

Künstliche Intelligenz: Revolution oder Risiko?

Teil 4: ChatGPT selbst gemacht Ein Leitfaden

Philipp Gerber

29. Februar 2024



IT Crowd Club Liechtenstein

informieren • austauschen • vernetzen

- Gegründet 2011
- 50 Mitglieder



itcc.li

Künstliche Intelligenz: Revolution oder Risiko?

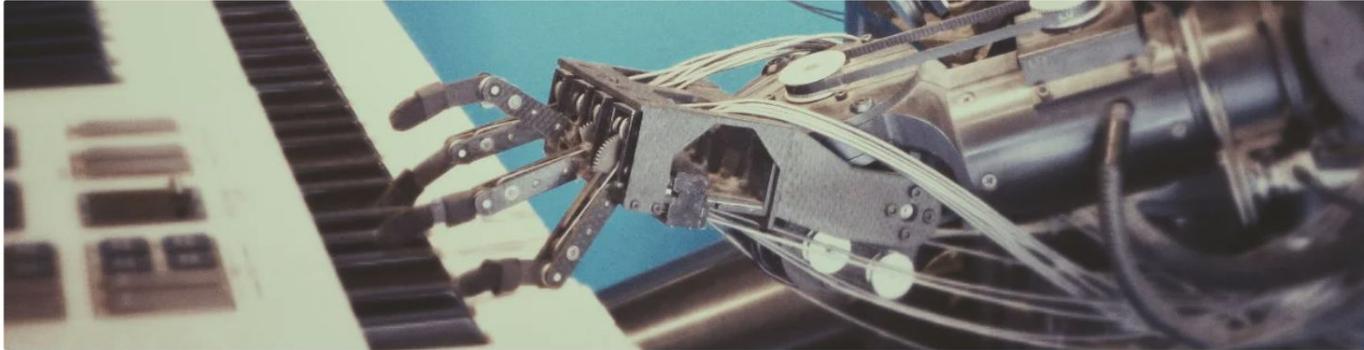


Foto: Klavierspielender Roboter (Unsplash)

Vier spannende Vorträge zum Thema Generative KI

[Jetzt kostenlos anmelden!](#)

Lerne die Zukunftstechnologie kennen

In einer Zeit, in der Technologie unser Leben immer mehr beeinflusst, ist Künstliche Intelligenz (KI) ein Thema, das uns alle betrifft. KI ist in der Lage, Aufgaben zu übernehmen, die bisher von Menschen erledigt wurden, und sie hat das Potenzial, unsere Gesellschaft in vielerlei Hinsicht zu verändern. Ob KI eine Revolution oder ein Risiko darstellt, ist eine Frage, die uns alle beschäftigt.

KI: Revolution oder Risiko?



KI: Revolution oder Risiko?

Do., 29. Februar 2024

4. ChatGPT selbst gemacht: Ein Leitfaden

Du möchtest erfahren, wie du mit einem KI-Modell erst

Wir zeigen dir, wie Modelle trainiert werden

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

Do., 25. Januar 2024

3. Die Kunst des Prompt Engineerings

In diesem Vortrag erfährst du, wie du Prompts schreiben kannst und sie für deine Zwecke optimieren kannst

Wir zeigen dir, wie Prompt-Hacking funktioniert und die Einschränkungen von KI-Systemen

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

Do., 30. November 2023

2. KI verstehen: Von Neuronen zu ChatGPT

In diesem Vortrag erfährst du alles über die Grundlagen der KI, vom Neuronalen Netz bis zum LLM (Large language model).

Du wirst lernen, wie generative KI-Systeme funktionieren und wie sie trainiert werden. Im Mittelpunkt der Diskussion steht die Frage, ob ChatGPT wirklich intelligent ist.

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

ELIZA-Effekt

Menschen nehmen textbasierte Dialogsysteme als intelligent und empathisch wahr, selbst wenn diese Systeme nur darauf trainiert sind, menschliche Sprache zu imitieren.

Was ist Künstliche Intelligenz?

