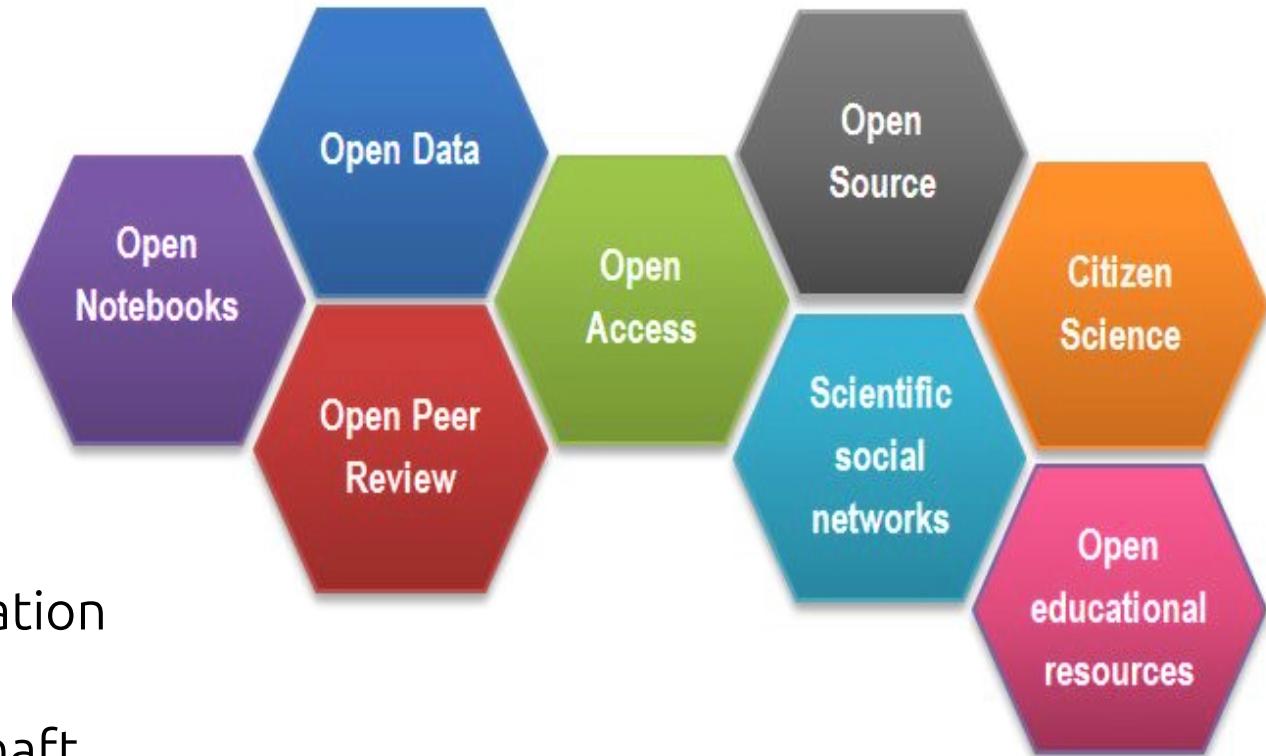


KI – demokratisch und frei?

Peter Sandrini
Universität Innsbruck

Künstliche Intelligenz und Translation
Vortragsreihe SS 2024
Institut für Translationswissenschaft

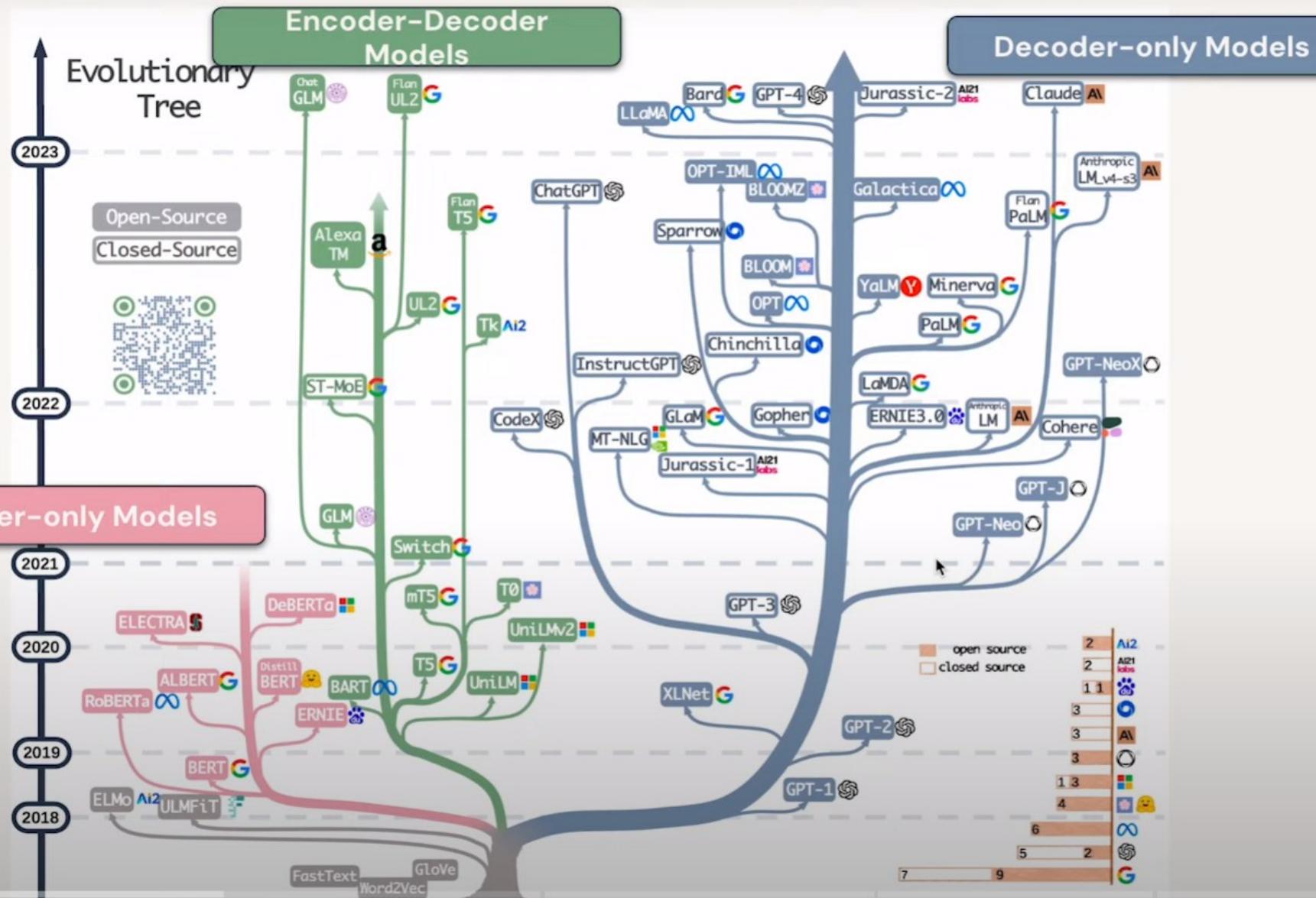


Übersicht

- 1) Warum ist Offenheit wichtig?
- 2) Vor- und Nachteile freier KI?
- 3) Welche freien lokal installierbaren KI-Systeme gibt es?
- 4) Wofür kann ich freie lokale KI anwenden?
- 5) Wo liegen die Grenzen?



The Transformer Family Tree



Foundation models – Basismodelle

(German Mistral)

1. Definition von Basismodellen (LLM):

In einem Sprachmodell, insbesondere einem großen Sprachmodell (LLM), bezieht sich das Basismodell auf die grundlegende Architektur und Parameter der Modellierung. Es handelt sich um ein vorab trainiertes neuronales Netzwerk mit einer bestimmten Anzahl von Parametern, das durch maschinelles Lernen oder Transfer Learning generiert wurde.

2. Erklärung des Konzepts:

Ein Basismodell ist eine grundlegende Struktur für Sprachmodelle und wird oft als "Basis" verwendet, um andere Modelle zu erweitern oder anzupassen. Das Basismodell enthält die Kernkomponenten wie Eingabe- und Ausgabeschichten sowie Verstecksschichten und Parameter. Durch das Training des Basismodells auf einer großen Menge an Textdaten wird eine grundlegende Repräsentation der Sprache erlernt, was es dem Modell ermöglicht, verschiedene Aufgaben im Zusammenhang mit der natürlichen Sprachverarbeitung (NLP) zu lösen.

Foundation models – Basismodelle

(*Llama 3 instruct*)

In the context of Large Language Models (LLMs), a "foundational model" refers specifically to a type of pre-trained language model that serves as a foundation or starting point for other models. Foundational models are designed to capture general knowledge, common sense, and linguistic patterns in natural language processing tasks.

The key characteristics of foundational LLMs include:

1. Large scale: They typically have millions or billions of parameters.
2. General-purpose training data: They're trained on large datasets that cover a wide range of topics, styles, and genres (e.g., books, articles, conversations).
3. Robustness to task-specific fine-tuning: Foundational models are designed to be robust to different downstream tasks or applications, allowing them to adapt quickly without overfitting.
4. Ability to learn from diverse data sources: They can incorporate knowledge and patterns learned from various text sources (e.g., books, articles, social media).

Filter Tasks by name

Multimodal

Visual Question Answering

Document Question Answering

Computer Vision

Image-to-Text Image Classification

Text-to-Image

Natural Language Processing

Text Generation Token Classification

Translation Text2Text Generation

Text Classification Fill-Mask

Sentence Similarity Feature Extraction

Zero-Shot Classification Question Answering

Summarization

Audio

Automatic Speech Recognition Text-to-Speech

Audio Classification Audio-to-Audio

Text-to-Audio

Reinforcement Learning

Reinforcement Learning

Robotics

HuggingFace.com

Repository

Offenheit



Geschlossen / proprietär	Beschränkter Zugang	Quelle Verfügbar	Open Source	KI-System
Kein Zugang	angestrebte Offenheit		Volle Offenheit	Daten (Trainings- und Testdaten) Aussagekräftige Beschreibung der Merkmale des Datensatzes (Herkunft, Struktur, Verarbeitungsschritte), z.B. Datenkarten, Datenblätter
	Teilmengen / Synthetische Daten / Gehosteter Zugriff	(ethische) zugangsbeschränkte Lizenz (z.B. OpenRAIL-D)	CC- oder Open-Data-Lizenz (z.B. CC-BY-4.0)	
	Keine Quellen Verfügbar	Ethische Offenheit (Ethische) nutzungsbeschränkende Lizenz (z.B. OpenRAIL-M)	Volle Offenheit OSI-genehmigte Lizenz (z. B. MIT)	Modell (Modellgewichte; Modellarchitektur; idealerweise; Hyperparameter etc.) Aussagekräftige Dokumentation, z.B. Modellkarten
	Keine offene Lizenz		Volle Offenheit OSI-genehmigte Lizenz (z. B. MIT)	Code (Trainings- und Inferenzcode) Standard- Softwaredokumentation

Vorteile durch offene Basismodelle

- Reduction in procurement costs.
- Easy identification of bugs and security issues.
- Making it easier to impart knowledge and educate people.
- Promotes innovation while also making way for reproducible research.
- Ensuring the accountability of systems through audits and the transparency of systems.

