



Grundlagen digitaler Mehrsprachigkeits- prozesse

Oliver Czulo
dikomm-Workshop
14. November 2024

Ausprobieren :)

mit einem Übersetzungstool deiner Wahl



Wohnberechtigungsschein

(und danach: „Antrag auf Erteilung eines
Wohnberechtigungsscheines“)



Probe 2

jurisdiction



Beobachtungen

- KI hat kein Weltverständnis
 - während der Texteingabe ins Übersetzungstool zeigt sich: KI-Ausgaben ändern sich je nach *Ko-Text* (also umgebendem Textmaterial)
 - je nach Übersetzungstool variiert das Ergebnis: KI kann keine terminologischen Entscheidungen treffen
- Hauptkenntnisziel: **Verlässliche** und **nachvollziehbare** Übersetzungsentscheidungen bedürfen **kontrollierbarer Ressourcen, klarer Prozesse** und **menschlicher Expertise**



Wie uns KI gemeinhin präsentiert wird





Die ganz weltlichen Probleme der KI

ntv

RESSORTS SPORT BÖRSE WETTER VIDEO AUDIO TV PROGRAMME

Samstag, 25. Mai 2024 13:10 Uhr Frankfurt

Zugteilung oder Explosion?

Handy-Übersetzung löst Polizei-Großeinsatz aus

14.05.2024, 17:47 Uhr

Artikel anhören



Er wollte eigentlich nur fragen, wo der Zug geteilt wird, fasst die Polizei den Einsatz zusammen.
(Foto: picture alliance / Maximilian Koch)



Folgen auf:

Zugteilungen sind eine Herausforderung für viele Bahnreisende - für nicht-deutschsprachige Passagiere sowie offenbar für Übersetzungsprogramme erst recht. Ein Iraker bat mithilfe seines Handys eine Mitfahrerin um Auskunft. Das ging gewaltig schief. Am Ende kommt der Zugverkehr zum Erliegen.

Er hat nur wissen wollen, wo sein Zug geteilt wird - aber sein Handy kündigte eine Zugexplosion an: Ein Übersetzungsfehler eines Hilfe suchenden irakischen Staatsbürgers auf seinem Handy hat im fränkischen Oberkotzau einen Großeinsatz der Polizei ausgelöst. Wie ein Sprecher der Bundespolizei in Selb sagte, musste deshalb am Montagabend eine Regionalbahn mit 80 Reisenden evakuiert werden. Der 36 Jahre alte Mann kam nach einer zwischenzeitlichen Festnahme wieder frei.

Andere Investoren wie Aaron Fleishman von Tola Capital verweisen wiederum darauf, dass man sich zu sehr auf "prognostiziertes exponentielles Wachstum" verlassen habe, um die extrem hohen Bewertungen von KI-Start-ups zu rechtfertigen. Firmen wie Stability AI seien da ein warnendes Beispiel: Dieses sei bereits Ende 2022 mit einer Milliarde Dollar bewertet worden, nur um dann im Jahr 2023 gerade einmal elf Millionen US-Dollar an Umsatz zu generieren. Das übrigens bei Betriebskosten in der Höhe von 153 Millionen US-Dollar.

aus Der Standard, 01.04.2024, "Der KI-Hype trifft auf die ersten Spuren aus Realität"

verbessert. Und das hat einen ganz trivialen Grund, den man in der Mathematik "Model Collapse" nennt. In ChatGPT steckt bereits das gesamte Internet – ein zweites Internet haben wir nicht. Ein großer Teil der Inhalte, die jetzt und in Zukunft im Internet dazukommen, werden bereits durch KI erzeugt. Es wurden jetzt schon durch Künstliche Intelligenz mehr Bilder erzeugt, als jemals Fotos von Kameras geschossen wurden. Wenn ich aber nun ein KI-System mit KI-generiertem Zeug füttere, dann wird es schlechter. Die Informationsdichte in dem Ding sinkt. Es wird dümmmer. Ein Bekannter von mir nennt es immer "Habsburg-KI".

aus turi2, 19.02.2024, "KI ist keine magische Box, die alle Probleme löst"



Akshat Rathi
@akshatrathi@mastodon.green

2020: Microsoft sets goal to be carbon negative by end of the decade.

2023: Microsoft's emissions are 30% higher than in 2020.