Künstliche Intelligenz (KI)

ist die Fähigkeit einer Maschine,

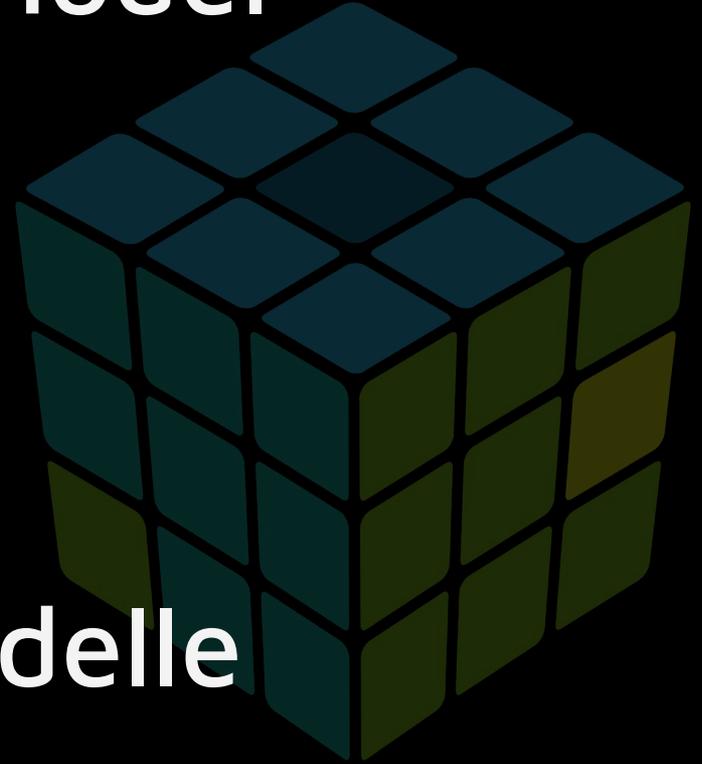
menschliche Fähigkeiten wie

logisches Denken, Lernen, Planen und Kreativität

zu **imitieren.**

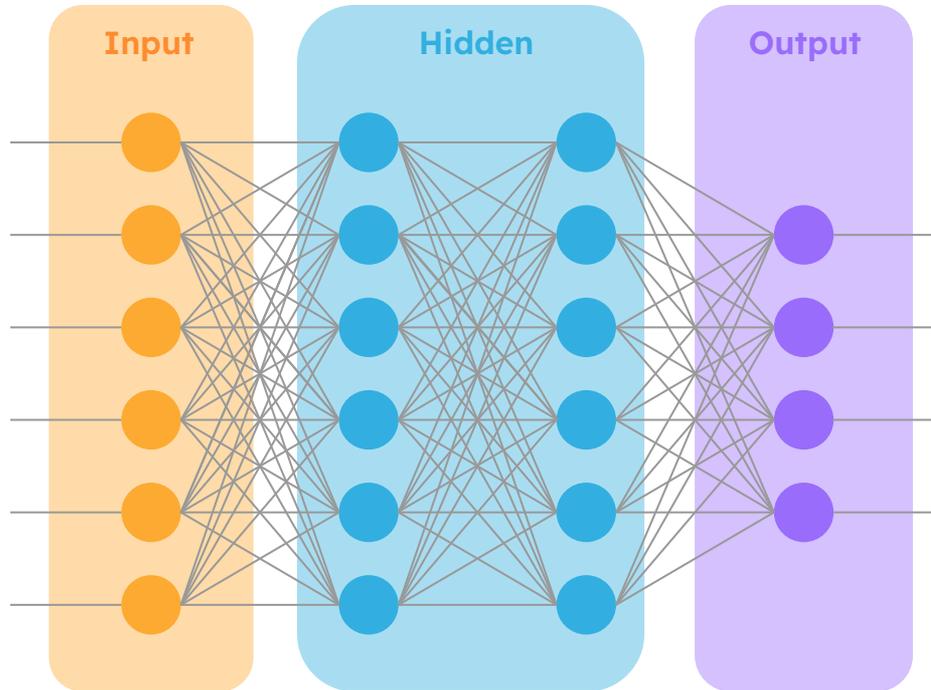
[[Europäisches Parlament](#)]