Mozilla (2024): Towards a Framework for Openness in Foundation Models

https://assets.mofoprod.net/network/documents/Towards_a_Framework_for_Openness_in_Foundation_Models.pdf?ref=news.itsfoss.com

Research, Innovation and Development	Safety and Security	Equity, Access and Usability	Broader Societal Aspects
+ Open Source Advances Research	+ Open Source Enables Technological Innovation for Safety	+ Open Source Increases Model Usability and Accessibility	+ Open Source Improves Public Trustworthiness Through Transparency
+ Open Source Could be More Affordable	- Open Models Can Also be Made to Generate Unsafe Content	+ Open Models Could Help Tackle Global Economic Inequalities	+ Open Models Can Help Reduce Copyright Disputes
+ Open Models are More Flexible and Customizable	- Open Models Cannot be Rolled Back or Forced to Update	+ Open Models Can Serve the Needs and Preferences of Diverse Communities	+ Open Models Can Drive Sustainability in GenAI Development
+ Open Source Empowers Developers and Fosters Innovation		+ Open Source Helps Democratize AI Development	

Fig. 5: Near to Mid-term Impacts of Open-Source Models: specific impacts of open-source Gen AI models categorized by area of impact and whether they are positive (+) or negative (-).

Coalition paper

Supporting open source and open science in the EU AI Act



GitHub



Hugging Face



<https://sites.google.com/view/open-foundation-models>

Falcon-180B

- The Technology Innovation Institute (TII) in the United Arab Emirates (UAE) launched an open LLM, which performs close to its proprietary competitors.
- Parameters: 180 billion
- License: Falcon-180B TII License (based on Apache 2.0)
- Usage: Allowed for commercial use and should be fine-tuned for specific tasks

OpenLLaMa

- OpenLLaMA is a reproduction of Meta's popular LLaMa model trained with a different dataset to make it available under a permissive license, when compared to Meta's custom commercial license.

Parameters: 7 billion

- License: Apache 2.0
- Usage: Allowed for research and commercial use

Mistral 7B

- Mistral 7B is a powerful language model that claims to outperform LLaMa 1/2 in various ways. The large language model is capable of math, reasoning, word knowledge, QA, and more.

Parameters: 7 billion

- License: Apache 2.0
- Usage: Allowed for research and commercial use

Wie wird Offenheit für LLMs bewertet?

Score Level	Description	Model (weights)	Pre- training Dataset	Fine- tuning Dataset	Reward model	Data Processing Code
0	Closed	No access to any public information, data or asset				
1	Published research only	Research papers(s) published but with no more information, data or asset				
2	Restricted access	Access to asset is possible only with special agreement (commercial, research...)				
3	Open with limitations	Access and reuse of asset is possible but with certain limitations on usage (ex. Open RAIL)				
4	Totally open	Access and reuse of asset is possible without restriction (ex. open source license)				

OpenAI



OpenAI
GPT

GPT-1 & 2

Non/For-profit (US)		
Component	Score	Level description
Model	4	Totally open
Dataset	1	Published research only
Code	1	Published research only

GPT-3 & 4



0	Closed
---	--------

ChatGPT
research paper only



No training of other commercial LLMs

BERT

PaLM 2 & Gemini

Enterprise (US)		
Component	Score	Level description
Model	4	Totally open
Dataset	2	Restricted access
Code	4	Totally open



1	Published research only
1	Published research only
0	Closed

Collaborative foundational LLMs



	Non-profit (US)	Research (UAE)	Research (EU)	Research (US)	Enterprise (FR)					
	EleutherAI GPT-J	Falcon	BLOOM	OpenLLaMa	Mistral/Mixtral					
Model	4	Access and reuse without restriction	3	Open with limitations	3	Open RAIL license	4	Access and reuse without restriction	4	Access and reuse without restriction
Dataset	3	Open with limitations	4	Access and reuse without restriction	3	Open with limitations	4	Access and reuse without restriction	0	No public information or access
Code	4	Completely open	1	General instructions	4	Completely open	1	Just examples	4	Completely open

Wie bezieht sich Freie Software auf Bildung?

Bildungseinrichtungen aller Bereiche sollten Freie Software nutzen und lehren, weil es die einzige Software ist, die ihnen ihre maßgeblichen Aufgaben ermöglicht zu erledigen: Menschliches Wissen zu vermitteln und Schülerinnen und Schüler darauf vorbereiten, gute Mitglieder ihrer Gesellschaft zu sein. Der Quellcode und die Methoden für Freie Software sind Teil menschlichen Wissens. Im Gegensatz dazu ist proprietäre Software geheimes, beschränktes Wissen, was das Gegenteil der Aufgabe von Bildungseinrichtungen ist.

Freie Software fördert Bildung, proprietäre Software verbietet Bildung.

Freie Software ist nicht nur eine technische, sondern eine ethische, soziale und politische Frage. Es ist eine Frage der Menschenrechte, die Softwarenutzer haben sollten. Freiheit und Zusammenarbeit sind unerlässliche Werte von Freier Software, da gemeinsame Nutzung gut und nützlich für den menschlichen Fortschritt ist.

<https://www.gnu.org/education/education.html>

Public Money? Public Code!



<https://fsfe.org/activities/activities.de.html>

<https://media.fsfe.org/w/7cH6EATpmPwaMsEZN8vfj>

<https://publiccode.eu/en/>

Explainable AI

training data analysis
causal intervention modeling
production observability

Accessible AI

open docs
open code
open weights
open data

low resource AI
private AI

Nachhaltigkeit von Software

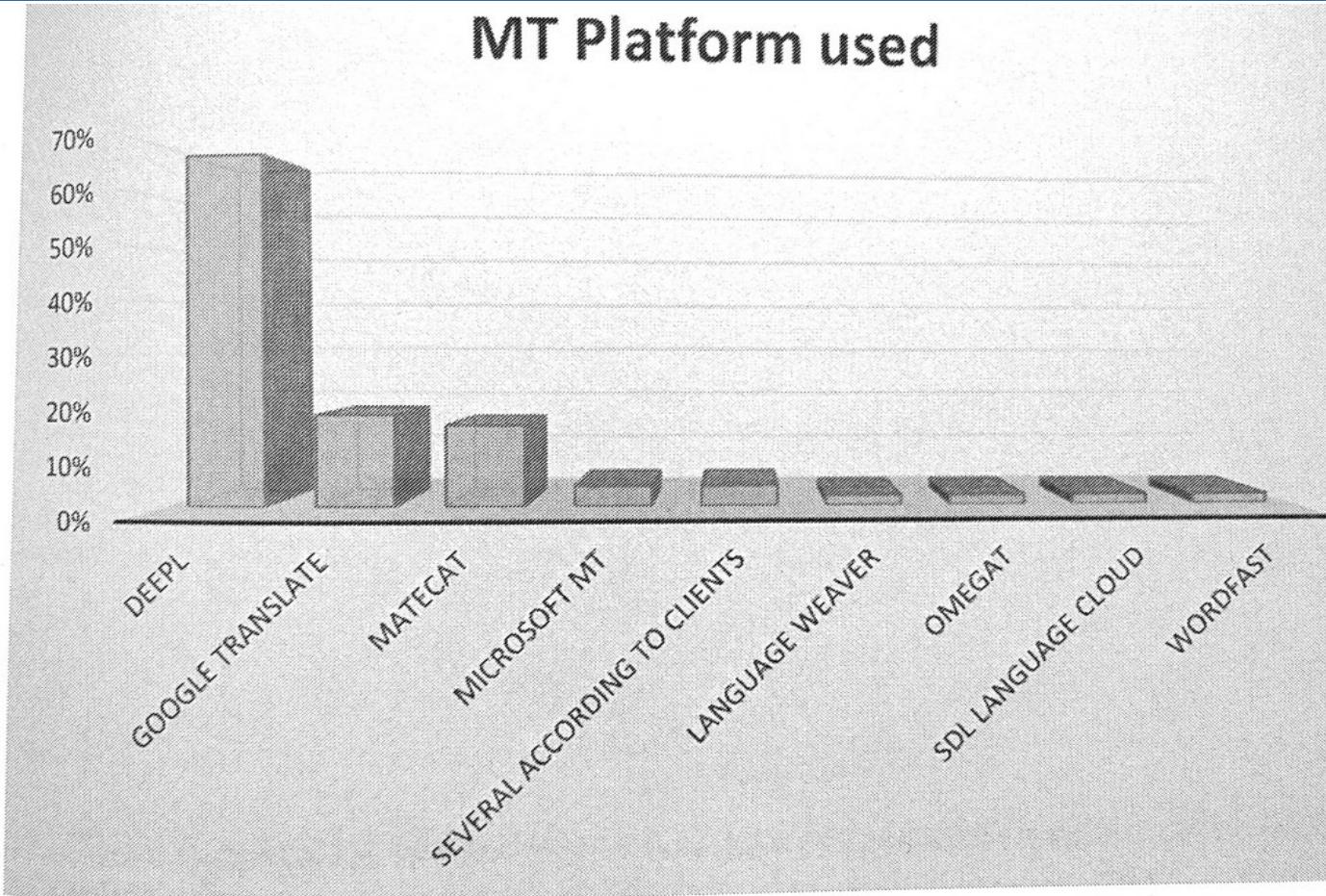
- Langlebigkeit
Support und Unterstützung
- Online oder lokal
- Kosten:
Zahlt sich ein Investment aus?
Investierte Zeit zum Erlernen des Tools,

(McDonough 2024: 9)

Lokale KI: Vorteile für Translator:innen

- Transparenz
- kein Vendor-Lock-in
- Nachhaltigkeit
- Frei für kommerzielle Zwecke
- Vollständige Kontrolle und Anpassung des Modells (Leistung, CO2-Abdruck, etc.)
- Einhalten der Geheimhaltungspflicht gegenüber Auftraggeber
- Einhalten von Copyright-Ansprüchen