Main cause? The relentless push to meet AI demand, which requires new data centers built out of carbon-intensive steel, cement, chips.

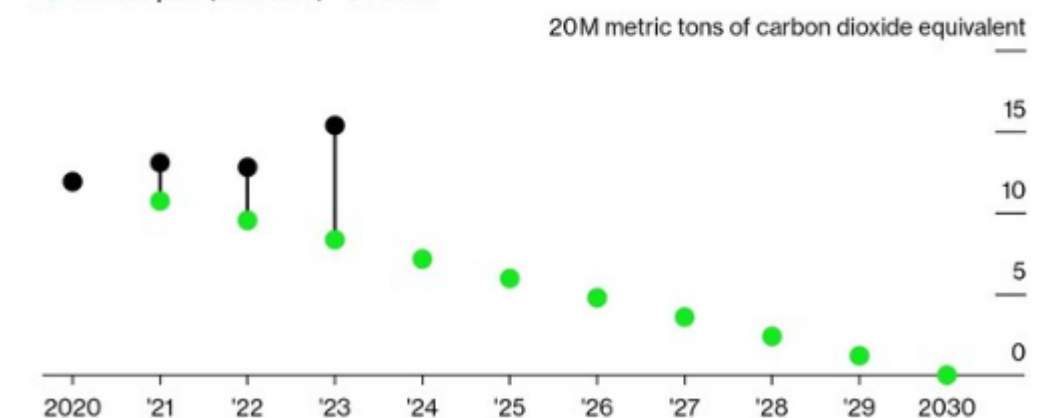
[bloomberg.com/news/articles/20...](https://www.bloomberg.com/news/articles/2024-05-15-microsoft-emissions-are-30-percent-higher-than-in-2020)



