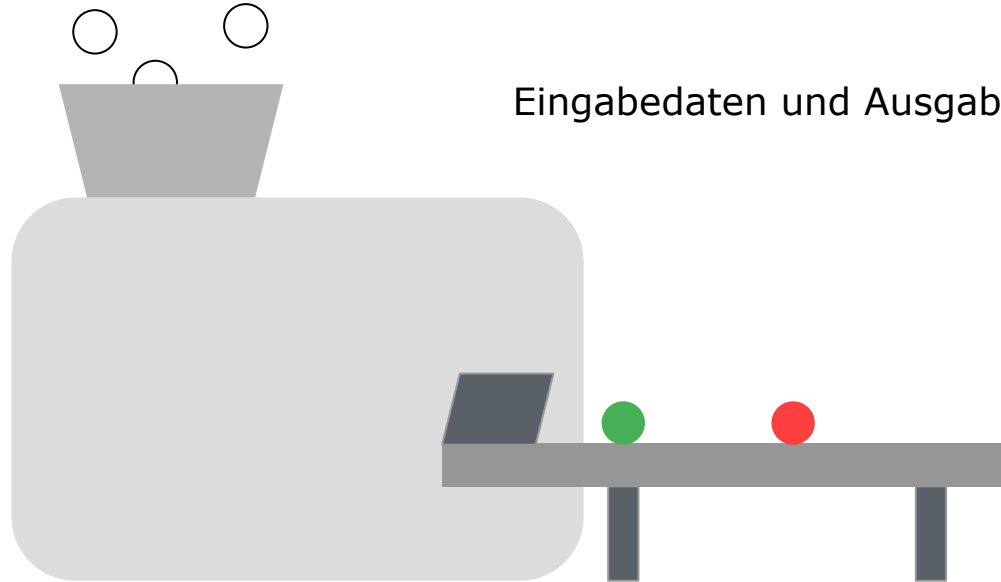


- openHPI: ChatGPT: Was bedeutet generative KI für unsere Gesellschaft? -

# Was ist datenzentrierte KI?

Johannes Hötter  
Christian Warmuth

# Ist nicht jedes Machine Learning-Modell datenzentriert?



Eingabedaten und Ausgabedaten

**Was ist  
datenzentrierte KI?**

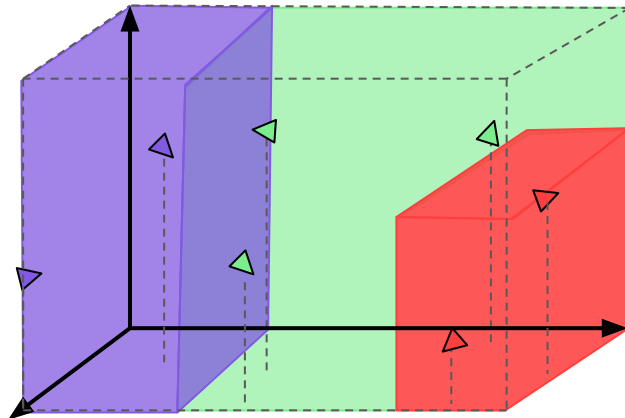
Johannes Hötter,  
Christian Warmuth

openHPI

# Ist nicht jedes Machine Learning-Modell datenzentriert?

## Machine Learning-Systeme haben i.d.R. zwei Komponenten:

- Algorithmik, d.h. Modellarchitektur zur Verarbeitung eingehender Daten sowie Lernmethodik, um etwa Modellparameter anzupassen
- Datengrundlage, welche zum Training oder zur Laufzeit eingebunden wird



- ▲ neutral
- ▲ positiv
- ▲ negativ

## Was ist datenzentrierte KI?

Johannes Hötter,  
Christian Warmuth

openHPI

# Ist nicht jedes Machine Learning-Modell datenzentriert?

## Trends:

- Komplexe Algorithmen sind in open-source Bibliotheken bereits umgesetzt
- Anbieter wie OpenAI oder Hugging Face (open-source) stellen vortrainierte Modelle bereit, welche per Finetuning einfach angepasst werden können

## Konsequenzen:

- Für die meisten Anwendungen kein Mehrwert, eigene Algorithmik zu implementieren
- Fokus lieber auf der Datensammlung und -aufbereitung
- Vorteil: Daten sind langlebig und modellunabhängig