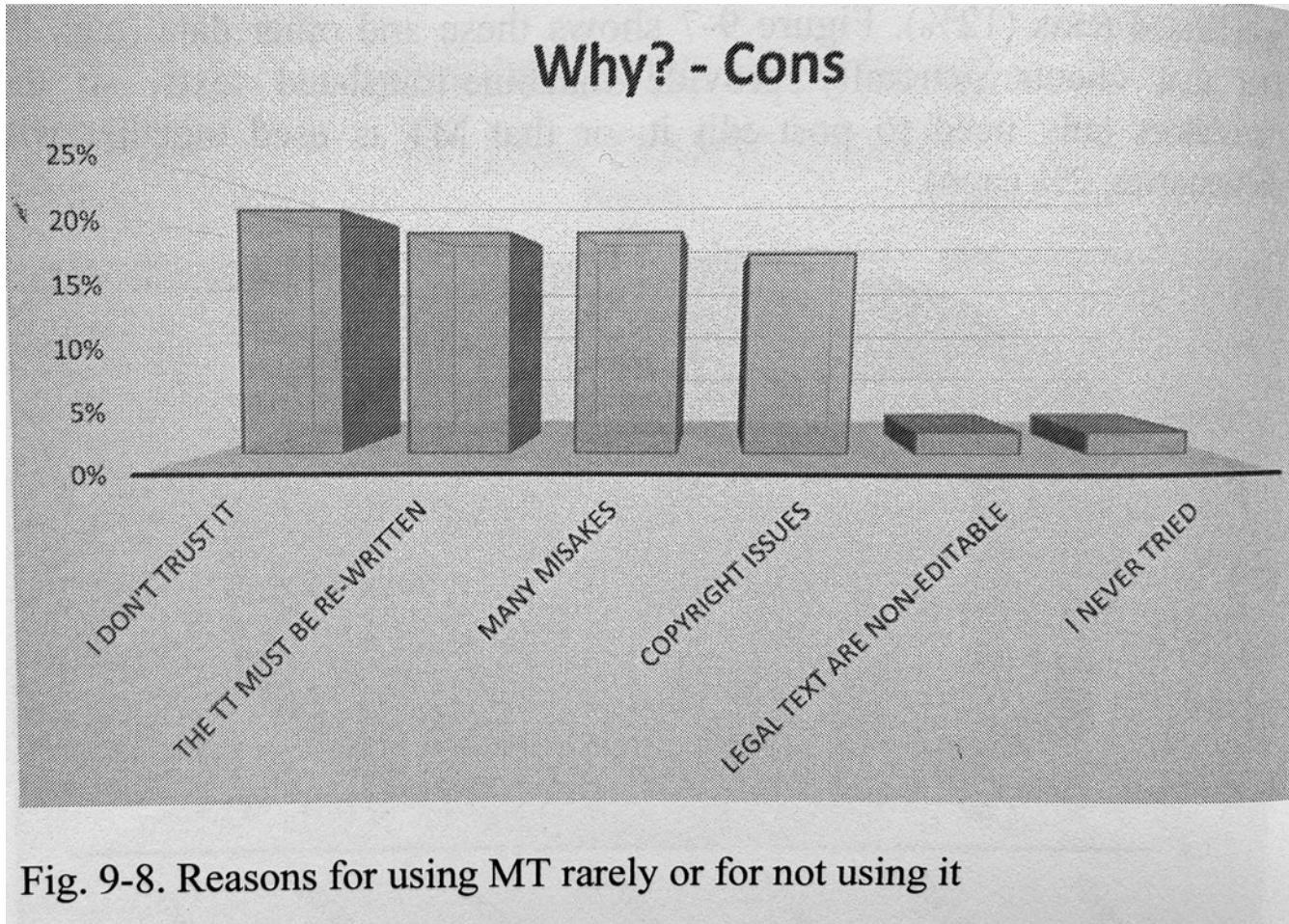
Online KI: Translator:innen im Recht?



Giampieri, P. (2023): Legal Machine Translation Explained. *MT in Legal Contexts*. Cambridge Scholars Publishing. 57.

Fig. 9-3. The MT platforms used by translators

Online KI: Translator:innen im Recht?



Giampieri, P. (2023): *Legal Machine Translation Explained. MT in Legal Contexts*. Cambridge Scholars Publishing. 61.

Wofür verwenden Translator:innen LLMs?

700 Translator:innen befragt:

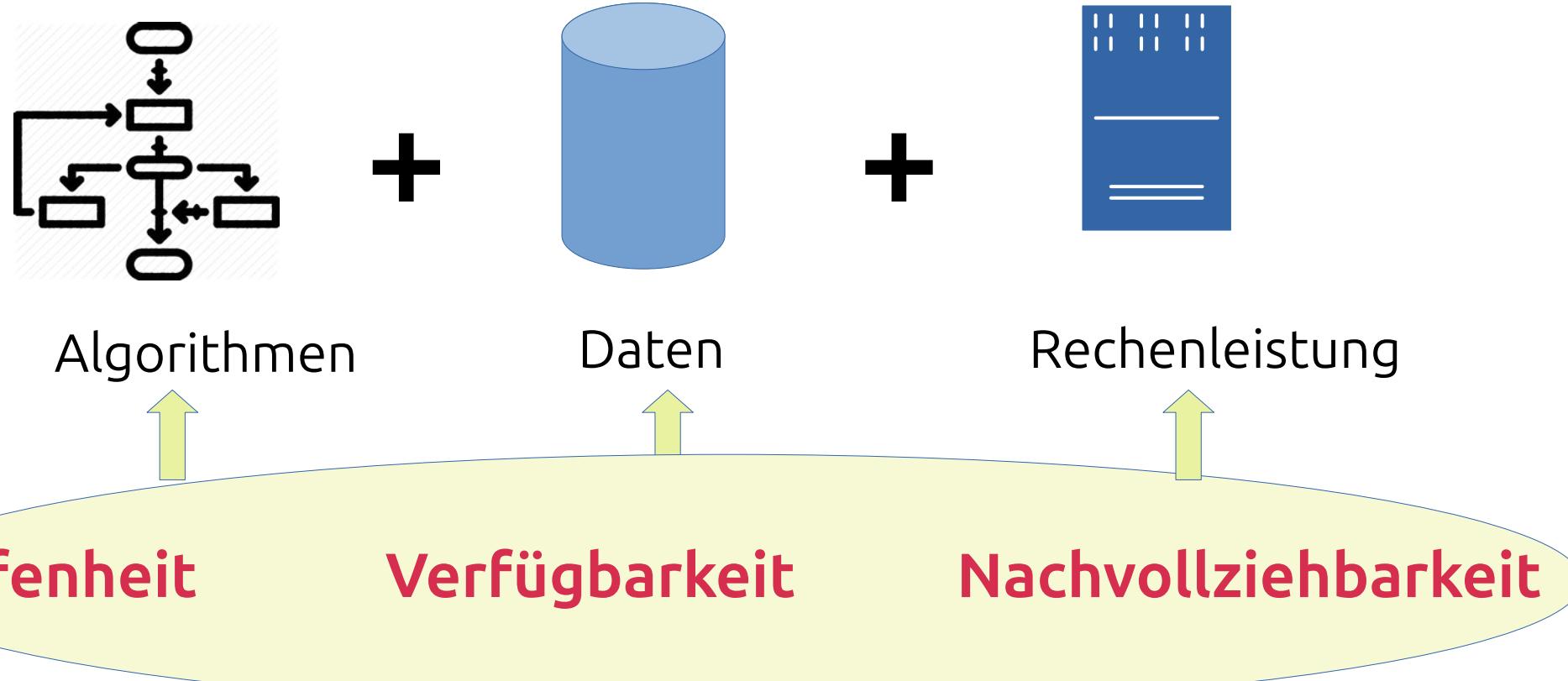
- > 30% Generating alternative target text options. MT puts out option 1, you don't like it, LLM creates option 2 or option 3 and voila, you have your preferred version for a final, light post-edit.
- 15% use AI to revise or improve fuzzy matches.
- 10 % use of AI's ability for creation and maintenance of glossaries, i.e. terminology extraction and review

223 LSPs (language service providers):

- 33 % have implemented large language models in their workflows. Almost all of the big Super Agencies have implemented LLMs at scale while just under 30% of the boutique LSPs have.

(Slator's 2024 Language Industry Market Report — Language AI Edition)

Voraussetzungen für lokale KI



Verarbeitung großer Datenmengen (NMT, LLM)