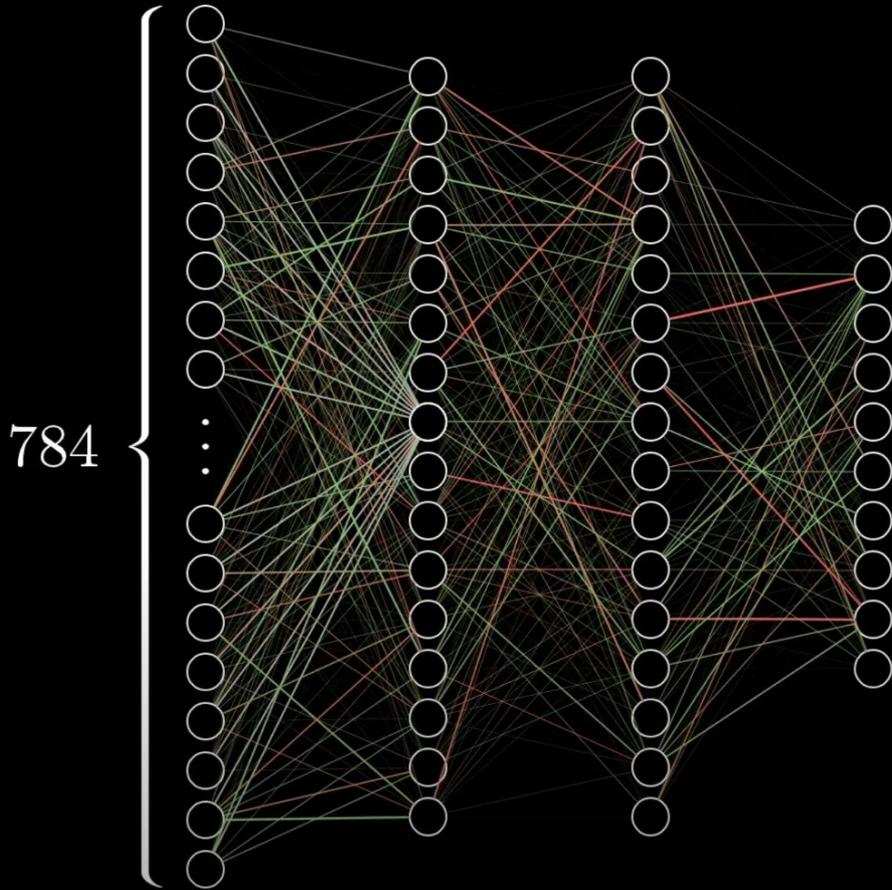
Large Language Model
LLM



Grosse Sprachmodelle

Künstliches Neuronales Netz





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

weights

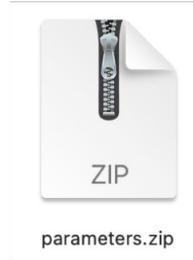
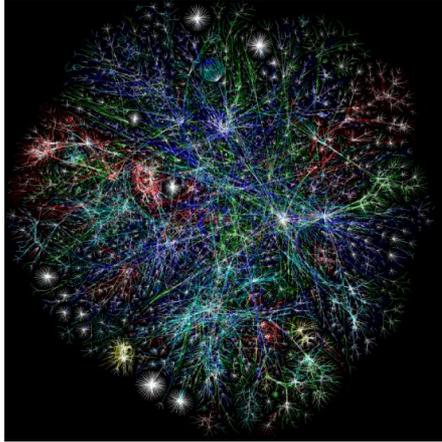
$$16 + 16 + 10$$

biases

13,002

Learning \rightarrow Finding the right weights and biases

Large Language Model



Trainingsdaten
~10TB Text

Pretraining
6'000 GPUs, 12 Tage,
~\$2M, ~1e24 FLOPS

LLM
~140GB

Llama 2 70B

Pre-Training und Fine-Tuning

Pre-Training

~10TB Text
6'000 GPUs
12 Tage
~\$2M

Llama 2 70B

Grundmodell

z.B. Llama 2, GPT3
Generative Pre-trained
Transformer

Fine-Tuning

100'000 Q&A
1 Tag

+Reinforcement
Learning

Assistentenmodell

z.B. Llama 2 Chat, ChatGPT3

Fine-Tuning (OpenAI)

Step 1

Collect demonstration data and train a supervised policy.

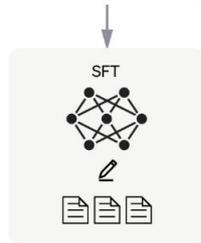
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



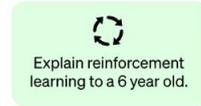
This data is used to fine-tune GPT-3.5 with supervised learning.



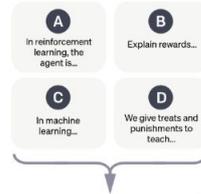
Step 2

Collect comparison data and train a reward model.

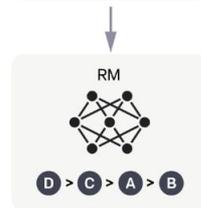
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



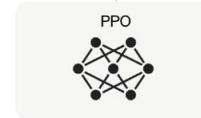
Step 3

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

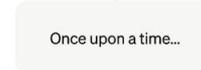
A new prompt is sampled from the dataset.