Microsoft's Emissions

Artificial intelligence is putting the tech giant's climate goals in peril

● Climate plan (simulated) ● Actual

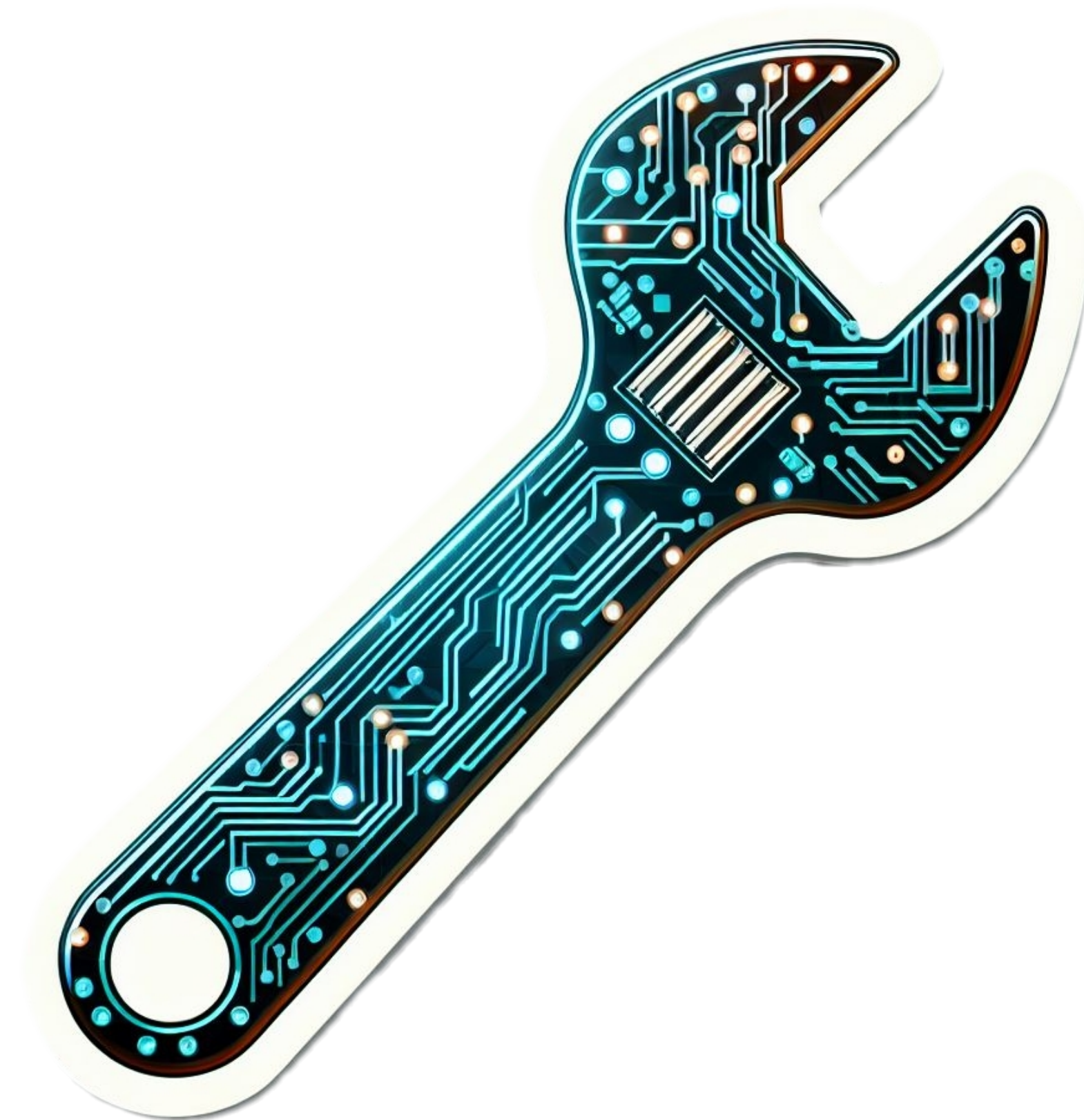


Source: Microsoft (Scope 1, 2 and 3 "management criteria" data)
Note: Green dots represent linear decline to carbon negative goal.

May 15, 2024, 18:11 · 1.4K · 972



Alles vergebens? Nein! Was KI eigentlich ist:



Fokus: Neuronale Maschinelle Übersetzung (NeuroMÜ)

Eine KI der *Translation*



Translation und Digitalität

- *Translation* (aus dem Lat.): Oberbegriff fürs Übersetzen und Dolmetschen in seinen verschiedensten Ausprägungen auch wie z. B. Softwarelokalisierung oder Schriftdolmetschen, sowie Schnittbereiche z. B. mit Barrierefreier Kommunikation, Fremdspracherwerb
- Algorithmische Modelle der Übersetzung können bis ins 9. Jhdt. zurückverfolgt werden, moderne Forschung zur Maschinellen Übersetzung (MÜ) beginnt nach dem 2. WK
- mehrsprachige Institutionen experimentieren schon in den 1950ern und 1960ern mit digitalen Ansätzen; die EU kauft Mitte der 1970er das erste MÜ-System und investiert anschließend massiv in die Forschung zu maschinellm Übersetzen und Dolmetschen
- spätestens ab den 2000ern wird der Übersetzer- und zunehmend auch Dolmetscherarbeitsplatz durchdigitalisiert: Terminologieverwaltung, Übersetzungsspeicher, Spracherkennungssoftware, ...
- seit Mitte/Ende der 2010er dominantes Paradigma der MÜ: „neuronale“ MÜ oder kurz NeuroMÜ, mit deutlichen Qualitätsverbesserungen gegenüber den Vorläufern



Vorteile der NeuroMÜ

- Daten werden produktiv nachgenutzt:
 - positiver psychologischer Effekt: Angst vor “dem weißen Blatt” wird gemindert (Macken et al. 2020)
 - wer eine sprachliche Grundkompetenz hat, kann, einfach gesagt, mit Hilfe von KI-Vorschlägen seine Kompetenz ausbauen (*Beurteilungskompetenz*)
→ mögliche Ergänzung zu Sprachlernapps
 - Effizienzgewinn bei der Übersetzung (allerdings niedriger als gedacht, wahrscheinlich max. ca. 20%, siehe Macken et al. 2020)
 - zur Systemerstellung nötige Daten eignen sich für weitere Anwendungen und sind zugleich ein Element der Sprachdokumentation
- durch sog. *zero shot*¹ können Systeme für Sprachenpaare erzeugt werden, für die es keine Trainingsdaten gibt: z.B. können ein englisch-japanisches und ein englisch-koreanisches MÜ-System zu einem japanisch-koreanischen MÜ-System zusammengeführt werden

