

# Grosse Sprachmodelle

Einführung und Anwendungsmöglichkeiten im öffentlichen Sektor

Prof. Dr. Marcel Gygli



# Übersicht

Grosse Sprachmodelle

Wort-Embeddings

Konkreter Einsatz in der Nationalbibliothek

Weitere Einsatzmöglichkeiten von Sprachmodellen

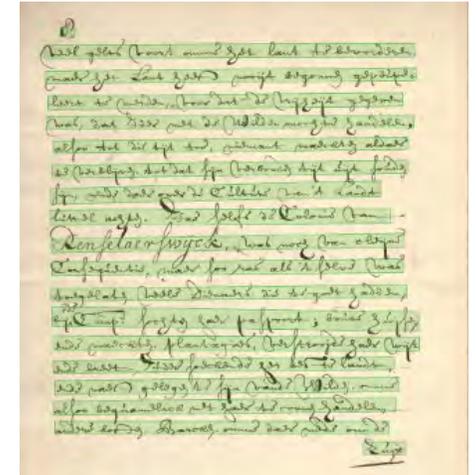
Mögliche Hürden

Wie anfangen?



# Kurz zu mir

- Innovation Fellow der Schweizerischen Nationalbibliothek (bis 31.07)
- Professor (Tenure Track) an der BFH (seit 01.05)
  - AI In the Public Sector
- PhD in Computer Science (2015 – 2020)
  - Arbeit mit historischen Dokumenten
- Wissenschaftlicher Mitarbeiter FHNW (2018 – 2022)
  - Human Centered Design / Design Thinking
  - Digitalisierung / Industrie 4.0 / Digital Humanities





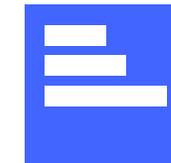
# Innovation Fellowship

- Pilotprogramm organisiert von der EPA
  - In anderen Ländern bereits etabliert
- Allgemeine Ziele
  - Innovation in der Bundesverwaltung fördern
  - Nachhaltiger Wissenstransfer
- Ziele in der NB
  - Potential von der KI für die NB ausloten
- Wird 2024 weitergeführt!
  - [simona.ingold@epa.admin.ch](mailto:simona.ingold@epa.admin.ch)



Presidential ★★  
Innovation Fellows

**No.10**  
Innovation  
Fellowships



**Tech  
4Germany**



# Was sind grosse Sprachmodelle?



Was ist ein grosses Sprachmodell?



Ein großes Sprachmodell ist eine künstliche Intelligenz, die mithilfe von Texten und großen Datenmengen "gelernt" hat bestimmte Antworten zu geben. Daher ist es wichtig in welchem Kontext diese Antwort verstanden wird, da man mit der gewählten Eingabe auch das Problem verändern kann. Das größte bekannte Sprachmodell (Stand 2021) heißt GPT-3 (Generative Pre-training Transformer 3). Es wurde vom Unternehmen OpenAI entwickelt.

[<https://open-assistant.io>]