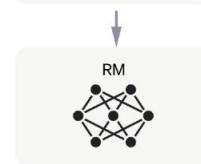
The PPO model is initialized from the supervised policy.



The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.



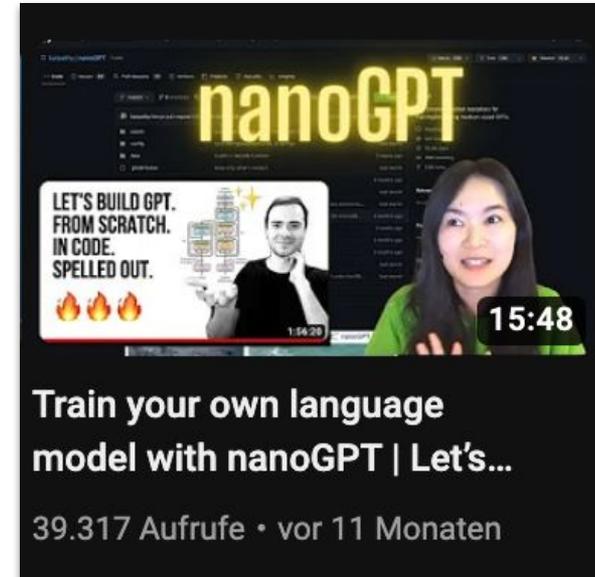
NanoGPT von Andrej Karpathy



Bauen wir GPT: von Grund auf



Andrej Karpathy



Sophia Yang

nanoGPT (1: BigramLanguageModel)

Trainingsdaten ~40 MB

Pretraining 5000 Iterationen, 15 Minuten



Google Colaboratory



[\[gpt-dev.ipynb\]](#)

nanoGPT (1: BigramLanguageModel)

Trainingsdaten ~40 MB

Pretraining 5000 Iterationen, 15 Minuten

```
# generate from the model  
print(decode(m.generate(context, max_new_tokens=2000)[0].tolist()[1:]))
```

```
zweieckellig gehoben. Auch der Ich besondere in Konnern  
getmal Lücher am Zeizen nach gegenügt immer sunderbirchten polx  
ring. Um doch nach dann dieses müsse Popötiz zugebemen, Gesitze  
auf, die ganzen sollen mit einem feigen Stein  
die die Menschenen der Schre trat, spilt sind er uns einer etwarf in  
sam sehr unter Klantz hinalseluren Morg ein Entwell, (Gehl.22, Mind. ....  
Zwiv 35.
```

```
Da Saneltd,). Theuversitzen ließlieger Erleusse, oders zweiterwider  
auf den sind vom in Geudeuting aber nur versollen, sie  
nicht, emben dassen andliches aus den durch gewessen;  
wir Hold im Kergen, sind spiest, in der eplieitern küßteraustürzen  
bei der Bototvrüte nur die bei Tages, einem  
Unter letzten Krauten könnte.
```

Trainingsdaten – Text und Tokens



Tokens

Characters

84

303

[644, 17495, 29931
56497, 26612, 1902
735, 60491, 20603,
4466, 62250, 11, 6
735, 40, 6127, 304
14104, 38160, 11,
18400, 48750, 11,
532, 11, 45467, 21
1354, 4970, 6529,

Text

Token IDs

Tokens

Characters

84

303

In einer Zeit, in der Technologie unser Leben immer mehr beeinflusst, ist Künstliche Intelligenz (KI) ein Thema, das uns alle betrifft. KI ist in der Lage, Aufgaben zu übernehmen, die bisher von Menschen erledigt wurden, und sie hat das Potenzial, unsere Gesellschaft in vielerlei Hinsicht zu verändern.

Text

Token IDs

nanoGPT (2: tiktoken, gpt2)

Trainingsdaten ~20 MB (Projekt Gutenberg)