¹ <https://research.google/blog/zero-shot-translation-with-googles-multilingual-neural-machine-translation-system/>



Probleme der (prädiktiven) NeuroMÜ

- großer Daten- und Energiehunger
- klebt sehr am Wortlaut des Ausgangstexts
 - in vielen, auch technischen Kontexten sind Umformulierungen, Hinzufügungen, Weglassungen üblich, um zielkulturellen Anforderungen zu genügen
- mehrsprachige Modelle haben aufgrund der Dominanz englischer Daten tendenziell einen „englischen Akzent“ (Papadimitriou et al. 2023)
 - Formulierungen z. B. in der deutschen Rechtssprache müssen aber sprachkonform sein; z. B. Verwendung des Strichpunkts, der im Englischen selten ist
- Systeme „raten“ semantische Näherungswerte, wo es keine eindeutigen Entsprechungen gibt
 - *Antrag auf Erteilung eines Wohnberechtigungsscheins* wird im Englischen zu *Application for a residence permit*, allerdings heißt *residence permit* auf Deutsch *Aufenthaltsgenehmigung* (getestet mit GoogleTranslate, 2024-05-25)
- verschiedene Unwuchten, z.B. *gender bias*
 - *clever pianist* wird zu *kluger Pianist*, *sexy pianist* wird zu *sexy Pianistin* (getestet mit GoogleTranslate, 2024-05-25)



Langzeitrisiken des (ungeprüften) KI-Einsatzes

- ! KI lernt Muster und neigt dazu, häufige Muster zu überproduzieren
→ Sprachverarmungshypothese (Toral 2019)
- ! Die Musterüberproduktion wird dadurch verstärkt, dass zunehmend KI-generierte Daten unterwegs sind; dies führt aufgrund des *Rekursionsfluchs* dauerhaft zum *Modellverfall* (Shumailov et al. 2023)
- ! Zudem kann die KI langfristig zum *Kompetenzverfall* führen: Das Sich-Verlassen auf KI-Lösungen schon in der Ausbildung mindert die Entwicklung *autonomer* Kompetenzen (z. B. Abbas et al. 2024)
- ! KI hat einen vergleichsweise hohen Energiebedarf und ist damit ökologisch belastend; nicht selten existieren einfachere, ebenso effiziente Lösungen
- ! KI kostet mehr als manche Betreiber zugeben; für Betreiber verlustreiche Lockangebote schaffen Abhängigkeiten, die später durch deutliche Preiserhöhungen für Kunden aufgefangen werden sollen



Manche haben die Risiken noch nicht erkannt ...

Aus dem Vitako/KGSt-Leitfaden zum Einsatz generativer KI im Verwaltungsalltag, S. 7
(https://vitako.de/wp-content/uploads/2024/09/2024-09-11_Generative_KI_in_Kommunalverwaltungen_Guideline_WEB.pdf)

- Der Grundsatz der Transparenz gilt nicht, wenn durch ein LLM-Tool kein Inhalt produziert wurde, sondern das LLM-Tool lediglich Korrekturen oder Sprachanpassungen vorgenommen hat. Beispiele sind Rechtschreibprüfungen oder Übersetzungen.

Wie funktioniert NeuroMÜ?



Wie wird ein Neuronales Netz trainiert?

→ Dr. Watson bei Youtube (https://youtu.be/ya_6l9IVMzY?t=154)



- Anstelle von Bildern und deren Einordnungskategorien werden *Parallelkorpora* fürs Training verwendet: Eingabe ist Ausgangssprachliche Satz, erwünschtes Ergebnis der Zielsprachliche Satz
- NeuroMÜ lässt sich gut in automatisierte, verteilte Abläufe integrieren und auf hohe Durchsatzmengen skalieren
- Maschinelles Dolmetschen, also eine direkte Übertragung von gesprochener zu gesprochener Sprache, noch in den Kinderschuhen
 - aktueller Prozess:
Spracherkennung → Maschinelle Übersetzung → Sprachausgabe
 - Nachteil: Fehlervererbung, d.h. ein Fehler an einer Stelle im Prozess kann zu Folgefehlern führen
- NeuroMÜ ist **kein** Wörterbuch: ein MÜ-System schlägt die als am wahrscheinlichsten errechnete Lösung vor (bei manchen Systemen kann man Alternativen auswählen), während ein *Wörterbuch* i.d.R. mehrere Lösungen vorschlägt
- Errechnung von Wortwahrscheinlichkeiten durch Sprachmodelle selbst ausprobieren:
<https://poloclub.github.io/transformer-explainer/>



Wichtige Daten: Parallelkorpora (1)

^ Externe Quellen (nicht geprüft)

[...] of Article 1 of the Convention of 27 September 1968 on
Jurisdiction and the Enforcement of Judgments in Civil and
Commercial [...]

