



# Kenne deine Daten: Wie frei verfügbare KI-Modelle bei der Analyse von großen Datenmengen die Erschließung unterstützen können

Martin Vogel,  
Niedersächsisches Landesarchiv

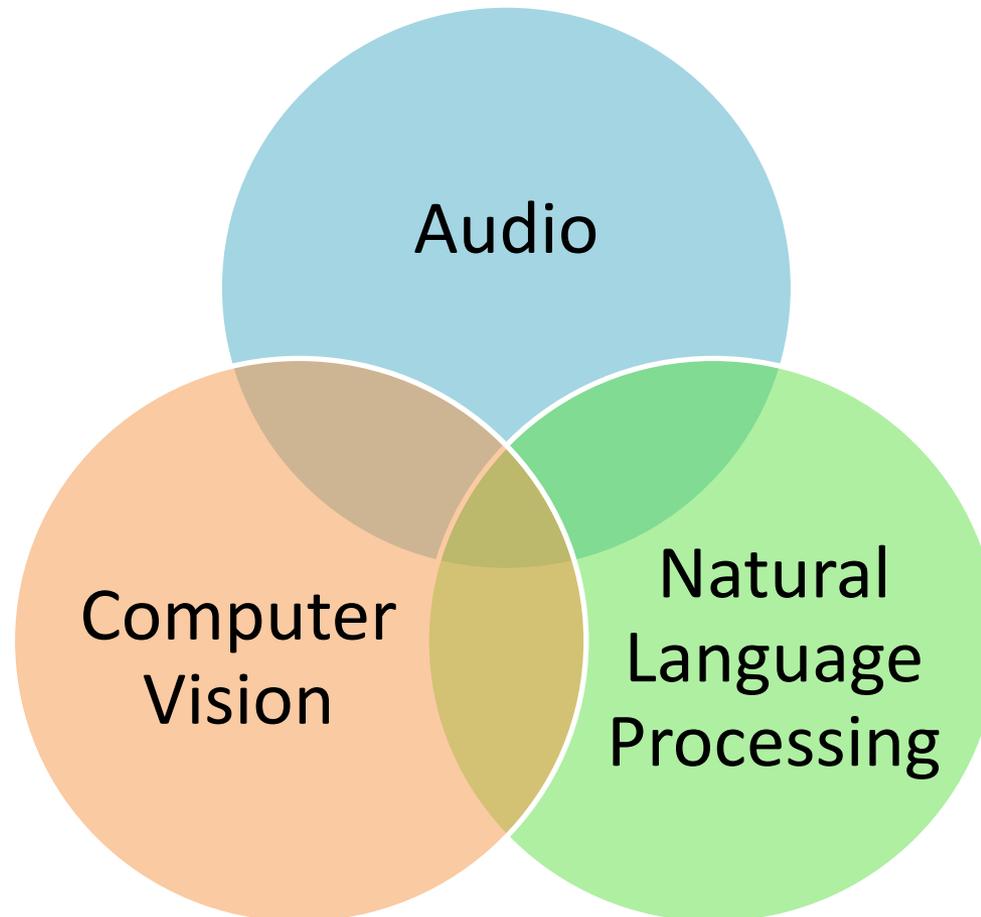


# Gliederung

- Übersicht der Modelle
- Beispiele
  - Segmentierung
  - Transkription
- Wie kann man Modelle nutzen?
- Zusammenfassung / Ausblick



# Übersicht der Modelle





# Beispiel: Segmentierung



Detected car with confidence 0.999 at location [338.93, 751.17, 546.29, 900.87]

Modell (Computer-Vision): [facebook/detr-resnet-50](https://github.com/facebookresearch/detr)



# Beispiel: Segmentierung



building	1.000
sky	1.000
tree	1.000
road	1.000
sidewalk	1.000
car	1.000
pole	1.000