Hardware

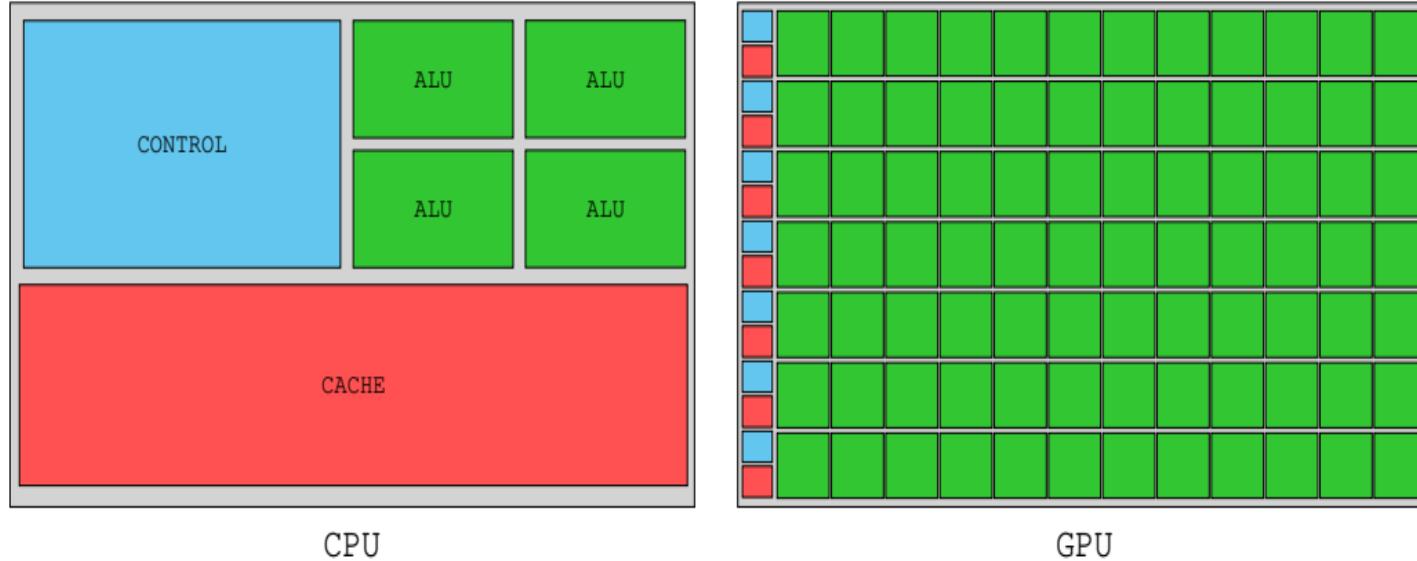
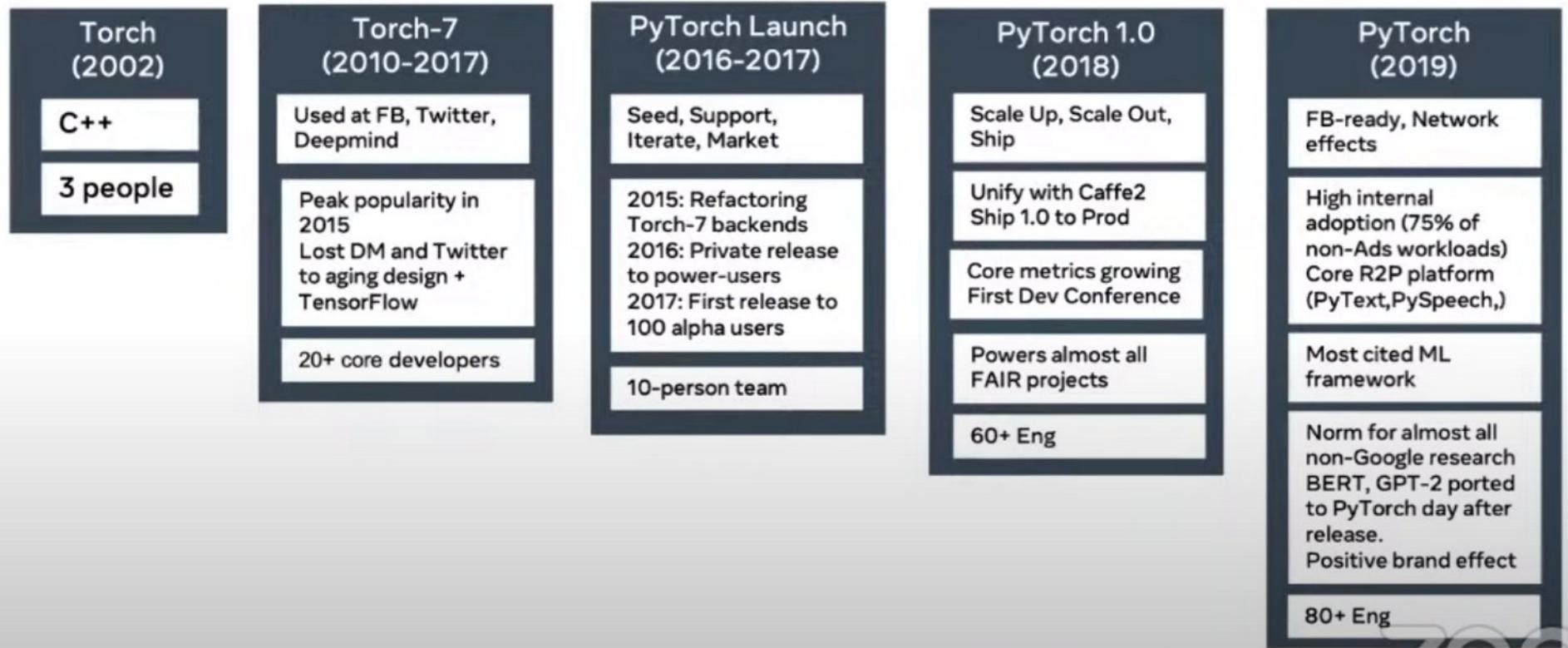


Figure 1: Visual comparison between a CPU and a GPU in terms of cache, control and processing units. Source: <https://fabiensanglard.net/cuda/>

Software: z.B. PyTorch: Python AI Library



Wahl eines LLM

- 1) Aufgabenstellung: z.B., natural language processing, computer vision, generative modeling
- 2) Fachgebiet: z.B., text classification, image segmentation, speech recognition
- 3) Hardwareressourcen: Wieviel Computer-Power und Speicher stehen zur Verfügung?
- 4) Prioritäten: Genauigkeit, Präzision, Effizienz, Latenz?

Frameworks (Container) für lokale LLMs

1) PrivateGPT

Setting Up PrivateGPT to Use AI Chat With Your Documents

<https://itsfoss.com/privategpt-setup/>

2) GPT4All

GPT4All <https://gpt4all.io/index.html>

3) Ollama

Running AI Locally Using Ollama on Ubuntu Linux

<https://www.jeremymorgan.com/blog/generative-ai/local-llm-ubuntu/>

<https://itsfoss.com/ollama-setup-linux/>

Run LLMs Locally on Raspberry Pi Using Ollama AI

<https://itsfoss.com/raspberry-pi-ollama-ai-setup/>

PRIVATE GPT

Mode

Query Docs

Search in Docs

LLM Chat

Upload a File

Ingested Files

File name

bitcoin.pdf

Chapter1-DigitalCurrency_DavidLee.p

LLM: local

According to David Lee's document, what is the major benefit of Bitcoin, and where in the document is it explained?



The major benefit of Bitcoin, as explained in the Handbook of Digital Currency, is its ability to enable fast transactions at low costs. This is mentioned in section 1.5.1 under the subheading "Freedom of payments".



Type a message...

Submit

Retry

Undo

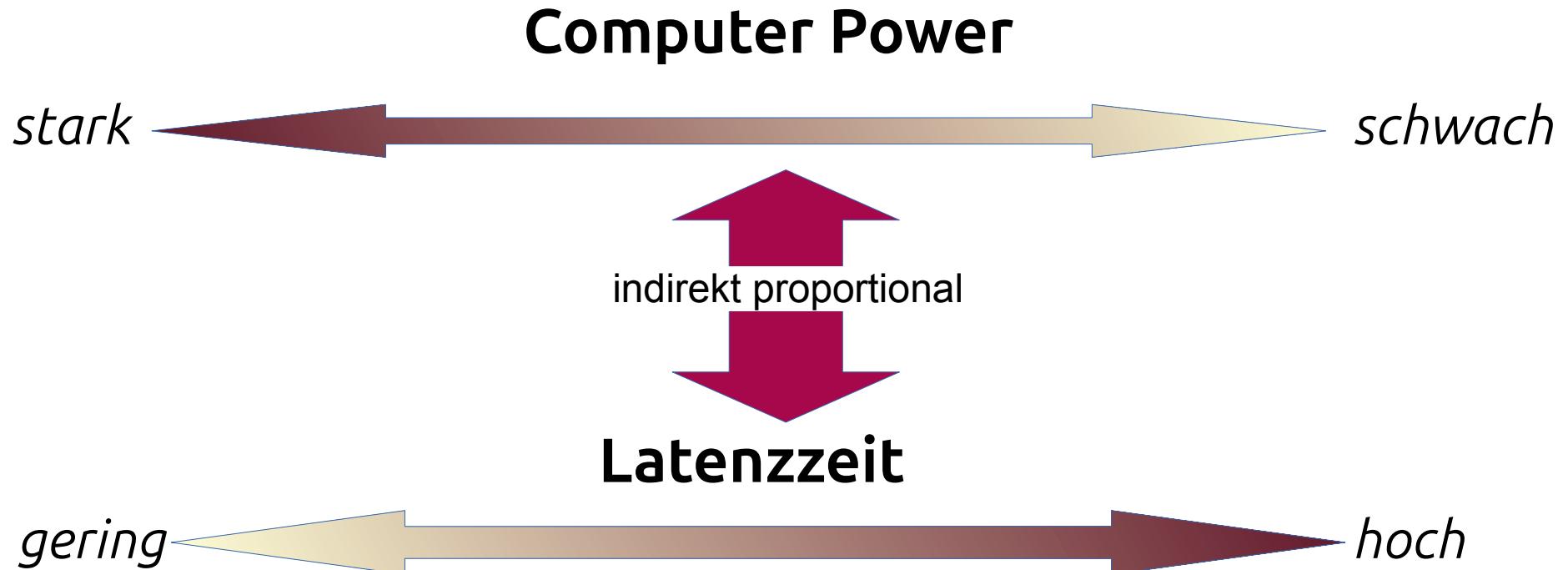
Clear

Verwendete Hardware für Beispiele (Büro:PC)

Prozessor	12th Gen Intel(R) Core(TM) i7-12700
Hauptspeicher	32,528 GB
Betriebssystem	XUbuntu 24.04 LTS
OpenGL-Renderer	Mesa Intel(R) UHD Graphics 770 (ADL-S GT1)

Alle Beispiele wurden auf dieser Hardware durchgeführt.

Lokale KI: Ein Balanceakt?



GPT4All

GPT4All v2.8.0

gpt4all.io | github nomic.ai | twitter | discord

+ New chat

New Chat

The conversati...

Electric vehicle...

Simple ...  

Updates

Downloads

About

Llama 3 Instruct

Llama 3 Instruct

Nous Hermes 2 Mistral DPO

GPT4All Falcon

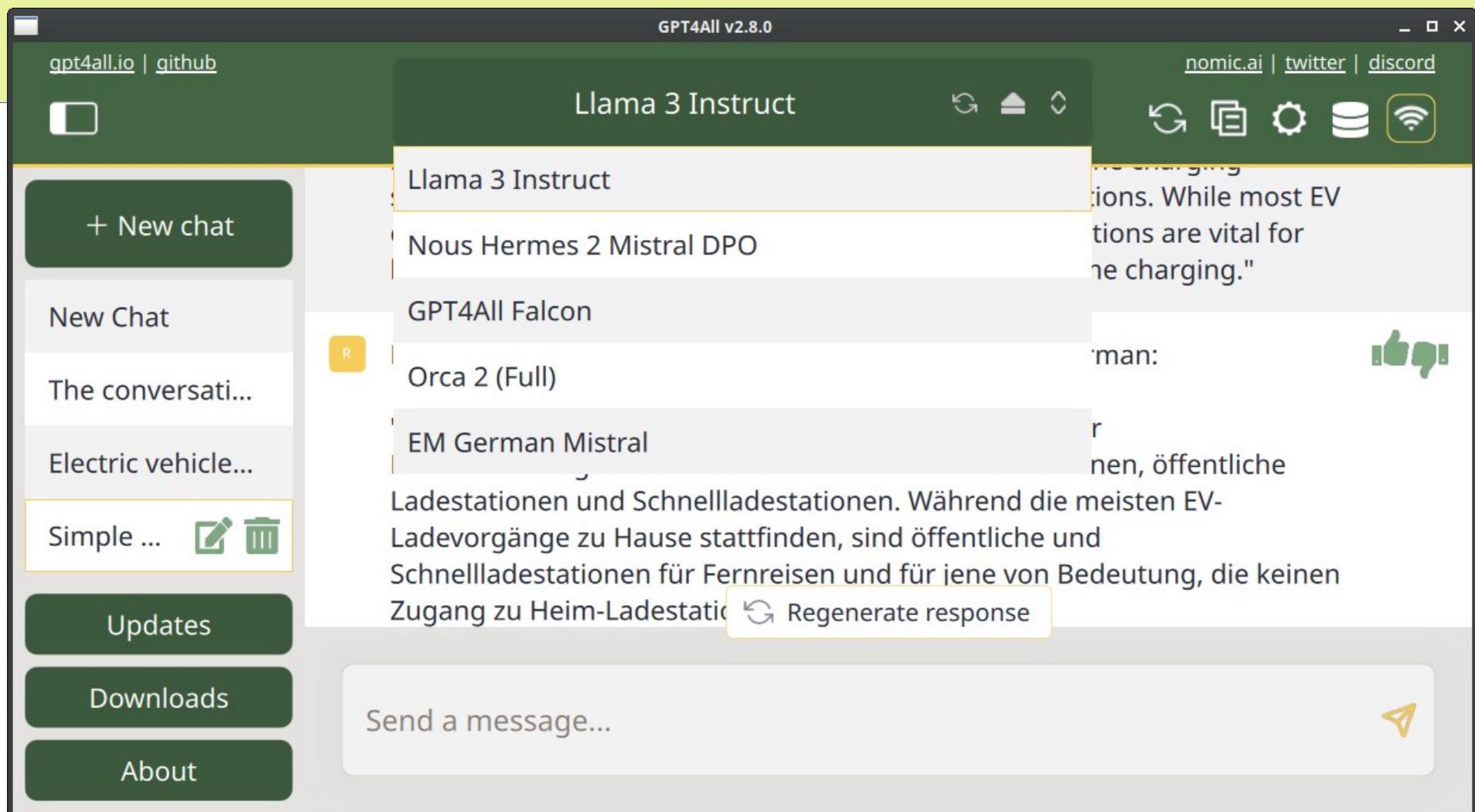
Orca 2 (Full)