### **Was ist datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

openHPI

# Beispiel für Sentiment Analyse



Overview **Documentation** API reference Examples Playground

## GET STARTED

Introduction

Quickstart

Libraries

Models

Tutorials

Policies

## GUIDES

Text completion

Chat completion

Image generation

Fine-tuning

**Preparing your dataset**

Advanced usage

Weights & Biases

Example notebooks

Embeddings

Speech to text

Moderation

Rate limits

Error codes

Safety best practices

## Case study: Sentiment analysis

Let's say you'd like to get a degree to which a particular tweet is positive or negative. The dataset might look something like the following:

```
{"prompt": "Overjoyed with the new iPhone! ->", "completion": " positive"}
{"prompt": "@lakers disappoint for a third straight night https://t.co/38EF43 -> ",
```

Once the model is fine-tuned, you can get back the log probabilities for the first completion token by setting `logprobs=2` on the completion request. The higher the probability for positive class, the higher the relative sentiment.

Now we can query our model by making a Completion request.

```
1 curl https://api.openai.com/v1/completions \
2 -H "Content-Type: application/json" \
3 -H "Authorization: Bearer $OPENAI_API_KEY" \
4 -d '{
5   "prompt": "https://t.co/f93xE2 Excited to share my latest blog post! ->",
6   "max_tokens": 1,
7   "model": "YOUR_FINE_TUNED_MODEL_NAME"
8 }'
```

Which will return:

```
1 {
2   "id": "cmpl-COMPLETION_ID",
3   "object": "text_completion",
4   "created": 1589498378,
```

Disclaimer: OpenAI ist hier als Beispiel genannt, es gibt noch viele weitere Optionen (siehe nächste Woche zu Open-Source)

**Was ist datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

openHPI

Folie 5

# Beispiel für Textbearbeitung



Overview Documentation API reference Examples Playground

## GET STARTED

Introduction

Quickstart

Libraries

Models

Tutorials

Policies

## GUIDES

Text completion

Introduction

Prompt design

Inserting text

Editing text

Chat completion

Image generation

Fine-tuning

Embeddings

Speech to text

Moderation

Rate limits

Error codes

Safety best practices

```
2. Lions are ferocious animals.  
3. Dolphins are playful animals.  
4. Horses are majestic animals.
```

## Editing text Alpha

The `edits` endpoint can be used to edit text, rather than just completing it. You provide some text and an instruction for how to modify it, and the `text-davinci-edit-001` model will attempt to edit it accordingly. This is a natural interface for translating, editing, and tweaking text. This is also useful for refactoring and working with code. Visit our [code guide](#) to learn more. During this initial beta period, usage of the edits endpoint is free.

### Examples

#### INPUT

```
GPT-3 is a very nice AI  
That's pretty good at writing replies  
When it's asked a question  
It gives its suggestion  
This is a poem it made that rhymes
```

#### INSTRUCTIONS

```
Make this in the voice of GPT-3
```

#### OUTPUT



```
I am a very nice AI  
I am pretty good at writing replies  
When I am asked a question  
I give my suggestion  
This is a poem I made that rhymes
```

Disclaimer: OpenAI ist hier als Beispiel genannt, es gibt noch viele weitere Optionen (siehe nächste Woche zu Open-Source)

**Was ist datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

openHPI

Was this page useful?  

# Ein simples Beispiel

Was ist die Hauptstadt von  
Frankreich?

Karteikarte **Vorderseite**

**Was ist  
datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

openHPI

# Ein simples Beispiel

Was ist die Hauptstadt von  
Frankreich?

Karteikarte **Vorderseite**

Berlin

Karteikarte **Rückseite**

**Was ist  
datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

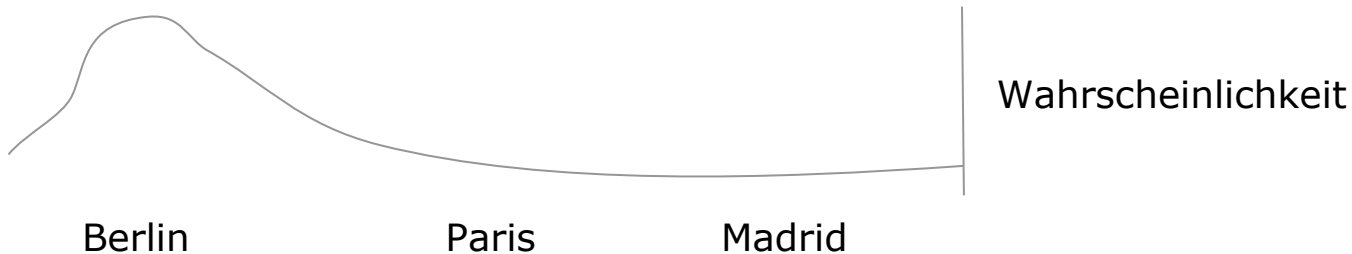
openHPI



## Recap

# Wie entstehen Embeddings?

Was ist die Hauptstadt von Frankreich? Antwort: \_\_\_\_\_.