Modell (Computer-Vision): [nvidia/segformer-b0-finetuned-ade-512-512](https://huggingface.co/nvidia/segformer-b0-finetuned-ade-512-512)



# Beispiel: Segmentierung



building	1.000
sky	1.000
tree	1.000
road	1.000
sidewalk	1.000
person	1.000
plant	1.000
fence	1.000
minibike	1.000
bicycle	1.000



# Beispiel: Transkription

Input: Warum Umweltschutz? (wav-Datei)

Output: 1466 Wörter

Im Jahre 1970 eröffnete das Senckenberg-Museum in Frankfurt am Main eine Umweltausstellung und brachte gleichzeitig eine Broschüre mit dem Titel Umwelt 2000 heraus. Eigentlich erst seit diesem Jahre wurde bei uns in der Bundesrepublik der Umweltschutz aktueller. Auf dem Büchermarkt in Mitteleuropa erschienen die verschiedensten Titel. Schutz unseres Lebensraumes, Vorträge an der Eidgenössischen Hochschule in Zürich im Jahre 1970, Belastete Landschaft, Gefährdete Umwelt, herausgegeben von Olschowi 1971, Die Erde hat keinen Notausgang von Dahmen 1971, Umwelt heute von Schwabe 1973. Und schließlich der ausgezeichnete Ökologieband in Jimmocks Tierleben mit einem Artikel. Abschnitt über die Umwelt des Menschen im Jahre 1973. In diesem Zusammenhang dürfen nicht unerwähnt bleiben...

Im Jahre 1970 eröffnete das **Senckenberg-Museum** **LOC** in **Frankfurt am Main** **LOC** eine Umweltausstellung und brachte gleichzeitig eine Broschüre mit dem Titel Umwelt 2000 heraus. Eigentlich erst seit diesem Jahre wurde bei uns in der **Bundesrepublik** **LOC** der Umweltschutz aktueller. Auf dem Büchermarkt in Mitteleuropa erschienen die verschiedensten Titel. Schutz unseres Lebensraumes, Vorträge an der **Eidgenössischen Hochschule** **ORG** in **Zürich** **LOC** im Jahre 1970, Belastete Landschaft, Gefährdete Umwelt, herausgegeben von **Olschowi** **LOC** 1971, **Die Erde** **LOC** hat keinen Notausgang von **Dahmen** **LOC** 1971, Umwelt heute von **Schwabe** **PER** 1973. Und schließlich der ausgezeichnete Ökologieband in **Jimmocks** **PER** Tierleben mit einem Artikel. Abschnitt über die Umwelt des Menschen im Jahre 1973. In diesem Zusammenhang dürfen nicht unerwähnt bleiben die

Modell (Audio): OpenAI / Whisper (large)