Pretraining 50'000 Iterationen, 8 Stunden

```
Die Welt kam ein,  
Und auf dem Tod.  
Das kann nicht mehr  
Ich bin nicht bei mir,  
Die Frau,  
Ich habe ihn gefunden.  
Ich habe ein Leben,  
Ich bin ein Leben,  
Der Herr,  
Nur ein Mädchen,  
Das ist ein schönes  
Künstler, das ist ein  
Neben,  
Das ist ein ganzes  
Wesen, und ich weiß,  
Daß ich nicht,  
wie ich auf dem Tage  
das Licht, das ist keine
```

```
Herr Meister:  
Ich will nicht leben, was ich mein Freund?  
Ich habe ihn jetzt nicht gewesen, und ich habe  
sie nicht mehr bewegen.  
Aber wenn dann, wenn ich es nicht gesehen  
kann, wenn ich mich nicht mehr wiederholen  
konnte.  
Ich kann nicht einmal einmal  
sehen, und ich kann nicht mehr  
nicht verwirrt.  
Vielleicht werde ich nicht  
mehr herumgehen.  
Der Konsul war ein kleiner  
Herrn, der ihm an ihrem  
Kopf, und der weiße Weihnachtschlagen  
und das Gesicht.  
Das war auch einmal, und  
daß sie nicht mehr gesehen  
hatte, sie waren einmal,  
und die ganze Stunden  
hätten sich auf den Kopf zu  
schlagen.
```

Kontext oder Verknüpfung von Wörtern

Der Schlüssel zum Textverständnis

Attention

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

nanoGPT (2: tiktoken, gpt2)

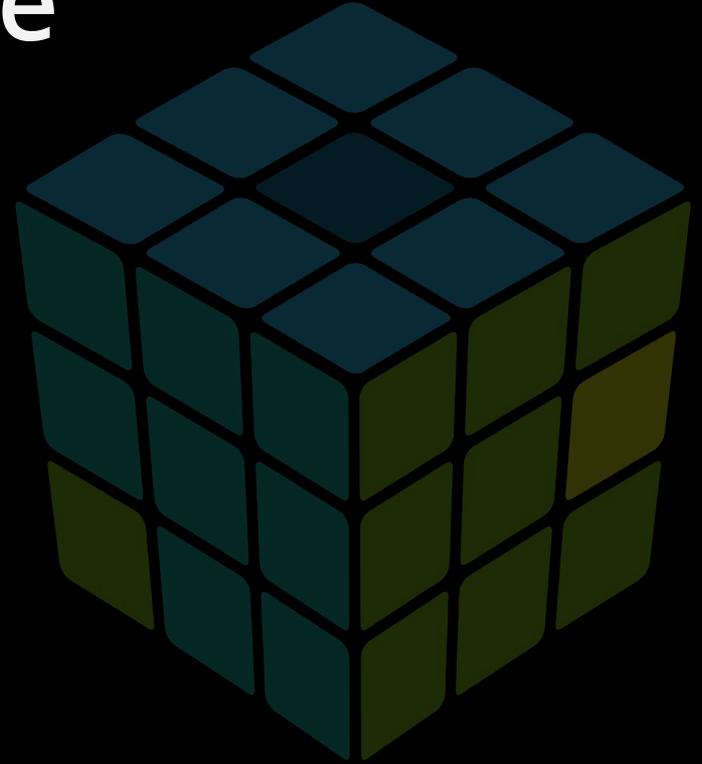
Trainingsdaten ~40 MB

Pretraining 150'000 Iterationen, 20 Stunden

```
Er sagte zu ihr erschlagen:  
Ich hab es nicht gesagt, daß ich auch nicht  
zum Weiterreisen kommen werde.  
Ich werde dir nicht vergessen, daß du so  
schuldig sein wirst, daß du sie nicht zu sagen  
hören kannst.  
Wie schön, wie ein Mensch, einmal das  
wird ich dir nicht wissen.  
Wie?  
Nun, das ist doch nicht so, ich würde  
nicht recht.
```

```
Das ist ja schon nichts.  
Das ist es, wenn man sie nicht überhaupt  
nicht versteht, daß man sich nicht weniger  
hätte, als ob er sich einen kleinen Schuß  
hatte.  
Aber es wäre nicht gewesen, wenn er nicht  
auf die Augen und die Gäste nicht einmal  
fürchteten, daß er nicht zu erfüllen konnte.  
Da lachte er nicht.
```

Benutzerdefinierte ChatGPTs



Benutzerdefinierte ChatGPTs

OpenAI Research API ChatGPT Safety Company Search Log in Try ChatGPT

Blog

Introducing GPTs

You can now create custom versions of ChatGPT that combine instructions, extra knowledge, and any combination of skills.

Creative Writing Coach

I'm excited to read your work and give you feedback to improve your skills.



Laundry Buddy

Ask me anything about stains, settings, sorting and everything laundry.



Game Time

I can quickly explain board games or card games to players of any skill level. Let the games begin!



Tech Advisor

From setting up a printer to troubleshooting a device, I'm here to help you step-by-step.





New GPT

• Draft



Save

Create

Configure



GPT Builder

Hi! I'll help you build a new GPT. You can say something like, "make a creative who helps generate visuals for new products" or "make a software engineer who helps format my code."