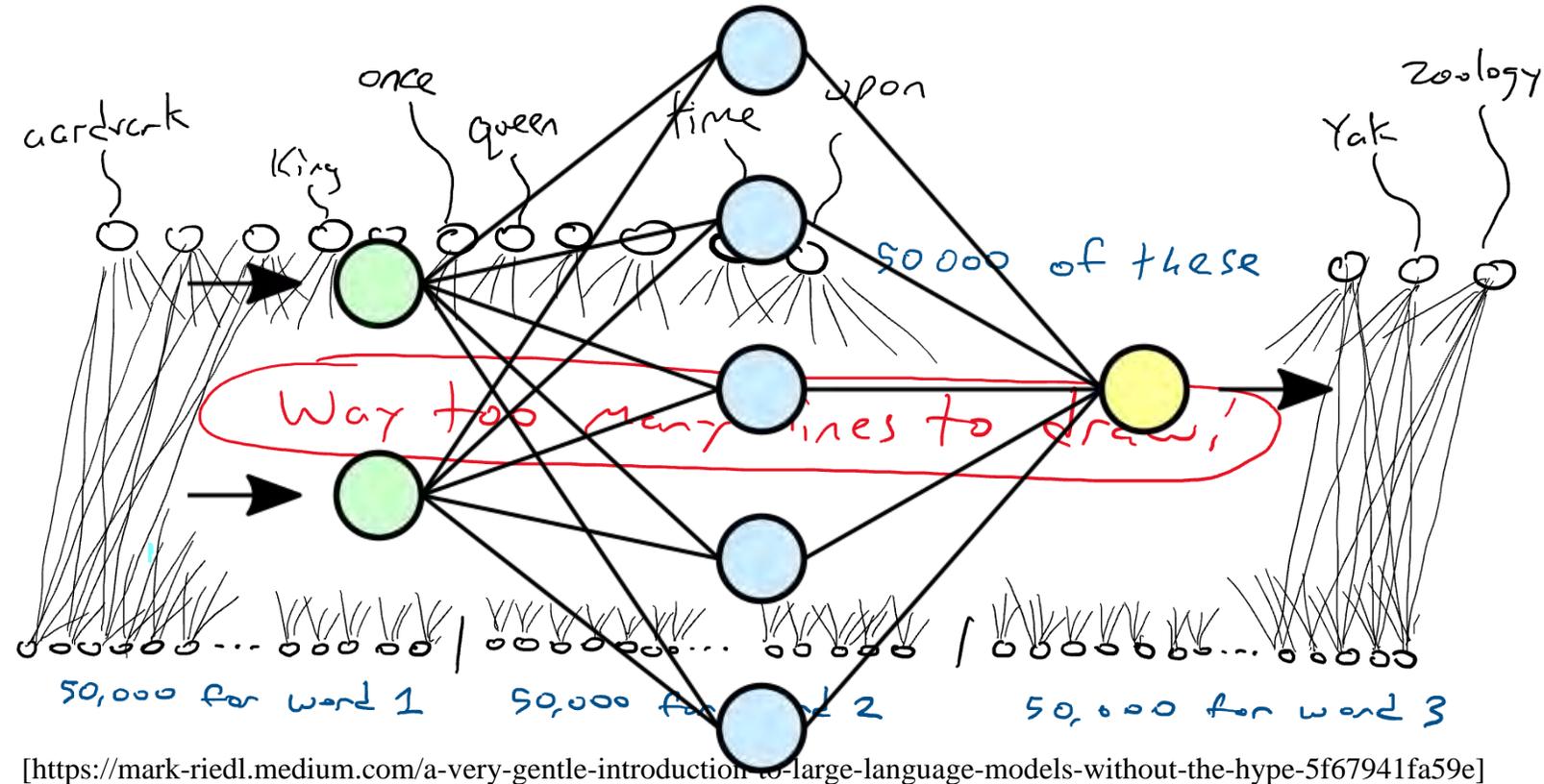


# Was machen Grosse-Sprachmodelle?

Once upon a ...