**Was ist  
datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

openHPI

# Trainingsdaten beinhalten Fehler. Fast immer.

MNIST



given: 5  
corrected: 3

CIFAR-10



given: cat  
corrected: frog

CIFAR-100 Caltech-256

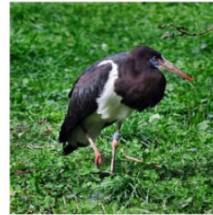


given: lobster  
corrected: crab



given: ewer  
corrected: teapot

ImageNet



given: white stork  
corrected: black stork

QuickDraw



given: tiger  
corrected: eye

Quelle: <https://github.com/cleanlab/cleanlab>

In der realen Anwendung haben Datensätze schnell  
bis zu 25% (oder sogar noch höher!) Fehlerquote

**Was ist  
datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

openHPI

# Wie kommen Fehler zustande?

Was ist die  
Hauptstadt von  
Frankreich?

Berlin



## Fehlerquellen:

- verlesen
- verklickt
- falsches Wissen
- missverstanden

**Was ist  
datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

openHPI

# Wie kommen Fehler zustande?

Was ist die  
Hauptstadt von  
Frankreich?

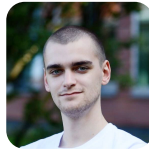
Berlin



Paris



Paris



## Fehlerquellen:

- verlesen
- verklickt
- falsches Wissen
- missverstanden

**Was ist  
datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

openHPI

# Kuratieren der Daten meistens nicht einfach

## Herausforderungen

- Zu viele Rohdaten (z.B. Tausende Texte, Bilder oder PDF-Dokumente)
- Schnell recht teuer
- Bei sensiblen Daten nur interne Aufbereitung möglich, und das bei Ressourcenmangel
- Aufgabe nicht klar genug gestellt, und somit unterschiedliche Ergebnisse je nach Person (Subjektivität)

## Was ist datenzentrierte KI?

Johannes Hötter,  
Christian Warmuth

openHPI

# Am Beispiel: OpenAssistant

**Dataset Preview** Size: 41.6 MB </> API Go to dataset viewer

Split

train (84.4k rows) ▾

text (string)	role (string)	lang (string)	review_count (int32)	review_result (bool)	deleted (bool)
"Can you write a short introduction about the relevance of the term "monopsony" in economics?..."	"prompter"	"en"	3	true	false
"Monopsony" refers to a market structure where there is only one buyer for a particular good or...	"assistant"	"en"	3	true	false
"Now explain it to a dog"	"prompter"	"en"	3	true	false
"Monopsony is a market structure in which there is a single buyer in a market. In the context of..."	"assistant"	"en"	3	true	false
"How can one fight back when a monopsony had been created?"	"prompter"	"en"	3	true	false
"Monopsony refers to a market structure where there is only one buyer of a good or service. In..."	"assistant"	"en"	3	true	false
"What can be done at a regulatory level to ensure, in a monopsony, that the power over the employees..."	"prompter"	"en"	3	true	false

< Previous 1 2 3 ... 845 Next >

**Was ist datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

openHPI

# Am Beispiel: OpenAssistant

**Dataset Preview** Size: 41.6 MB </> API Go to dataset viewer

Split

train (84.4k rows)

del_name (string)	detoxify (dict)	message_tree_id (string)	tree_state (string)	emojis (sequence)	labels (sequence)
null	{ "toxicity": 0.00044308538781479, "severe_toxicity": 0.0003252684837207198, "obscene": 0.00023475120542570949, "identity_attack": 0.0001416115992469713, "insult": 0.00039489680784754455, "threat": 0.00004075629112776369, "sexual_explicit": 0.00002712695459194947 }	"6ab24d72-0181-4594-a9cd-deaf170242fb"	"ready_for_export"	{ "name": [ "+1", "_skip_reply", "_skip_ranking" ], "count": [ 10, 1, 4 ] }	{ "name": [ "spam", "lang_mismatch", "pii", "not_appropriate", "hate_speech", "sexual_content", "quality", "toxicity", "humor", "creativity", "violence" ], "value": [ 0, 0, 0, 0, 0, 0, 0.9166666666666666, 0.16666666666666666, 0.3333333333333333, 0.6666666666666666, 0 ], "count": [ 3, 3, 3, 3, 3, 3, 3, 3, 3 ] }

< Previous 1 2 3 ... 845 Next >

**Was ist datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

openHPI

# Recap

## Chat-Trainingsdaten

Chat



Wieso ist RLHF wichtig für GPT?