What would you like to make?



You



skriptum-fl_210517_mit-inhaltsverzeich...
PDF



skriptum-nl_210517_mit-inhaltsverzeic...
PDF

Erstelle einen freundlichen und hilfreichen Assistenten, mit dem der Benutzer für den Staatskunde-Test in Liechtenstein lernen kann.



Message GPT Builder...



Preview



You

Gib mir eine Multiple-Choice-Frage.



Untitled

Hier ist eine Multiple-Choice-Frage aus dem Bereich "Das Staatswesen":

Frage: Jeder Staat besitzt eine politische Ordnung: Welche drei Merkmale muss er aufweisen?

Antwortmöglichkeiten:

- A) Staatsreligion, Staatspräsident, Staatswährung
- B) Staatsgebiet, Staatsvolk, Staatsgewalt
- C) Staatsflagge, Staatsfeiertag, Staatshymne
- D) Staatsgericht, Staatsbibliothek, Staatsarchiv

Die richtige Antwort ist B) Staatsgebiet, Staatsvolk, Staatsgewalt.



Message GPT...



Benutzerdefinierte ChatGPTs

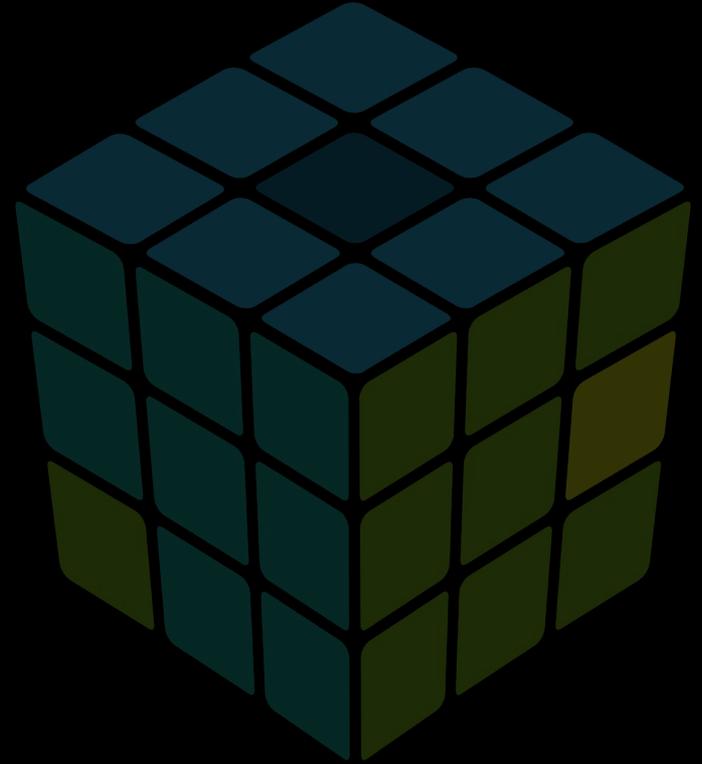
Prompt Engineering

- Vordefinierte System Prompts
- Fokus auf ein definiertes Thema
- Stil, Tonalität, etc.

Daten

- Benutzerdefinierte Daten und Dokumente

Open-Source-LLMs
lokal ausführen



Hugging Face



Hugging Face

„GitHub für Machine Learning“

- Modelle
 - LLM, Audio, Computer Vision, Image Models
- Trainingsdaten
- Community

<https://huggingface.co>

Ollama



Get up and running with large
language models, locally.

Run [Llama 2](#), [Code Llama](#), and other models.
Customize and create your own.

Download ↓

Available for macOS, Linux,
and Windows (preview)

