Finetuning von Modellen

Johannes Hötter
Christian Warmuth

Warum ist Finetuning wichtig?

"Ich hatte am gestrigen Abend einen Unfall mit meinem Auto. [...] Können Sie mir weiterhelfen?"

Prognose vortrainiertes Modell

negativ

Prognose finetuned Modell

neutral

Finetuning

Johannes Hötter,
Christian Warmuth

openHPI

Warum ist Finetuning wichtig?

"Mit welcher Art von Finanzinstrument gewährt man zum Beispiel dem Staat einen Kredit?"

Prognose vortrainiertes Modell
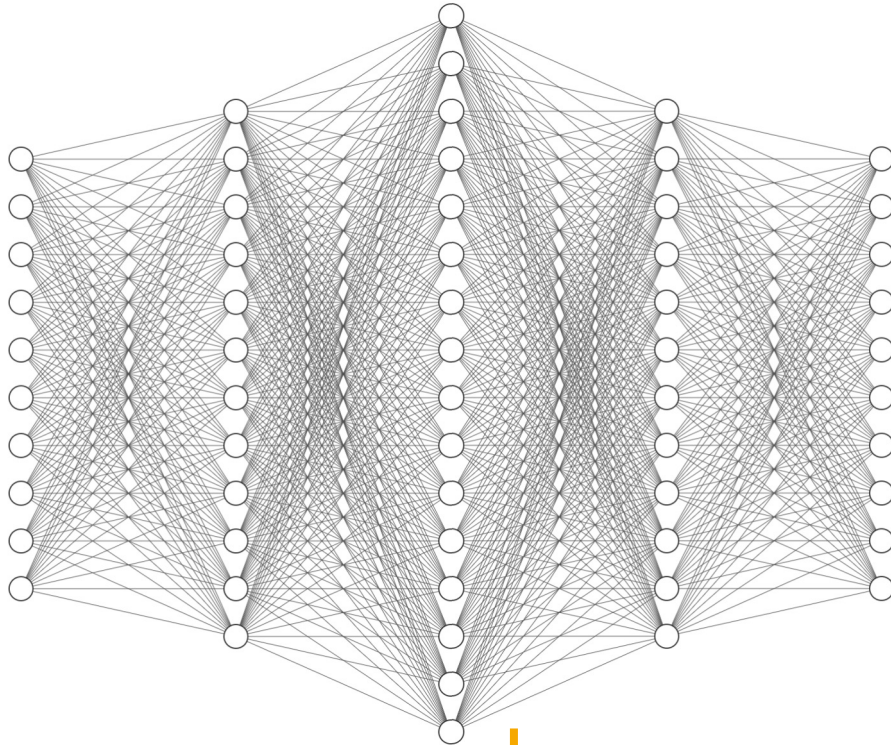
Aktie

Prognose finetuned Modell

Anleihe

Finetuning

Johannes Hötter,
Christian Warmuth

openHPI

Unterschied zum Prompting

Langfristige Veränderung des Modells

Beispiel für Klassifikationen

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

Finetuning

Johannes Hötter, Christian Warmuth

openHPI

Quelle: https://platform.openai.com/docs/guides/fine-tuning/preparing-your-dataset

Beispiel für Klassifikationen

Text	Contains PII?
My name is John Smith.	True
I live at 123 Main St.	True
You can contact me at john.smith@email.com.	True
My favorite color is blue.	False
I was born on January 1, 1990.	True
My social security number is 123-45-6789.	True
I work at Acme Corporation.	False
My phone number is (555) 123-4567.	True
My credit card number is 1234-5678-9012-3456.	True
I graduated from Harvard University in 2012.	False

Je nach Modell welches eingesetzt wird, kann das Datenformat einfach programmatisch angepasst werden.