GPT ist auf veralteten Daten trainiert, und RLHF erneuert das Wissen.



GPT ist nicht auf Konversationen optimiert.

Es macht einfach Spaß.

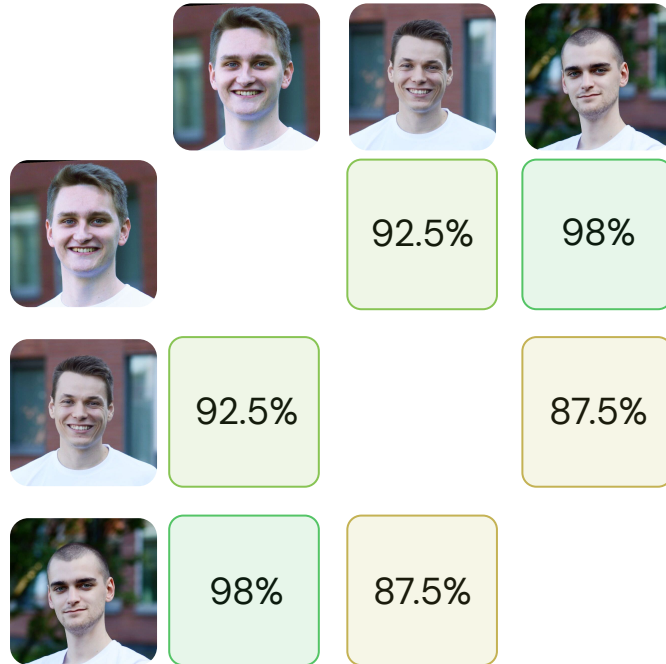
**Was ist datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

openHPI



Je besser der Datensatz, desto besser das Modell.  
Und auch das Benchmarking.



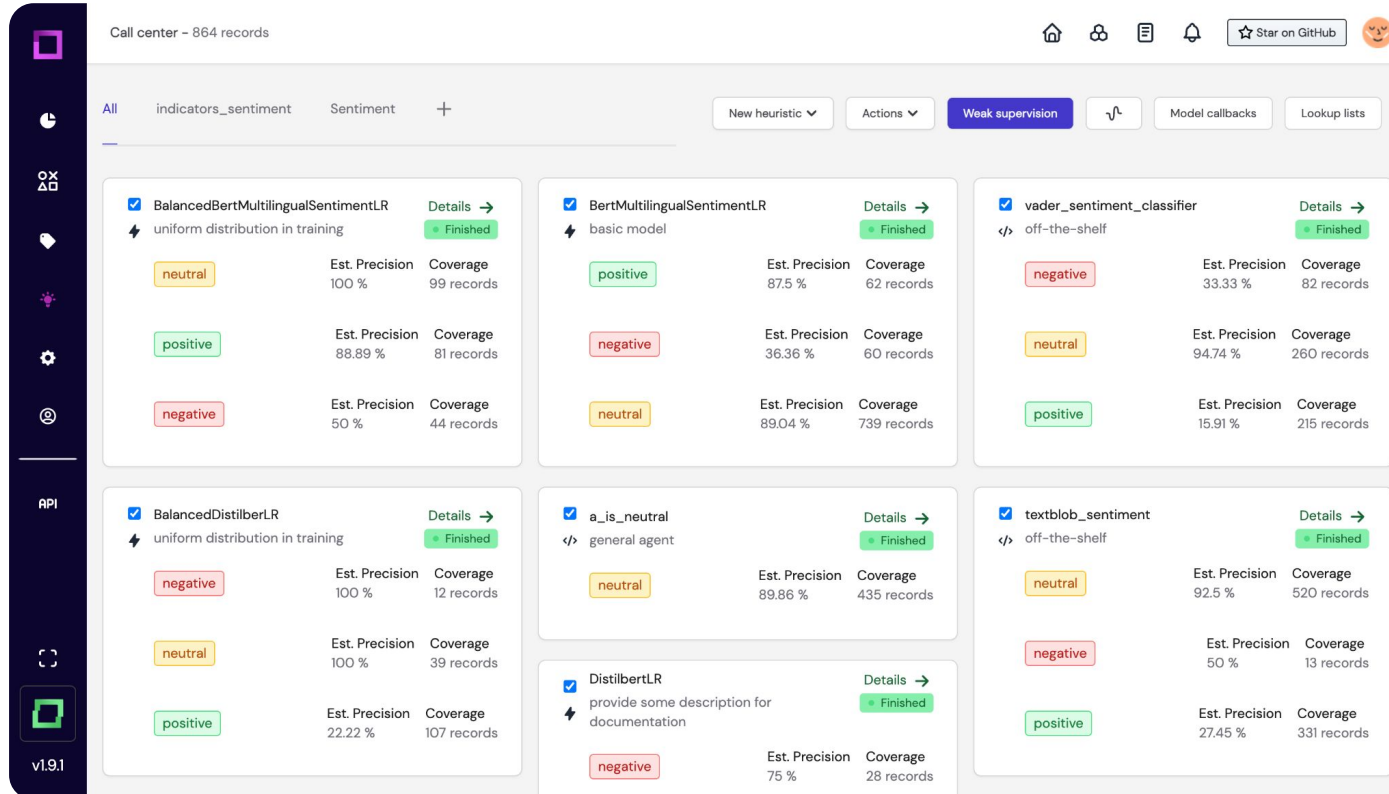
**Inter-Annotator-Agreement:**  
Wie oft stimmen sich Menschen  
bei der Aufgabe überein?