⇒ eur-lex.europa.eu

Art. 1 Abs. 1 Satz 1 des Übereinkommens vom 27. September
1968 über die gerichtliche Zuständigkeit und die Vollstreckung
gerichtlicher Entscheidungen in [...]

⇒ eur-lex.europa.eu

[...] operated in a manner that complies with the insurance laws and
regulations applicable to its jurisdiction of organization; that
company provides procedures for processing claims that do not
permit [...]

⇒ knorr-venture-capital-group.de

[...] Vorschriften übereinstimmt, die für die Gerichtsbarkeit bzgl. der
Organisation gelten; dieses Unternehmen stellt Verfahren zur
Bearbeitung von Ansprüchen bereit, die es nicht erlauben, [...]

⇒ knorr-venture-capital-group.de

Article 16(1)(a) of the Convention of 27 September 1968 on
Jurisdiction and the Enforcement of Judgments in Civil and
Commercial [...]

⇒ eur-lex.europa.eu

[...] Übereinkommens vom 27. September 1968 über die
gerichtliche Zuständigkeit und die Vollstreckung gerichtlicher
Entscheidungen in [...]

⇒ eur-lex.europa.eu

[...] upset the institutional balance on which the division of
jurisdiction between OHIM and the Court is based (see, to that
effect, [...])

⇒ oami.europa.eu

[...] Gleichgewicht zuwiderliefe, das dem Grundsatz der
Zuständigkeitsverteilung zwischen dem HABM und dem Gericht
zugrunde liegt [...]

⇒ oami.europa.eu

[...] applicable in the United States, the European Union and the
jurisdiction in which Seller has its registered office or from which [...]

⇒ mobrey.de

[...] Vereinigten Staaten, der Europäischen Union und in der
Jurisdiktion gelten, in der sich der eingetragene Sitz des
Lieferanten [...]

⇒ mobrey.de

However, it also includes a regulation concerning the place of
jurisdiction (cf. Article 60 CIV 1980) and is therefore composed
following [...]

⇒ otif.org

Er umfaßt allerdings darüber hinaus auch eine
Zuständigkeitsregelung (vgl. Art. 60 CIV 1980) und ist somit ähnlich
wie Artikel [...]

⇒ otif.org

Dividendes of 16 February 2000 in Case T-122/00 Procter &

Mit Urteil vom 16. Februar 2000 in der Rechtssache T-122/00



Wichtige Daten: Parallelkorpora (2)

- *Alinierung* (d.h. Zuordnung von einander entsprechenden Segmenten wie Sätzen, Phrasen, Wörtern) läuft zu großen Teilen automatisiert (meist nach statistischen Algorithmen) ab
- hohe Zuverlässigkeit der Alinierung auch bei $1:m$ -Beziehungen oder $n:m$ -Beziehungen
- eine vergleichsweise günstige Technologie bzw. Technik
- hohe Zahl an Parallelkorpora bereits jetzt verfügbar, wobei Ausgangssprache überwiegend Englisch ist (Nachteil dadurch: ein deutsch-polnisches Parallelkorpus hat keine direkte Übersetzungsbeziehung, wenn Englisch Ausgangssprache für beide; Übersetzungsnuancen können verloren gehen)
- eigene Erstellung von Parallelkorpora stellt einen Ressourcenaufbau für spätere Wertschöpfung dar und eröffnet verschiedene Anwendungsmöglichkeiten; oft existieren in Organisationen schon so genannte *Translation Memories* (Übersetzungsspeicher oder TM, d. h. Datenbanken an bestehenden Übersetzungen)
- wo keine Parallelkorpora existieren, können einsprachige Korpora u.a. als Recherchemöglichkeiten aushelfen; auch diese sind wertvolle Ressourcen

Erkenntnis 1

**Hinter dem / in *KI*
stecken Menschen**

(mehrheitlich nicht Programmierer)

**Eigene Ressourcen
erstellen**



Alinierung: MateCat Aligner, Intertext (Bsp.)