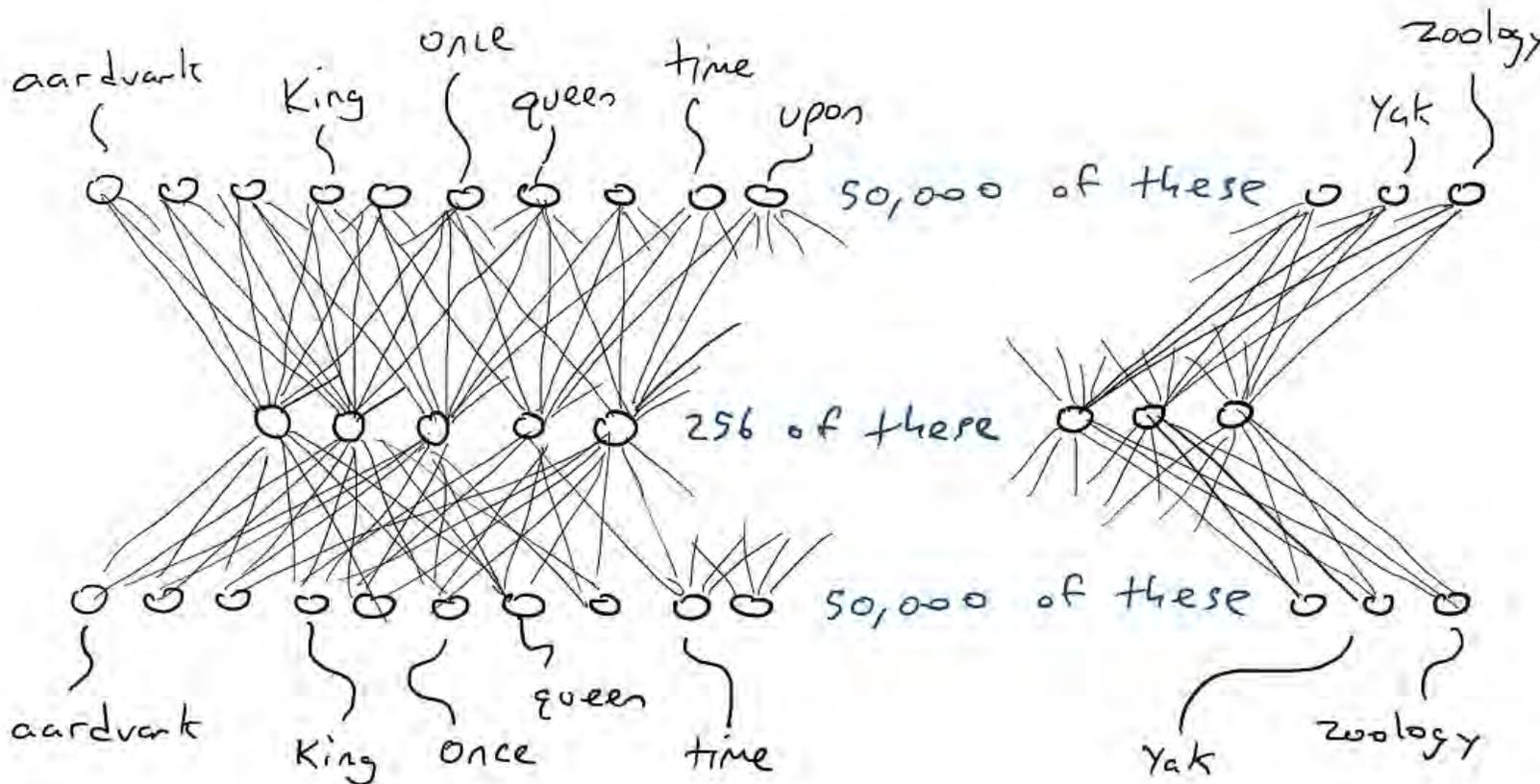


# Als Neutrales Netzwerk nicht umsetzbar



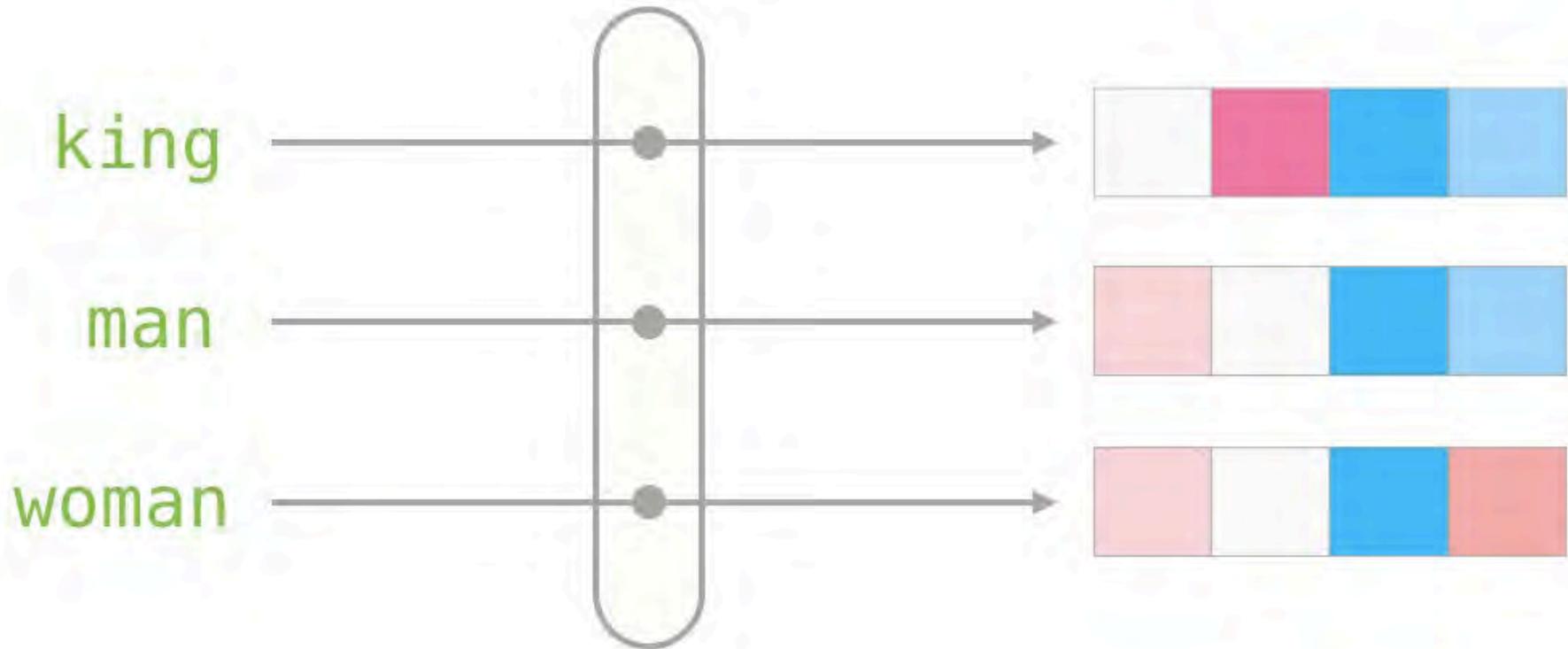


# Wir benötigen eine Zwischenrepräsentation





# Word-Embeddings als Repräsentation

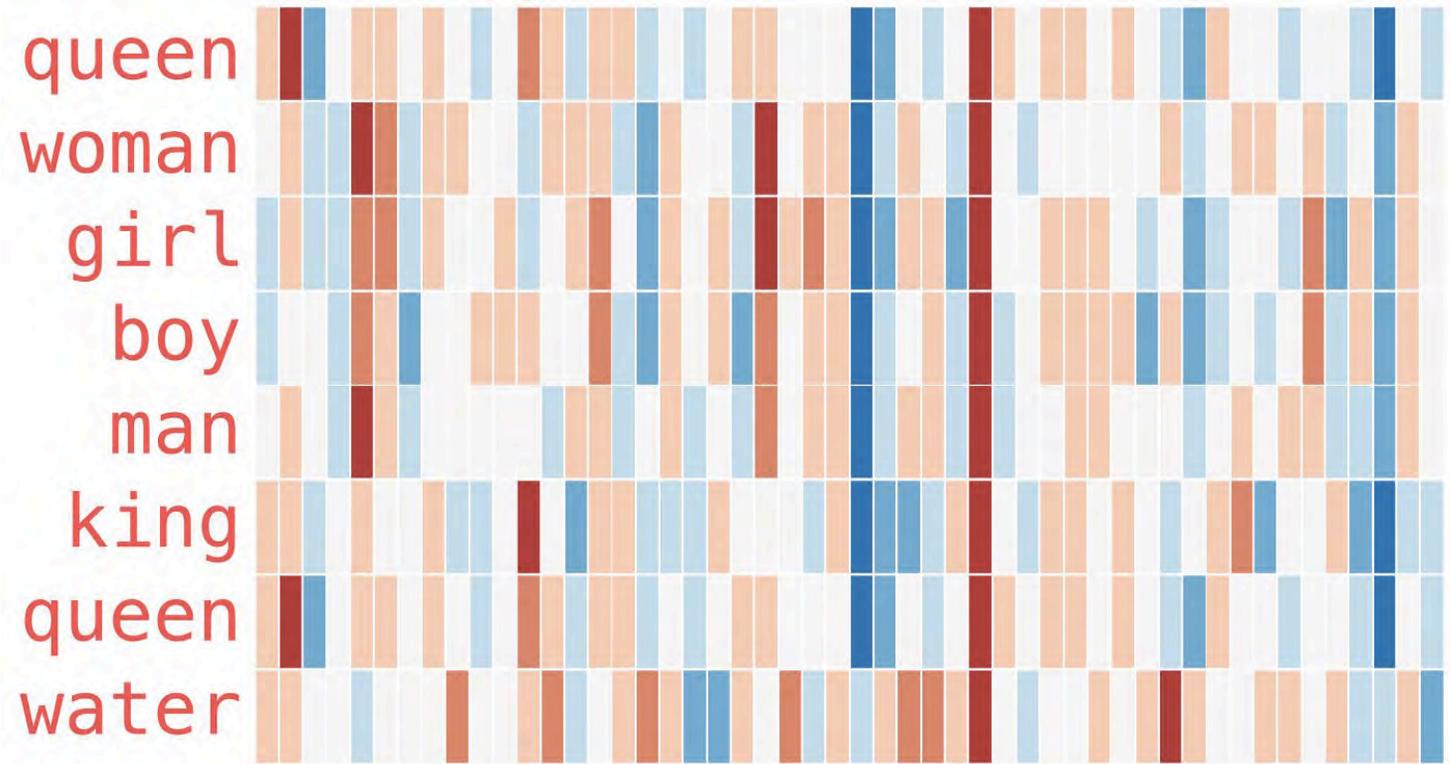




# Word-Embeddings im Details

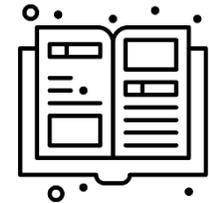
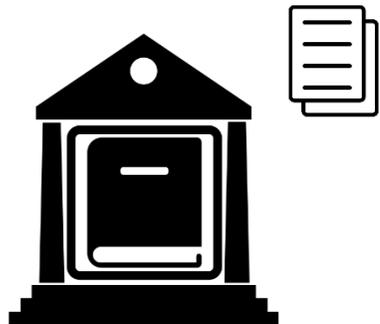
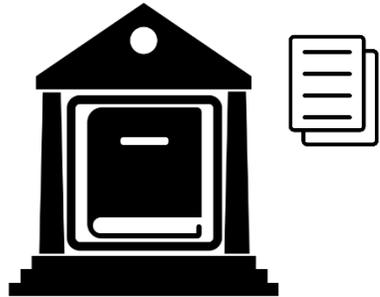
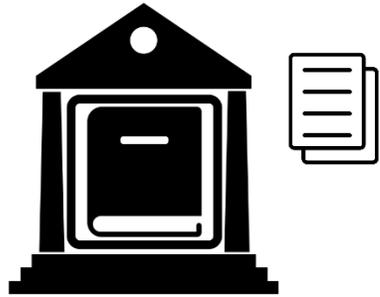
King

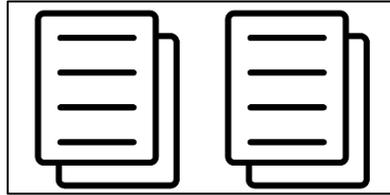
[ 0.50451 , 0.68607 , -0.59517  
, -0.022801, ... , -1.6106 , -  
0.64426 , -0.51042 ]





Berner  
Fachhochschule



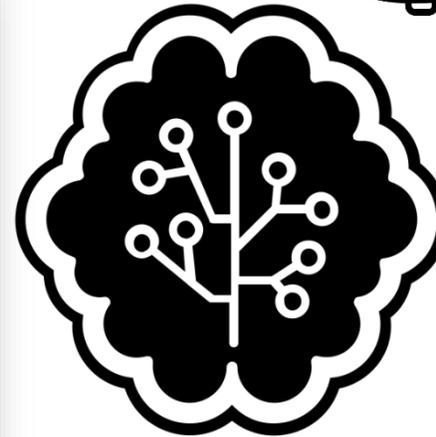
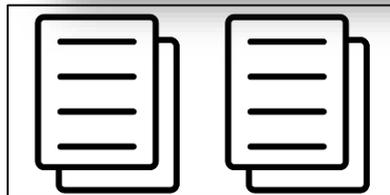


**R**eproducible Research (RR) has been a research topic within the computational sciences for a long time. It deals with the problem of ensuring that other researchers can independently verify results generated on a computer.

A longstanding problem that reproducible research tries to address is that many published scientific results can not be verified by reading the publication itself. This is due to a set of various reasons: not enough space in the publication to explain all implementation details or aversion to sharing source code publicly. Thus, it becomes ever more difficult for researchers to reproduce results from a presented method, compute results on new data, or use it as part of a new system. In this thesis, we provide an overview of these issues with a particular focus on the area of Document Image Analysis (DIA), where these problems arise as well. We analyze approaches proposed as solutions to these problems, and identify open issues, especially in the area of providing access to executable versions of research code. Using this analysis, we design the features of a system that will be able to support DIA researchers in performing reproducible research. This design is implemented in a Web Service framework called DIVAServices that allows for the publication and execution of research methods through a unified interface.

The contribution of this thesis is threefold: First, we analyze how Web Services can be used in the domain of reproducible research and what benefits they offer. This is brought into production by the development of the DIVAServices framework as an implementation of such a concept. Second, we design a concept for bringing workflows into the frameworks as they are an essential extension. We show that this concept works by developing a proof-of-concept for the execution of linear workflows. Lastly, we demonstrate the need for an ecosystem of tools and services around such a framework. An initial ecosystem is being developed within this work and we introduce different applications.