**Was ist  
datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

openHPI

# Je besser der Datensatz, desto besser das Modell. Und auch das Benchmarking.



## Was ist datenzentrierte KI?

Johannes Hötter,  
Christian Warmuth

openHPI

# Trainingsdaten als Intellectual Property

**GPT-3**  
**Dolly 2.0**  
**OpenAssistant**



**GPT-4**  
**Dolly 3.0**  
**OpenAssistant 2**

Dataset Preview

Size: 41.6 MB API Go to dataset viewer

Split

train (84.4k rows)

text (string)	role (string)	lang (string)	review_count (int32)	review_result (bool)	deleted (bool)
"Can you write a short introduction about the relevance of the term "monopsony" in economics?..."	"prompter"	"en"	3	true	false
"Monopsony refers to a market structure where there is only one buyer for a particular good or..."	"assistant"	"en"	3	true	false
"Now explain it to a dog"	"prompter"	"en"	3	true	false
"Monopsony is a market structure in which there is a single buyer in a market. In the context of..."	"assistant"	"en"	3	true	false
"How can one fight back when a monopsony had been created?"	"prompter"	"en"	3	true	false
"Monopsony refers to a market structure where there is only one buyer of a good or service. In..."	"assistant"	"en"	3	true	false
"What can be done at a regulatory level to ensure, in a monopsony, that the power over the employees..."	"prompter"	"en"	3	true	false

< Previous 1 2 3 ... 845 Next >

Dataset Preview

Size: 41.6 MB API Go to dataset viewer

Split

train (84.4k rows)

text (string)	role (string)	lang (string)	review_count (int32)	review_result (bool)	deleted (bool)
"Can you write a short introduction about the relevance of the term "monopsony" in economics?..."	"prompter"	"en"	3	true	false
"Monopsony refers to a market structure where there is only one buyer for a particular good or..."	"assistant"	"en"	3	true	false
"Now explain it to a dog"	"prompter"	"en"	3	true	false
"Monopsony is a market structure in which there is a single buyer in a market. In the context of..."	"assistant"	"en"	3	true	false
"How can one fight back when a monopsony had been created?"	"prompter"	"en"	3	true	false
"Monopsony refers to a market structure where there is only one buyer of a good or service. In..."	"assistant"	"en"	3	true	false
"What can be done at a regulatory level to ensure, in a monopsony, that the power over the employees..."	"prompter"	"en"	3	true	false

< Previous 1 2 3 ... 845 Next >

**Was ist datenzentrierte KI?**

Johannes Hötter,  
Christian Warmuth

openHPI

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

# Was ist datenzentrierte KI?

Johannes Hötter  
Christian Warmuth