Modell (Natural Language Processing): Spacy / de\_core\_news\_lg



# Beispiel: Transkription

## Input: Warum Umweltschutz? 1466 Wörter

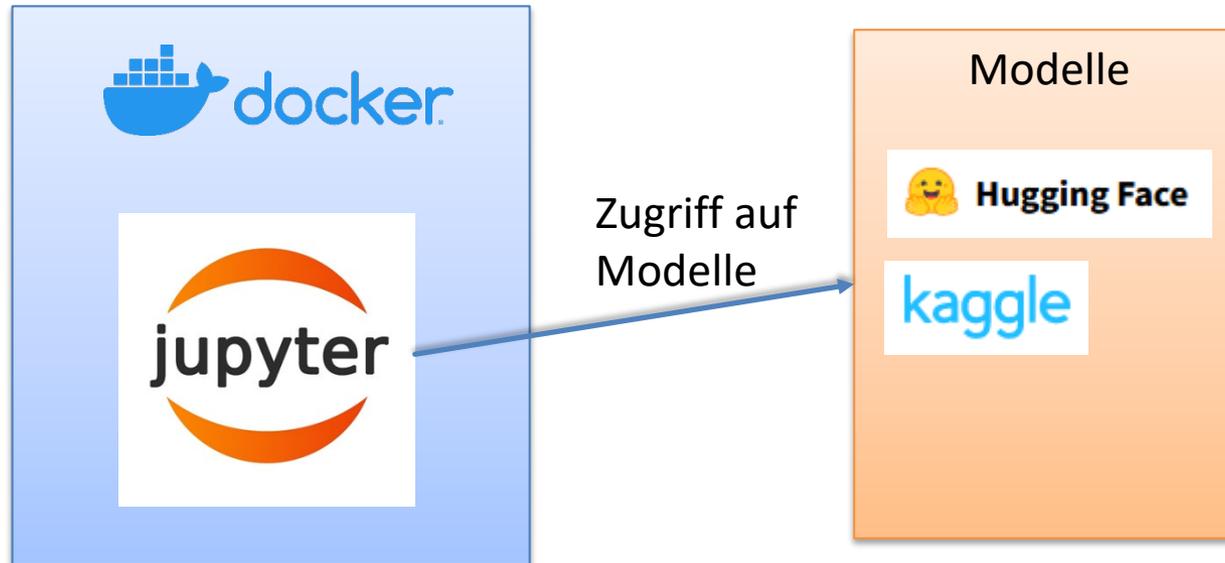
Im Jahre 1970 eröffnete das Senckenberg-Museum in Frankfurt am Main eine Umweltausstellung und brachte gleichzeitig eine Broschüre mit dem Titel Umwelt 2000 heraus. Eigentlich erst seit diesem Jahre wurde bei uns in der Bundesrepublik der Umweltschutz aktueller. Auf dem Büchermarkt in Mitteleuropa erschienen die verschiedensten Titel. Schutz unseres Lebensraumes, Vorträge an der Eidgenössischen Hochschule in Zürich im Jahre 1970, Belastete Landschaft, Gefährdete Umwelt, herausgegeben von Olschowi 1971, Die Erde hat keinen Notausgang von Dahmen 1971, Umwelt heute von Schwabe 1973. Und schließlich der ausgezeichnete bebilderte Ökologieband in Jimmocks Tierleben mit einem Artikel. Abschnitt über die Umwelt des Menschen im Jahre 1973. In diesem Zusammenhang dürfen nicht unerwähnt bleiben...

## Zusammenfassung:

Im Jahre 1970 eröffnete das Senckenberg-Museum in Frankfurt am Main eine Umweltausstellung und brachte gleichzeitig eine Broschüre mit dem Titel Umwelt 2000 heraus. In diesem Zusammenhang dürfen nicht unerwähnt bleiben die Materialien zum Umweltprogramm der Bundesregierung 1971 und die Berichte des Club of Rome, die Grenzen des Wachstums 1972 und Menschheit am Wendepunkt 1974. Heute hat fast jede Tageszeitung dann und wann eine Seite mit der Überschrift Umweltschutz.

Modell (Natural Language Processing): Falconsai/text\_summarization

# Wie kann man Modelle nutzen?



- Docker Desktop für Windows
- Programm: Jupyter-Notebook
- Einfache Python-Kenntnisse
- Hugging-Face
- Kaggle
- Handelsüblicher PC/Laptop

```
[2]: from datetime import datetime
import whisper
print("Start", datetime.now())
base_model_medium = whisper.load_model("large")
whisper.DecodingOptions(fp16 = False, language = 'de')
result_medium = base_model_medium.transcribe("/home/jovyan/work/audio/
print("Ende", datetime.now())
print(result_medium["text"])
```



# Zusammenfassung / Ausblick

- KI-Modelle können heute mit wenig Aufwand genutzt werden
- Die Qualität der Modelle wird stetig verbessert
- Modelle aus den unterschiedlichen Bereichen dienen der Analyse der Daten und können bei großen Datenmengen einen guten Überblick verschaffen
- Diese Analysedaten können für die Erschließung genutzt werden