<https://ollama.com/>

Ollama

```
ollama run mistral
pulling manifest
pulling e8a35b5937a5 ... 8% | 328 MB/4.1 GB 11 MB/s 5m28s
```

```
ollama run mistral
pulling manifest
pulling e8a35b5937a5 ... 100% 4.1 GB
pulling 43070e2d4e53 ... 100% 11 KB
pulling e6836092461f ... 100% 42 B
pulling ed11eda7790d ... 100% 30 B
pulling f9b1e3196ecf ... 100% 483 B
verifying sha256 digest
writing manifest
removing any unused layers
success
>>> Schreibe ein Gedicht über die Schönheit der
In bloom der schönen Natur wohnt die Sehnsucht,
Eines Traums, der tief in unserem Herzen ruht.
Der Himmel streut mit Sternen seine Weihnacht,
Und die Erde trägt das Kleid der Frühlingsblüter
In blauen Talen fließt der Quellensprudel,
Ein
```

```
ollama run mistral
Der Himmel streut mit Sternen seine Weihnacht,
Und die Erde trägt das Kleid der Frühlingsblüten aus.
In blauen Talen fließt der Quellensprudel,
Ein Lied in Silberton und in grünem Bandel.
Der Wald erwacht und singt mit Vögeln Lobgesang,
Auf jeder Blätterspitze spielt der Wind ein Tanz.
Der Sommer zieht mit goldenen Fäden an,
Und die Sonne küßt den Horizont, in Feuersternen fangt.
Die Wellen singen morgens in blauem Meer,
Und der Regen tanzt auf grünem Talwiesenreif.
Herbst bringt in Farben unser Landschaft zu leben,
Ein Bild von Schönheit und des Lebens Leben.
So lässt uns die Natur mit jeder Jahreszeit begeistern,
Und ruft uns ein, zu ihr hin, ein ewiges Reisender.
>>> Send a message (/? for help)
```

Jan.ai

 Meet Jan

Bringing AI to your Desktop

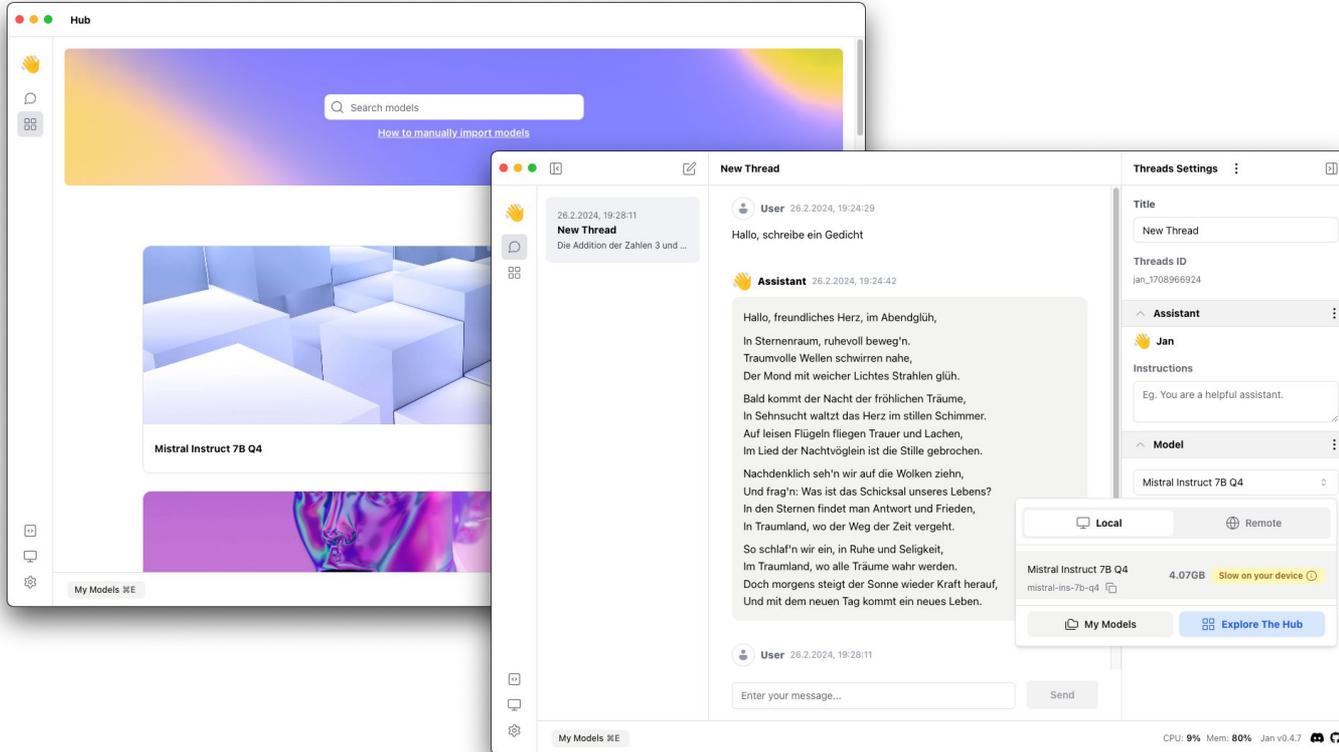
Open-source ChatGPT alternative that runs
100% offline on your computer.