EM German Mistral

Ladestationen und Schnellladestationen. Während die meisten EV-Ladevorgänge zu Hause stattfinden, sind öffentliche Schnellladestationen für Fernreisen und für jene von Bedeutung, die keinen Zugang zu Heim-Ladestatio

Regenerate response

Send a message... 



ScreenCast 1

- GPT4All: „Who is Slavoj Zizek?“
<file:///screen-gpt4all-llama3-zizek.mp4>
- Model Llama 3 Instruct
- Latenzzeit:
30 Sek. 3 Zeilen Antwort
2,10 Min. insgesamt

Status: Installed
File size: 4.34 GB
RAM required: 8 GB
Parameters: 8 billion
Quantization: q4_0
Type: LLaMA3

ScreenCast 1

- GPT4All: „Who is Slavoj Zizek?“
<file:///screen-gpt4all-orca-zizek.mp4>

- Model Orca 2 (Full)

Status: Installed
File size: 6.86 GB
RAM required: 16 GB
Parameters: 13 billion
Quantization: q4_0
Type: LLaMA2

- Latenzzeit:
50 Sek. 3 Zeilen Antwort
2,45 Min. insgesamt mit 10 Literaturangaben

Lokale LMs: Ollama

- Ollama CLI App installieren (<https://ollama.com/download?ref=hackernoon.com>)
<https://github.com/ollama/ollama?tab=readme-ov-file#web--desktop>
curl -fsSL https://ollama.com/install.sh | sh
- LLMs aus Bibliothek auswählen
<https://ollama.com/library?ref=hackernoon.com&sort=popular>
und downloaden
- Verfügbare lokale LLMs listen
ollama list

```
c61302@ThinkCentre-M70s:~$ ollama pull llama3
pulling manifest
pulling 6a0746alec1a... 100% [REDACTED] 4.7 GB
pulling 4fa551d4f938... 100% [REDACTED] 12 KB
pulling 8ab4849b038c... 100% [REDACTED] 254 B
pulling 577073ffcc6c... 100% [REDACTED] 110 B
pulling 3f8eb4da87fa... 100% [REDACTED] 485 B
verifying sha256 digest
writing manifest
removing any unused layers
success
```

```
c61302@ThinkCentre-M70s:~$ ollama list
NAME          ID      SIZE    MODIFIED
llama3:70b    786f3184aec0  39 GB   2 minutes ago
gemma:latest  a72c7f4d0a15  5.0 GB  7 hours ago
llama3:latest 365c0bd3c000  4.7 GB  7 hours ago
```

ScreenCast 2

- Ollama CLI
file:///screen-ollamaCLI-democracy.mp4 2:43 Min

```
>>> /show info
Model details:
Family           gemma
Parameter Size   9B
Quantization Level Q4_0
```

Ollama run llama3:70b

>>> please describe the professional outlook for human translators in the age of AI

The professional outlook for human translators in the age of AI is complex and multifaceted. While AI-powered translation tools have made significant progress, they also pose challenges and opportunities for human translators. Here are some key trends and observations:

****Challenges:****

1. ****Automation****: AI can automate routine, repetitive, and simple translation tasks, potentially replacing some human translators.
2. ****Commodification****: The increasing availability of free or low-cost machine translation (MT) tools may lead to downward pressure on prices and profit margins for human translators.
3. ****Changing client expectations****: Clients may expect faster turnaround times, lower costs, and higher quality translations, which can be challenging for human translators to meet.

Ollama run llama3:70b

Opportunities:

1. **Specialization**: AI is not yet capable of replicating the nuance, creativity, and cultural understanding that human translators bring to complex, high-stakes, or specialized projects (e.g., literary translations, marketing campaigns).
2. **Post-editing and review**: As MT output improves, human translators can focus on post-editing and reviewing AI-generated content, ensuring accuracy, fluency, and quality.
3. **Consulting and strategy**: Human translators can offer consulting services to help clients develop effective language strategies, select the right tools, and integrate AI solutions into their workflows.
4. **Niche areas**: Human translators can focus on niche areas, such as:
 - * Transcreation (adapting content for cultural and linguistic nuances)
 - * Localization (adapting software, apps, and websites for specific markets)
 - * Interpreting (real-time spoken language translation)

Ollama run llama3:70b

****Skills and adaptations:****

1. ****Technical skills****: Human translators need to develop proficiency in using AI-powered tools, such as MT platforms, terminology management systems, and project management software.
2. ****Domain expertise****: Specializing in a particular domain or industry can help human translators maintain their competitive edge.
3. ****Continuous learning****: Staying up-to-date with the latest developments in AI, language technology, and industry trends is crucial for human translators to remain relevant.
4. ****Collaboration and project management****: Human translators need to be able to work effectively with clients, project managers, and other stakeholders to deliver high-quality translations efficiently.

Ollama run llama3:70b

Industry outlook:

1. **Hybrid approach**: The translation industry will likely adopt a hybrid model, where AI-powered tools are used for routine tasks, while human translators focus on high-value tasks requiring nuance, creativity, and expertise.
2. **Growing demand**: Despite the rise of AI, the global demand for language services is expected to continue growing, driven by globalization, digitalization, and the increasing importance of international communication.

In conclusion, while AI-powered translation tools pose challenges for human translators, they also create opportunities for specialization, consulting, and high-value tasks. By developing technical skills, domain expertise, and collaborative project management capabilities, human translators can thrive in an industry that will increasingly rely on a hybrid approach combining the strengths of both humans and machines.

Ollama GUI

The screenshot shows a web-based interface for "Chat with Llama2". At the top, there's a header bar with a back button, forward button, refresh button, and a search bar containing "localhost:8000/?model=gemma%3Alatest". To the right of the search bar are zoom controls (170%), a star icon, and several small icons for user settings. Below the header, the main title "Chat with Ollama" is displayed. On the right side of the interface, there are configuration options: "System Prompt" (containing "You are a helpful assistant"), "Model" (set to "gemma:latest"), "Hostname" (set to "http://localhost:11434"), "History" (set to "Select a chat"), and buttons for "Reset", "Save", and "Delete". A blue input field contains the question "Who is Slavoj Zizek?". Below this, a dark gray box displays a response: "Slavoj Zizek is a renowned Slovenian philosopher, cultural theorist, and literary critic. He is known for his complex and thought-provoking theories on subjects such as:" followed by a bulleted list of his theories. At the bottom left, there's a text input field with placeholder text "Type your question here..." and a blue "Send" button. At the bottom right, a green circular badge contains the number "43".

Chat with Ollama

System Prompt
You are a helpful assistant

Model:
gemma:latest

Hostname
http://localhost:11434

History:
Select a chat

Reset
Save
Delete

Who is Slavoj Zizek?

Slavoj Zizek is a renowned Slovenian philosopher, cultural theorist, and literary critic. He is known for his complex and thought-provoking theories on subjects such as:

- **Radical philosophy:** Zizek argues that traditional philosophies are inadequate to address the complexities of our contemporary world and proposes radical approaches to understanding reality.
- **Cultural studies:** He explores the influence of popular culture, media, and consumerism on society, often using concepts like "enjoyment" and "simulation" to analyze these phenomena.
- **Psychoanalysis:** Zizek employs psychoanalytic theories to understand the workings of the unconscious mind and its influence on behavior and thought.
- **Political theory:** He critiques contemporary political systems and ideologies, arguing for a more nuanced and subversive understanding of power.

His work is often controversial due to its radical and unconventional ideas, but he is also highly respected for his intellectual depth and originality.

Type your question here... Send

43

ScreenCast 2

- Ollama GUI

file:///screen-ollama-gui-zizek.mp4 1:45 Min

- Model Gemma latest 4,7 GB

- Latenzzeit:

30 Sek. 3 Zeilen Antwort
1,45 Min. insgesamt

```
>>> /show info
Model details:
Family          gemma
Parameter Size  9B
Quantization Level Q4_0
```

Übersetzungsversuch mit Ollama LLama3

30 Sek.

>>> please translate this text from English into German: "Charging infrastructure has a significant role in the EV revolution. It includes home charging stations, public charging stations, and fast-charging stations. While most EV charging happens at home, public and fast-charging stations are vital for long-distance travel and for those without access to home charging."

Hier ist die Übersetzung ins Deutsche:

"Die Ladetechnologie spielt bei der Elektrofahrzeug-Revolution eine bedeutende Rolle. Sie umfasst Heim-Ladestationen, öffentliche Ladestationen und Schnell-Ladestationen. Während die meisten Ladevorgänge im **Heim** stattfinden, sind öffentliche und Schnell-Ladestationen für Fernreisen und für jene von Bedeutung, die keinen Zugang zu Heim-Ladestationen haben."

Ich hoffe, das ist korrekt!

GPT4All Orca 2 Full 50 Sek.

The screenshot shows the GPT4All v2.8.0 application window. The title bar reads "GPT4All v2.8.0". The top menu bar includes links to "gpt4all.io | github" and "nomic.ai | twitter | discord". The main content area displays two messages. The first message is a prompt from the user (P) asking for a German translation of a text about EV charging infrastructure. The second message is a response from the model (R) translating the same text into German. A "Regenerate response" button is visible at the bottom of the message list. The input field at the bottom left contains the placeholder "Send a message...".