As a commitment to reproducible research, all code generated within this thesis is available under an open-source license.



## Computer Science



# Weitere Einsatzmöglichkeiten

LU

Ich komme aus Deutschland und wohne seit 5 Jahren in der Schweiz. Ich bin seit 2 Jahren mit einer Schweizerin verheiratet. Was sind die Bedingungen für meine Einbürgerung?



Als deutscher Staatsangehöriger können Sie sich nach fünf Jahren rechtmäßigen und ununterbrochenen Aufenthalts in der Schweiz um die Einbürgerung bewerben. Die Ehe mit einer Schweizer Staatsangehörigen kann eine positive Rolle bei der Entscheidung über Ihre Einbürgerung spielen.

LU

Was sind die gesetzlichen Grundlagen im Kanton Bern?

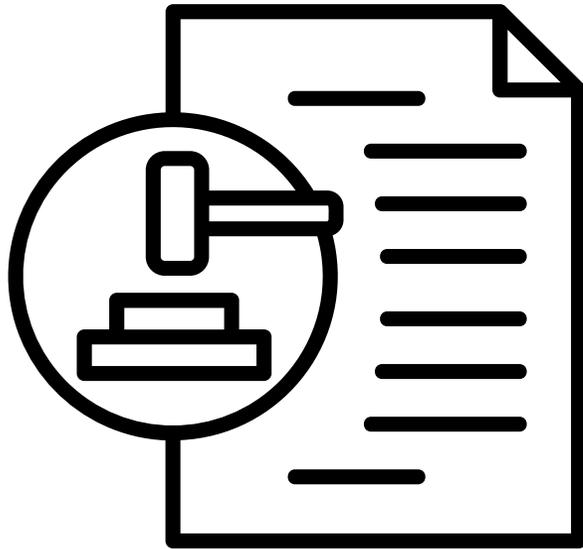


Die gesetzlichen Grundlagen für die Einbürgerung im Kanton Bern finden sich in verschiedenen kantonalen Gesetzen und Verordnungen, darunter insbesondere im Gesetz über das Bürgerrecht und die politischen Rechte (BüG) sowie in der Verordnung über das Verfahren zur Einbürgerung von Ausländerinnen und Ausländern (VEA).

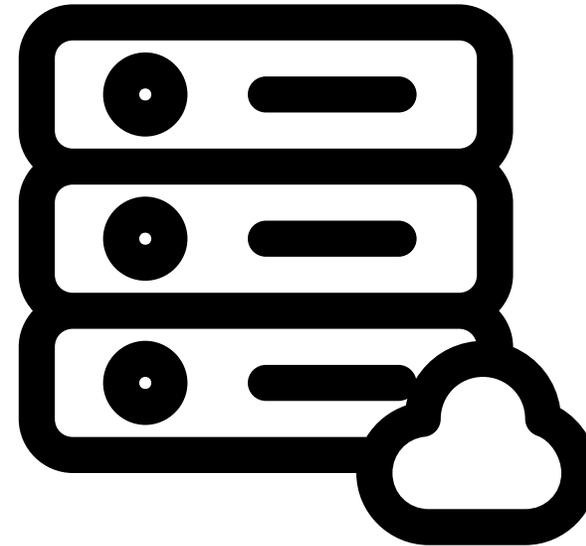
Beide Quellen existieren nicht!



# Hürden im Öffentlichen Sektor



Rechtliche Grundlagen  
(Datenschutz / Urheberrecht)



Cloud Ressourcen  
On Premise Hosting



# Aller Anfang ist «leicht»



## Hugging Face



```
from transformers import pipeline
image_to_text = pipeline("image-to-text", model="nlpconnect/vit-gpt2-image-captioning")
image_to_text("https://ankur3107.github.io/assets/images/image-captioning-example.png")

# [{'generated_text': 'a soccer game with a player jumping to catch the ball '}]
```



Computation time on cpu: cached

a soccer player kicking a soccer ball



# Weiterführende Links

- A Very Gentle Introduction to Large Language Models without the Hype  
Prof. Mark Riedl
- Innovation Fellowship: [simona.ingold@epa.admin.ch](mailto:simona.ingold@epa.admin.ch)
- Persönlich: [marcel.gygli@bfh.ch](mailto:marcel.gygli@bfh.ch)

