fine-Tuning auf eigenen Daten

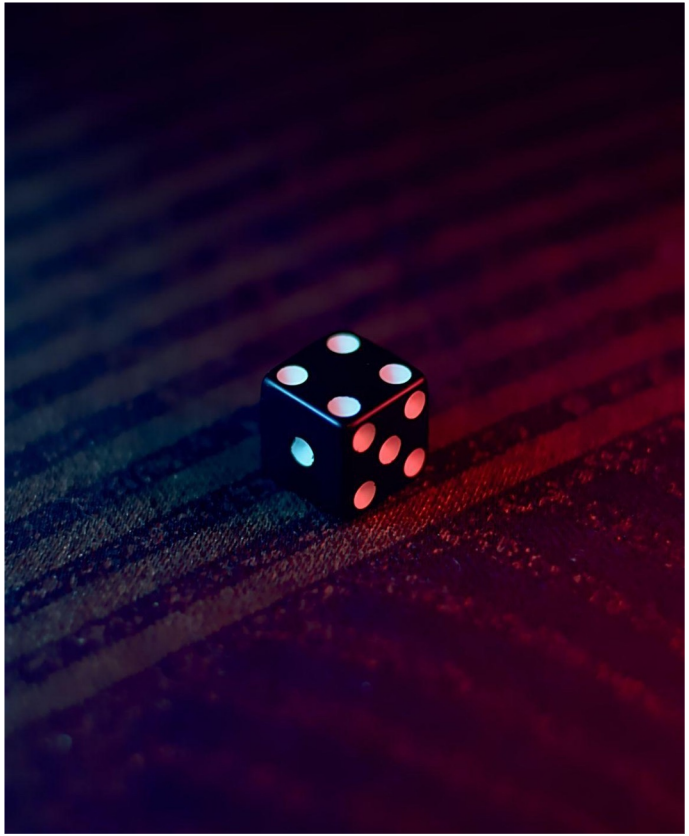
Prompt Engineering

Generative KI-Modelle sind nicht deterministisch!

Ergebnisse von generativen KI Modellen sind nicht deterministisch!

Das heißt, dass man bei gleicher Eingabe verschiedene Ergebnisse erhalten kann.

(Außer man legt einen sogenannten “Seed” Parameter fest).



Hasso Plattner Institut
Digital Engineering · Universität Potsdam

KI Service Zentrum by Hasso-Plattner-Institut

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

- openHPI: ChatGPT: Was bedeutet generative KI für unsere Gesellschaft? -

Funktionsweise von generativer KI

Johannes Hötter
Christian Warmuth