P please translate this text from English into German: "Charging infrastructure has a significant role in the EV revolution. It includes home charging stations, public charging stations, and fast-charging stations. While most EV charging happens at home, public and fast-charging stations are vital for long-distance travel and for those without access to home charging."

R Ladeinfrastruktur spielt eine bedeutende Rolle in der Elektroauto-Revolution. Sie umfasst Ladegeräte für den privaten Gebrauch, öffentliche Ladevorgänge und Schnellladestationen. Während die meisten Ladevorgänge zu Hause erfolgen, sind öffentliche und Schnellladestationen für Fernreisen unverzichtbar und für diejenigen, die keinen Zugang zu einem Ladegerät im eigenen Haus haben.

Regenerate response

Send a message...

Übersetzungsversuch mit Ollama Gemma

25 Sek.

>>> please translate this text from English into German: "Charging infrastructure has a significant role in the EV revolution. It includes home charging stations, public charging stations, and fast-charging stations. While most EV charging happens at home, public and fast-charging stations are vital for long-distance travel and for those without access to home charging."

Die Ladinfrastruktur spielt eine entscheidende Rolle in der Elektromobil-Revolution. Sie umfasst Hausladestationen, öffentliche Ladestationen und Schnellladestationen. Während die meisten EV-Ladung zu Hause erfolgt, sind öffentliche und Schnellladestationen von entscheidender Bedeutung für Fernreisen und diejenigen ohne Zugang zu Hausladung.

GPT4All Falcon

The screenshot shows the GPT4All v2.8.0 application window. The top bar includes links to gpt4all.io and GitHub on the left, and nomic.ai, Twitter, and Discord on the right. The title bar says "GPT4All Falcon". The main area has a sidebar with a "+ New chat" button and a list of previous conversations: "The conve...", "Electric vehicle in...", "Translation: cultu...", and "Simple Time Disp...". At the bottom of the sidebar are buttons for "Updates", "Downloads", and "About". The main content area shows a prompt from the user (P) asking to translate a legal text from English to German. The AI response (R) provides the German translation, with some parts highlighted in blue. A "Regenerate response" button is at the bottom of the AI's message. A "Send a message..." input field is at the bottom, and a green circular icon with the number 48 is in the bottom right corner.

P please translate this legal text from English into German: "§6 CONFIDENTIALITY
1. During the employment relationship the Employee will not disclose to any third party including unauthorized employees of the Employer any confidential information with regard to the business of the Employer, in particular business and operational secrets and all information, data and documents which are marked as confidential or will not give any third party including unauthorized employees of the Employer access to such information and documents.
2. These obligations continue to apply after the termination of contractual relations."

R §6 Verpflichtung zur Geheimhaltung
1. Im Arbeitsverhältnis wird der Mitarbeiter nicht geglaubt, von Dritter zu verraten, einschließlich unautorisierte Angestellten des Arbeitgebers über die Geschäftsangelegenheiten des Arbeitgebers, insbesondere Geschäfts- und Betriebsgeheimnisse und alle Informationen, Daten und Dokumente, die als vertraulich eingeschätzt oder unerlaubt für Dritter zugänglich sind.
2. Diese Verpflichtungen gelten auch nach der Beendigung der Arbeitsbeziehungen.

Send a message... 

Regenerate response

48

Übersetzungsversuch 2 mit Ollama Gemma

32 Sek

>>> please translate this legal text from English into German: "§6 CONFIDENTIALITY

1. During the employment relationship the Employee will not disclose to any third party including unauthorized employees of the Employer any confidential information with regard to the business of the Employer, in particular business and operational secrets and all information, data and documents which are marked as confidential or will not give any third party including unauthorized employees of the Employer access to such information and documents.
2. These obligations continue to apply after the termination of contractual relations."

§ 6 CONFIDENTIALITÄT

1. Im Verlauf des Arbeitsverhältnisses verpflichtet sich der Mitarbeiter gegenüber jeder dritten Person einschließlich nicht autorisierter Mitarbeiter des Arbeitgebers, keine vertrauliche Informationen im Zusammenhang mit dem Geschäfts der Firma zu offengeben, insbesondere Geschäfts- und Betriebsgeheimnisse sowie alle Informationen, Daten und Dokumente, die als vertraulich gekennzeichnet sind oder denen die Zugang zu solchen Informationen und Dokumenten für Dritte einschließlich nicht autorisierter Mitarbeiter des Arbeitgebers versagt wird.
2. Diese Verpflichtungen bleiben nach Beendigung des Vertragsverhältnisses bestehen.

Übersetzungsversuch 2 mit Ollama LLama3

53 Sek.

>>> please translate this legal text from English into German: "§6 CONFIDENTIALITY

1. During the employment relationship the Employee will not disclose to any third party including unauthorized employees of the Employer any confidential information with regard to the business of the Employer, in particular business and operational secrets and all information, data and documents which are marked as confidential or will not give any third party including unauthorized employees of the Employer access to such information and documents.
2. These obligations continue to apply after the termination of contractual relations."

Here is the translation of the legal text from English into German:

§6 VERTRAULICHKEIT

1. Während der Dauer des Beschäftigungsverhältnisses wird der Arbeitnehmer keine vertraulichen Informationen an Dritte, einschließlich unautorisierten Arbeitnehmern des Arbeitgebers, offenbaren, soweit es sich um Geschäfts- und Betriebsgeheimnisse oder andere vertrauliche Informationen handelt, die als solche gekennzeichnet sind, oder jeglichen Dritten, einschließlich unautorisierten Arbeitnehmer des Arbeitgebers, keinen Zugang zu solchen Informationen und Dokumenten gewähren.
2. Diese Verpflichtungen gelten auch nach Beendigung der Vertragsbeziehung fort.

Note: I translated "confidential information" to "vertrauliche Informationen", and "business and operational secrets" to "Geschäfts- und Betriebsgeheimnisse".

Übersetzungsversuch 2 mit Ollama LLama3:70b

32 Min.

>>> please translate this legal text from English into German: "§6 CONFIDENTIALITY

1. During the employment relationship the Employee will not disclose to any third party including unauthorized employees of the Employer any confidential information with regard to the business of the Employer, in particular business and operational secrets and all information, data and documents which are marked as confidential or will not give any third party including unauthorized employees of the Employer access to such information and documents.
2. These obligations continue to apply after the termination of contractual relations."

§6 VERTRAULICHKEIT

1. Während des Arbeitsverhältnisses wird der Mitarbeiter keine vertraulichen Informationen über das Geschäft des Arbeitgebers, insbesondere Geschäfts- und Betriebsgeheimnisse sowie alle als vertraulich gekennzeichneten Informationen, Daten und Dokumente an Dritte, einschließlich nicht befugter Mitarbeiter des Arbeitgebers, weitergeben oder diesen Zugang zu solchen Informationen und Dokumenten gewähren.
2. Diese Verpflichtungen bleiben auch nach Beendigung der vertraglichen Beziehungen wirksam."

Please note that legal translations should always be done by a qualified translator or a lawyer who is fluent in both languages, as the nuances of legal language can be complex and culturally specific.

Übersetzungsversuch 2 mit GPT4All Orca 2 (Full)

1:10 Min.

The screenshot shows the GPT4All v2.8.0 application window. The title bar says "GPT4All v2.8.0" and "Orca 2 (Full)". The left sidebar has buttons for "+ New chat", "Legal t...", "Electric vehicl...", "Updates", "Downloads", and "About". The main area shows a message from the user: "please translate this legal text from English into Italian: \"§6 CONFIDENTIALITY
1. During the employment relationship the Employee will not disclose to any third party including unauthorized employees of the Employer any confidential information with regard to the business of the Employer, in particular business and operational secrets and all information, data and documents which are marked as confidential or will not give any third party including unauthorized employees of the Employer access to such information and documents.
2. These obligations continue to apply after the termination of contractual relations.\"". The AI's response is: "§6 CONFIDENZIALITÀ
1. Durante il rapporto di lavoro, l'Impiegato non rivelerà a nessuna terza parte, compresi dipendenti non autorizzati dell'Imprenditore, alcuna informazione riservata riguardo agli affari dell'Imprenditore, in particolare segreti commerciali e operativi e tutte le informazioni, i dati e i documenti che sono marchiati come riservati o non darà a nessuna terza parte, compresi dipendenti non autorizzati dell'Imprenditore, accesso a tale informazione e documenti.
2. Queste obbligazioni continuano ad applicarsi dopo la fine delle relazioni contrattuali." A "Regenerate response" button is visible below the AI's text. At the bottom, there is a "Send a message..." input field and a send icon. The top right corner shows links to "nomic.ai", "twitter", and "discord".

Übersetzungsversuch 2 mit GPT4All Llama 3 Instruct Min.

GPT4All v2.8.0

gpt4all.io | github nomic.ai | twitter | discord

Llama 3 Instruct

P Please translate this legal text from German into Italian
"§6 VERTRAULICHKEIT
1. Während des Arbeitsverhältnisses wird der Mitarbeiter keine vertraulichen Informationen über das Geschäft des Arbeitgebers, insbesondere Geschäfts- und Betriebsgeheimnisse sowie alle als vertraulich gekennzeichneten Informationen, Daten und Dokumente an Dritte, einschließlich nicht befugter Mitarbeiter des Arbeitgebers, weitergeben oder diesen Zugang zu solchen Informationen und Dokumenten gewähren.
2. Diese Verpflichtungen bleiben auch nach Beendigung der vertraglichen Beziehungen wirksam."

R Here is the translation of the legal text from German into Italian:

"§6 CONFIDENZIALITÀ
1. Durante il rapporto di lavoro, l'insegnante non divulgherà informazioni confidenziali relative all'azienda del datore di lavoro, in particolare segreti commerciali e segreto industriale, nonché tutte le informazioni, i dati e documenti contrassegnati come confidenziali a terzi, inclusi dipendenti non autorizzati dell'azienda.
2. Queste obbligazioni restano efficaci anche dopo la fine delle relazioni contrattuali."

Note: I translated "Mitarbeiter" as "insegnante", which is the most common translation for this term in an employment context, but it could also be translated as "employee". If you need a different translation, please let me know.

Regenerate response

Send a message... 

53

Überprüfen von Übersetzungsvorschlägen aus MT (Opus-Cat-MT)

Prompt:

Please check and proofread this legal translation from English to German and evaluate accuracy and fluency of the target text

source:

"§6 CONFIDENTIALITY

1. During the employment relationship the Employee will not disclose to any third party including unauthorized employees of the Employer any confidential information with regard to the business of the Employer, in particular business and operational secrets and all information, data and documents which are marked as confidential or will not give any third party including unauthorized employees of the Employer access to such information and documents.
2. These obligations continue to apply after the termination of contractual relations."

target:

"§6 VERTRAULICHKEIT

1. Während des Beschäftigungsverhältnisses wird der Mitarbeiter keine vertraulichen Informationen in Bezug auf das Geschäft des Arbeitgebers, insbesondere Geschäfts- und Betriebsgeheimnisse und alle Informationen, Daten und Dokumente, die als vertraulich gekennzeichnet sind, an Dritte weitergeben oder Dritten, einschließlich nicht autorisierter Mitarbeiter des Arbeitgebers, keinen Zugang zu diesen Informationen und Dokumenten gewähren.
2. Diese Verpflichtungen gelten auch nach Beendigung des Vertragsverhältnisses weiter."