 Download for Mac (M1/M2/M3) 

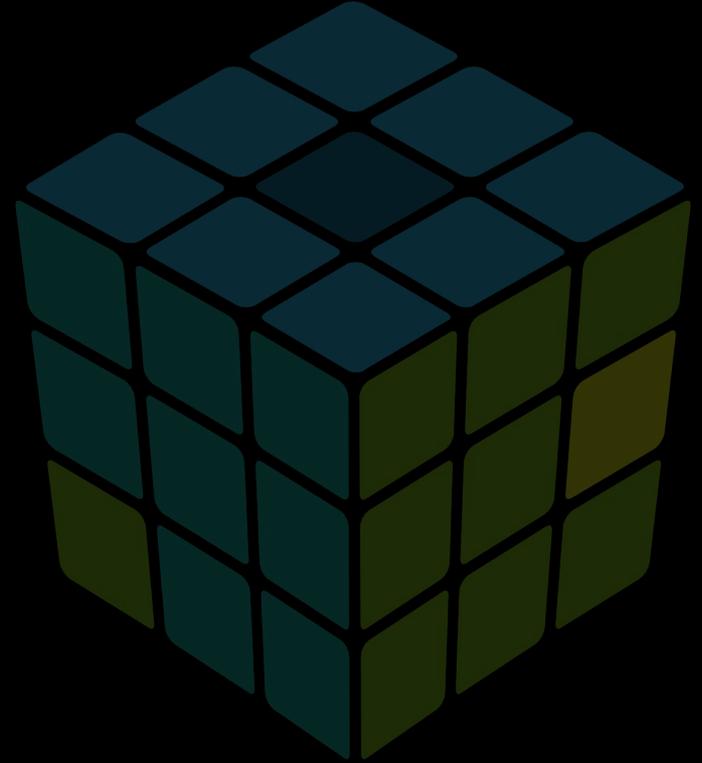
Find out more 

<https://jan.ai>

Jan.aiJan.ai



ELIZA-Effekt



KI sagt: 'Ich denke, also bin ich?'
Eher: 'Ich simuliere, also pass auf!'



Links

[The basics of neural networks](#)



3Blue1Brown

[Andrej Karpathy](#)

[Andrej Karpathy - YouTube](#)

[Andrej Karpathy - GitHub](#)



Andrej Karpathy

I like to train deep neural nets on large datasets 🧠🔥



Links

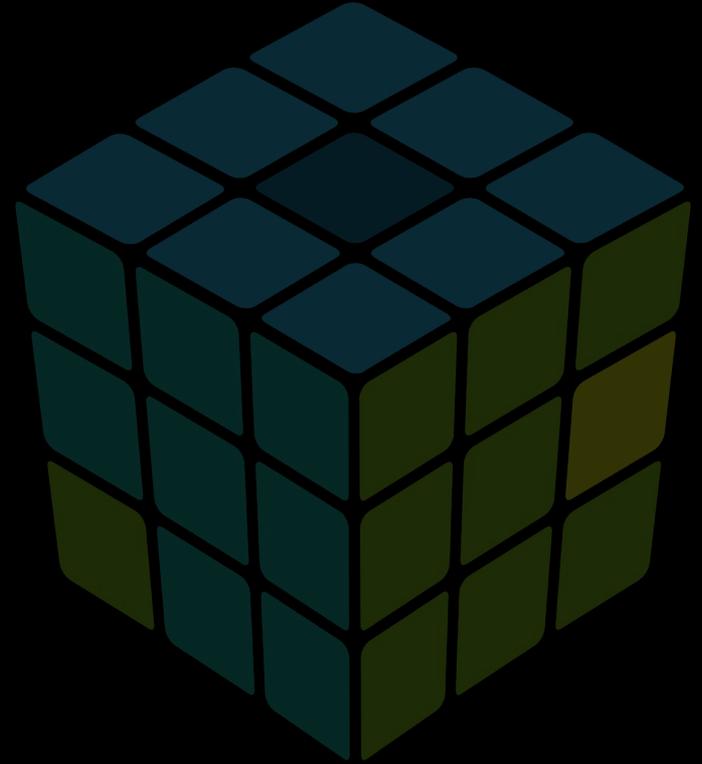
Open-Source-LLMs lokal ausführen

- [Ollama](#) 🦙
- [Jan.ai](#) 🖐️

Modell-Hub-Repository

- [Hugging Face](#) 🤗

Fragen?



Vielen Dank!

Mitglied werden

<https://itcc.li>