Quelle: https://hitchhiker.kern.ai/technical-best-practices

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam

KI Service Zentrum HPI
by Hasso-Plattner-Institut

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Finetuning

Johannes Hötter, Christian Warmuth

openHPI

Beispiel für Textbearbeitung

Quelle: https://platform.openai.com/docs/guides/fine-tuning/preparing-your-dataset

Beispiel für RLHF
(Reinforcement Learning with Human Feedback)

Question	Option 1	Option 2	Option 3	Best Fit
What is the capital of France?	Paris	London	New York City	1
What is the best way to the airport?	Take the train	Drive	Walk	1
Which phone should I buy?	iPhone	That depends.	Nokia	2
What should I wear to a wedding?	A suit or formal dress	Jeans and t-shirt	Swimwear	1
How can I better my programming?	Take a nap	Watch TV all day	Practice regularly	3

Finetuning

Johannes Hötter,
Christian Warmuth

openHPI

Quelle: https://hitchhiker.kern.ai/technical-best-practices

Beispiel für Extraktionen

Overview   Documentation   API reference   Examples   Playground

GET STARTED
Introduction
Quickstart
Libraries
Models
Tutorials
Policies

GUIDES
Text completion
Chat completion
Image generation
Fine-tuning

Case study: Entity extraction

This is similar to a language transformation task. To improve the performance, it is best to either sort different extracted entities alphabetically or in the same order as they appear in the original text. This will help the model to keep track of all the entities which need to be generated in order. The dataset could look as follows:

```
{"prompt":"<any text, for example news article>\n\n###\n\n", "completion":" <
```

For example:

```
{"prompt":"Portugal will be removed from the UK's green travel list from Tues
```

A multi-line separator works best, as the text will likely contain multiple lines. Ideally there will be a high diversity of the types of input prompts (news articles, Wikipedia pages, tweets, legal documents), which reflect the likely texts which will be encountered when extracting entities.

Finetuning

Johannes Hötter,
Christian Warmuth

openHPI

Quelle: https://platform.openai.com/docs/guides/fine-tuning/preparing-your-dataset

Beispiel für Extraktionen

Case study: Entity extraction

This is similar to a language transformation task. To improve the performance, it is best to either sort different extracted entities alphabetically or in the same order as they appear in the original text. This will help the model to keep track of all the entities which need to be generated in order. The dataset could look as follows:

```
ticle>\n\n###\n\n", "completion":" <list of entities, separated by a newline>
```

For example:

```
s...\n\n###\n\n", "completion":" Portugal\nUK\nNepal mutation\nIndian variant
```

A multi-line separator works best, as the text will likely contain multiple lines. Ideally there will be a high diversity of the types of input prompts (news articles, Wikipedia pages, tweets, legal documents), which reflect the likely texts which will be encountered when extracting entities.

Navigation (left sidebar):

- Overview
- Documentation
- API reference
- Examples
- Playground

GET STARTED
- Introduction
- Quickstart
- Libraries
- Models
- Tutorials
- Policies

GUIDES
- Text completion
- Chat completion
- Image generation
- Fine-tuning

Quelle: https://platform.openai.com/docs/guides/fine-tuning/preparing-your-dataset

Finetuning

Johannes Hötter, Christian Warmuth

openHPI

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Beispiel für Extraktionen

Text	PII Type	Start Position	End Position
My name is John Smith and my email is john.smith@email.com.	Name, Email	11	48
I live at 123 Main St, Anytown, USA.	Address	10	27
My phone number is (555) 123-4567.	Phone Number	19	32
My social security number is 123-45-6789.	Social Security Number	27	39
My credit card number is 1234-5678-9012-3456.	Credit Card Number	24	43
I was born on January 1, 1990.	Date of Birth	14	28
My passport number is ABC123456.	Passport Number	22	30
My driver's license number is A123-456-789-0.	Driver's License Number	26	43
I work at Acme Corporation.	None	-	-
My favourite colour is blue.	None	-	-

Je nach Modell welches
eingesetzt wird, kann
das Datenformat einfach
programmatisch
angepasst werden.

Finetuning

Johannes Hötter,
Christian Warmuth

openHPI

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

Johannes Hötter
Christian Warmuth