Überprüfen von Übersetzungsvorschlägen aus MT (Opus-Cat-MT)

I've reviewed the translation and provided feedback on accuracy and fluency:

****Accuracy: 9/10****

The translation is largely accurate, with a few minor suggestions for improvement:

- * "including unauthorized employees of the Employer" could be translated to "einschließlich nicht befugter Mitarbeiter des Arbeitgebers" to better reflect the nuance of "unauthorized".
- * "will not give any third party... access to such information and documents" is correctly translated as "gewähren keinen Zugang zu diesen Informationen und Dokumenten", but the word order could be improved for better fluency (see below).

****Fluency: 8.5/10****

The translation is clear and understandable, but there are a few areas where the sentence structure and wording could be improved for better readability:

- * In sentence 1, the phrase "Während des Beschäftigungsverhältnisses wird der Mitarbeiter keine vertraulichen Informationen..." is a bit long and convoluted. Consider breaking it up into two simpler sentences: "Während des Beschäftigungsverhältnisses ist der Mitarbeiter verpflichtet, keine vertraulichen Informationen zu offenbaren. Dies gilt insbesondere für Geschäfts- und Betriebsgeheimnisse..."
- * In sentence 1, the phrase "alle Informationen, Daten und Dokumente, die als vertraulich gekennzeichnet sind" could be shortened to "vertrauliche Informationen, Daten und Dokumente" for better brevity.
- * In sentence 2, the phrase "Diese Verpflichtungen gelten auch nach Beendigung des Vertragsverhältnisses weiter" is a bit formal. Consider rephrasing it to "Die Verpflichtung zur Vertraulichkeit bleibt auch nach Ende des Arbeitsverhältnisses bestehen."

Here's the revised translation incorporating these suggestions:

"§6 VERTRAULICHKEIT

1. Während des Beschäftigungsverhältnisses ist der Mitarbeiter verpflichtet, keine vertraulichen Informationen zu offenbaren. Dies gilt insbesondere für Geschäfts- und Betriebsgeheimnisse sowie alle vertraulichen Informationen, Daten und Dokumente. Er wird Dritten, einschließlich nicht befugter Mitarbeiter des Arbeitgebers, keinen Zugang zu diesen Informationen und Dokumenten gewähren.
2. Die Verpflichtung zur Vertraulichkeit bleibt auch nach Ende des Arbeitsverhältnisses bestehen."

Framework: Ollama GUI
Model: Llama3:70B
Zeit: 1:47:14

Überprüfen von Übersetzungsvorschlägen aus MT (Opus-Cat-MT)

Framework: Ollama GUI
Model: Llama3
Zeit: 1:55

I've reviewed the translation and provided feedback on accuracy and fluency:

****Accuracy:****

- * The translation is largely accurate, with only a few minor issues:
 - + In §6.1, "Geschäfts- und Betriebsgeheimnisse" should be translated as "business and trade secrets" instead of "operational secrets". This is because the word "Betriebsgeheimnisse" typically refers to internal company information, not external business secrets.
 - + In §6.2, it's better to use "ab dem Zeitpunkt des Vertragsende" instead of "nach Beendigung des Vertragsverhältnisses weiter". This phrase is more commonly used in German employment law contracts.

****Fluency:****

- * The translation reads smoothly and is easy to understand. However, there are a few areas where the language could be improved:
 - + In §6.1, "in Bezug auf das Geschäft des Arbeitgebers" could be rephrased as "betreffend das Geschäft des Arbeitgebers" or simply "betreffend das Unternehmen" to make it more natural-sounding.
 - + In §6.2, the phrase "Diese Verpflichtungen gelten auch nach Beendigung des Vertragsverhältnisses weiter" could be shortened to "Diese Verpflichtungen bestehen auch nach dem Vertragseinstieg fort".

****Additional suggestions:****

- * Consider adding a definition of what constitutes "confidential information" or "vertrauliche Informationen" to ensure clarity for all parties involved.
- * The translation uses the phrase "Dritten, einschließlich nicht autorisierter Mitarbeiter des Arbeitgebers" to refer to unauthorized employees. While this is accurate, it might be more concise to simply use "Dritten und unautorisierten Mitarbeitern".
- * If you want to maintain a consistent tone and style throughout the contract, consider revising the translation to match the formal, legal language used in the rest of the document.

Overall, the translation is good, but could benefit from some minor tweaks to improve fluency and accuracy.

Lokale KI: Ein Balanceakt?

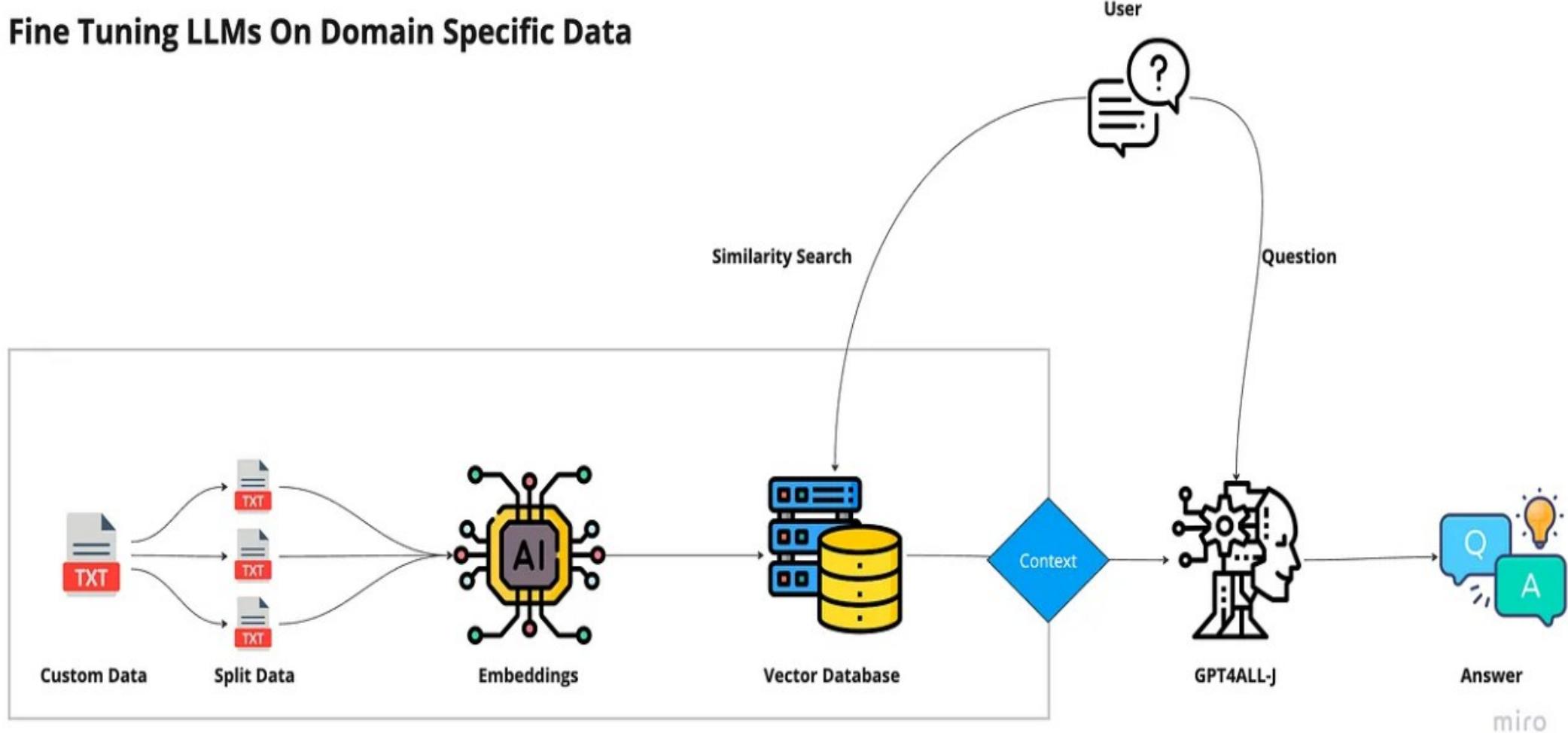


LLM und Korpus an Referenztexten

GPT4All und PrivateGPT (mit Ollama) können eigene Texte als Referenzmaterial verwenden

- 1) Texte werden in kleinere Segmente zerteilt
- 2) Wörter- und Wortkombinationen werden in Zahlenwerte umgewandelt (Embeddings) und in Relation gestellt (Vektoren)
- 3) Vektoren werden in Datenbank gespeichert, auf die das LLM bei Abfrage zurückgreift
- 4) Diese bilden den Kontext für spezifische Abfragen

Fine Tuning LLMs On Domain Specific Data



z.B. in GPT4All

GmbH-Texte: 14 Textdateien 1,4 MB

Investment-Fonds: 21 Textdateien 170 KB

KI als Alignment Tool

GPT4All Llama3 Instruct

please give me a translation memory in TMX-Format for all the sentences in the reference files. Entries should contain German sentence and English sentence, and translate the German sentences into English.

Here is the translation memory (TMX) file with the requested entries:

```
<?xml version="1.0" encoding="UTF-8"?>
<TMX xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="tmx-tm.dtd tmx-tm.xsd">
  <tuConversion>
    <!-- Entry 1 -->
    <tuv>
      <tuid>entry_1</tuid>
      <segSource><seg id="0" lang="de">Gesamte Rechtsvorschrift für GmbH- Gesetz, Fassung vom 09.10.2019 Beachte für folgende Bestimmung Im Titel der BGBl. I Nr. 114/1997 findet sich folgende Fußnote: Diese Kundmachung ersetzt die Kundmachung BGBl. I Nr. 106/1997.</seg></segSource>
      <segTarget><seg id="0" lang="en">Complete legal provisions for GmbH law, version of October 9th, 2019 Note that the following provision is applicable: In the title of BGBl. I No. 114/1997, there is a footnote stating that this publication replaces the announcement BGBl. I No. 106/1997.</seg></segTarget>
      </tuv>
    <!-- Entry 2 -->
    <tuv>
      <tuid>entry_2</tuid>
      <segSource><seg id="0" lang="de">Langtitel Gesetz vom 6. März 1906, BGBl. I Nr. 70/2008 treten mit 1. Juni 2008 in Kraft.</seg></segSource>
      <segTarget><seg id="0" lang="en">Long title Law of March 6th, 1906, BGBl. I No. 70/2008 enters into force on June 1st, 2008.</seg></segTarget>
      </tuv>
```

KI als Terminologie-Extraktionsstool

GPT4All Llama3 Instruct

please write me a list of the 20 most important terms in tab-delimited format with the German term, the English term, a brief definition, the frequency, from the reference texts; do also consider specialized terms consisting of multiple words

After analyzing the provided text and relevant references (§10-12 Verfugungen, Tod und Einziehung), I've compiled a list of 20 important terms in tab-delimited format. Please note that some terms might be specific to German law or financial regulations. Here is the list:

German Term	English Term	Definition	Frequency
Verfugung	Disposition	A decision-making process	2
Einziehung	Sequestration	The act of taking possession	1
Tod	Death	Refers to a person's passing away	3
Sitzungsleitung	Meeting Leadership	Leading or chairing a meeting	2
Protokollierung	Protocolization	Recording and documenting meetings	2
Vertretung	Representation	Acting on behalf of someone else	4
Einladung	Invitation	A formal request to attend an event	1
Form	Form	Refers to the structure or layout	3
Frist	Deadline	The time limit for completing a task	2
Mehrheit	Majority	Having more than half of the votes cast	5
Sondermehrheit	Special Majority	A specific type of majority required	1
Einigung	Agreement	Coming to an understanding or consensus	4
Übertragung	Transfer	The act of passing something from one person	2
Tätigkeit	Activity	Refers to a person's work, occupation, etc.	3
Definition	Definition	A clear and concise explanation of a term	5
Richtlinie	Directive	A set of rules or guidelines issued by the EU	4
Verordnung	Regulation	A formal rule or law governing an area	2
Ergänzung	Supplement	Adding to or expanding something existing	1 ...

Open Source quality is rapidly advancing – while fine tuning cost is rapidly decreasing

Dolly started the trend to open models with a commercially friendly license



Facebook Llama



Stanford Alpaca



Databricks Dolly



Mosaic MPT



TII Falcon

"Smaller, more performant models such as Llama ... democratizes access in this important, fast-changing field."

"Alpaca behaves qualitatively similarly to OpenAI ... while being surprisingly small and easy /cheap to reproduce"

"Dolly will help democratize LLMs, transforming them into a commodity every company can own and customize"

"MPT-7B is trained from scratch on 1T tokens ... is open source, GPT-3 for ... 75% of the training available for commercial use, and compute budget—and ... a fifth of the quality of LLaMA-7B" matches the quality of LLaMA-7B" the compute at inference time.'

February 24, 2023

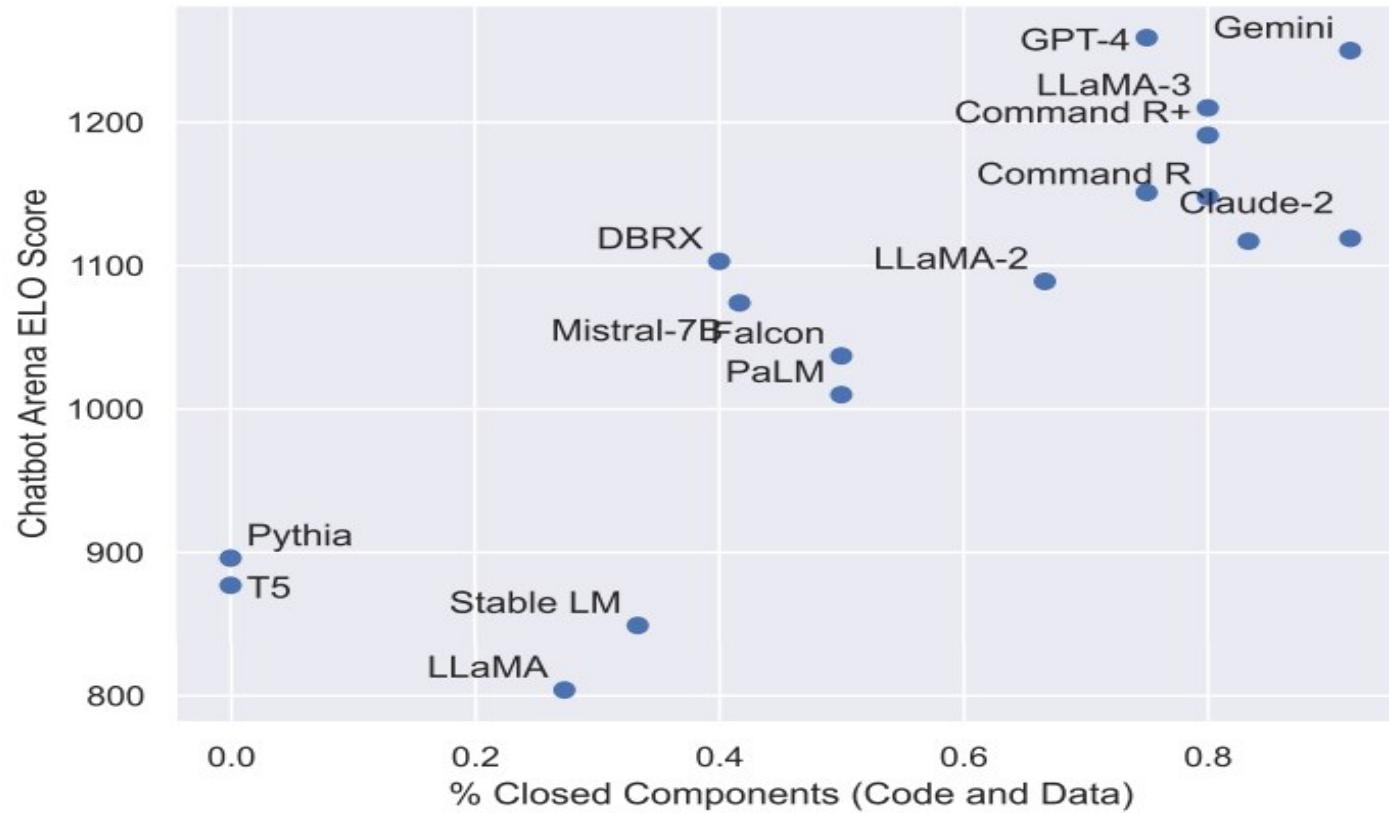
March 13, 2023

March 24, 2023

May 5, 2023

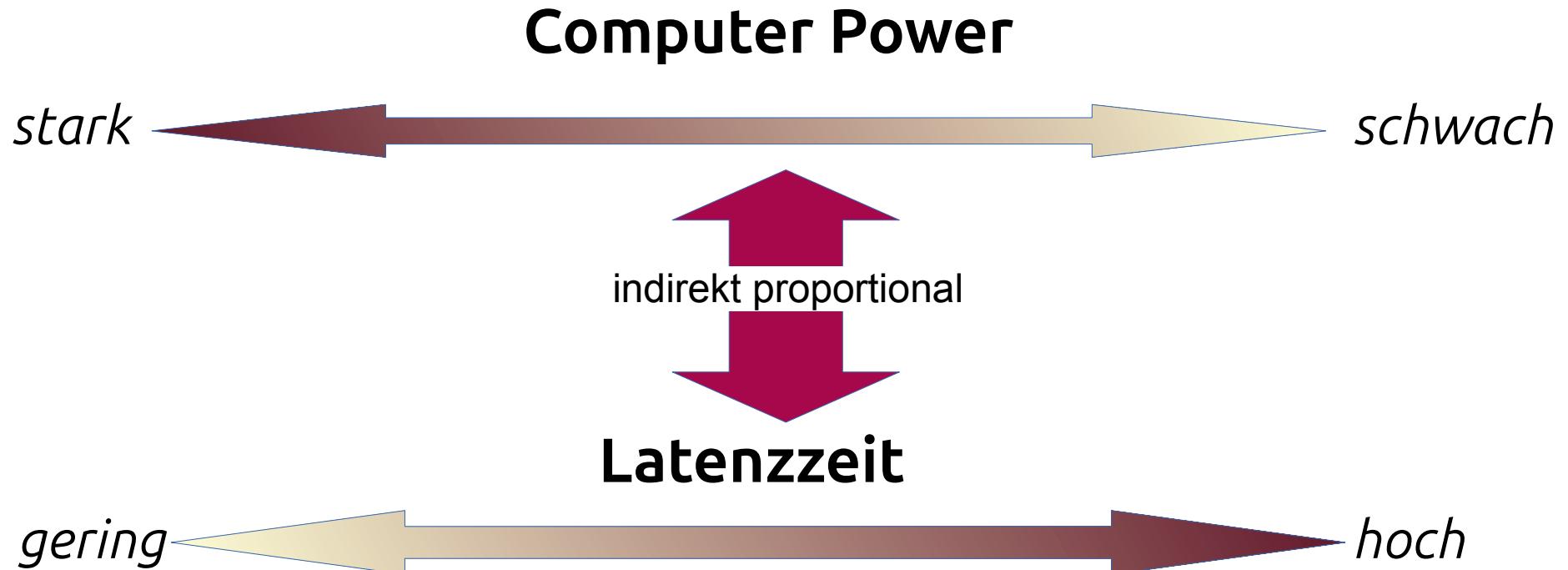
May 24, 2023

Chancengleichheit?



(b) Closed Models Currently Outperform Open Ones

Lokale KI: Ein Balanceakt?



Lokale KI: Ein Balanceakt?



Lokale KI: Ein Balanceakt?



Lokale LLM: Fazit

- Abwägen der lokalen Voraussetzungen
- Abwägen des optimalen Frameworks
- Testen und Evaluieren der zur Verfügung stehenden Basismodelle
- Testen und Evaluieren der Anwendungszecke

Unter optimalen Bedingungen kann eine lokal installiertes LLM eine nützliche Ergänzung der technischen Infrastruktur von Translator:innen sein.

Literatur

Eiras, F.; Petrov A. Et al. (2024): Risks and Opportunities of Open-Source Generative AI.
arXiv:2405.08597v3

Free Software Foundation Europe (2022): Public Money? Public Code!
<https://media.fsfe.org/w/7cH6EATpmPwaMsEZN8vfj>

Giampieri, P. (2023): Legal Machine Translation Explained. MT in Legal Contexts. Cambridge Scholars Publishing.

GNU.org (2016): Freie Software und Bildung. Wie bezieht sich Freie Software auf Bildung?
<https://www.gnu.org/education/education.html>

McDonough Dolmaya, Julie (2024): Digital Research Methods for Translation Studies. New York: Routledge.

Mozilla (2024): Towards a Framework for Openness in Foundation Models
https://assets.mofoprod.net/network/documents/Towards_a_Framework_for_Openness_in_Foundation_Models.pdf?ref=news.itsfoss.com

Semetey, Raphaël (2024): From OpenAI to Open Source AI. Fileshare.com

Kommentare, Fragen, Anregungen, Kritik?



LET'S TALK