Alignment project name (optional)

English US

French

+ Add Source file (or drop it here).

+ Add Target file (or drop it here).

Matecat Aligner supports [69 file formats](#)

START ALIGNING

<https://www.matecat.com/plugins/aligner/index#/>

naum-sen_otroka - InterText	
Alignment	Edit Search Options Help
M:	cs-00 ar-00
1	☆ ▶ Sen otroka
2	☆ ▶ Blízko jedné pyramidu si můj kamarád postavil chatu, na kterou občas chodíval, aby unikl městskému hluku, obvykle doprovázen manželkou a jedinou dcerkou, aby tak strávili víkend v klidu pouště.
3	☆ ▶ Několikrát jsem se omluvil, ale pak mě přesvědčil.
4	☆ ▶ Není to ani hodinka a jsi v úplně jiném světě."
5	☆ ▶ Rozprostřely se písky a obdivoval jsem, jak vytrvale pěstoval zahrádku.
6	☆ ▶ A pohled na pyramidu zblízka, no ten byl výjimečný.
7	☆ ▶ Jelikož věděl, že nemám rád život na samotě, začal vypočítávat přednosti toho místa:
8	☆ ▶ Čistý vzduch, klid mysli a čerstvá zelenina, která po velkém úsilí vyrostla, byl to opravdový pěstitelský úspěch.

<https://wanthalf.saga.cz/intertext>



Recherche, Termextraktion: SketchEngine (Bsp.)

DASHBOARD

Europarl UdS parallel, English



[Mehr Speicher](#)

EUROPARL UDS PARALLEL, ENGLISH

KORPUSINFORMATIONEN

KORPUS VERWALTEN



Word Sketch

Kollokationen und Wortkombinationen



Thesaurus

Synonyme und ähnliche Wörter



Parallelkonkordanz

Übersetzungen suchen



N-Gramme

Mehrwortlexeme



Trends

Diachrone Analyse, Neologismen



OneClick-Wörterbuch

Automatische Wörterbucherstellung



Word-Sketch-Vergleich

Kollokationen zweier Wörter vergleichen



Konkordanz

Verwendungsbeispiele im Kontext



Wortliste

Häufigkeitsliste



Schlüsselwörter

Terminologieextraktion



Texttypen-Analyse

Statistiken des gesamten Korpus



Zweisprachige Begriffspaare

Extraktion zweisprachiger Begriffspaare

KÜRZLICH VERWENDET

Europarl UdS parallel, Engli

VLVFK

Europarl UdS parallel, Gern

Firefox Sicherheit

< Le

An intensiv
lexicography
Málaga, Spair

R



Zum Mehrwert von Termdatenbanken und Korpora

- Eigene sprachlichen Datenressource: Kontrolle über **Eigentums- bzw. Nutzungsrechte** sowie über **Qualitäts- und Risikofaktoren**
- damit: höhere **Verlässlichkeit** und **Nachvollziehbarkeit**
- Aufbau einfacher Ressourcen wie z.B. von Glossaren vergleichsweise kostengünstig und mit erprobten Technologien möglich
- Erstellung einer Termdatenbank gibt u.a. Gelegenheit ...
 - den eigenen Wissensbestand terminologisch zu dokumentieren und
 - terminologische Bereinigungen durchzuführen.
- Kuratierung und Aufbereitung von Korpora ermöglicht u.a. ...
 - eine eigene Recherchequelle zu erstellen und
 - Daten (auch) für KI-Nutzungen zu erzeugen.

Erkenntnis 2

**Risikobehafteter Einsatz
von KI gehört in die
Hände von Menschen,
die wissen, was sie tun**



Risikomodell der Translation (1)

Gästeinformation eines kleinen Hotels

wahrscheinlich					
möglich	10				
entfernt vorstellbar		10			
unwahrscheinlich					
unmöglich			1	1	1
	Interne Kommunikation erschwert oder unmöglich	Imageschaden	Bedienung erschwert oder unmöglich	Gefahr von Sachschäden/ Vermögensschäden	Gefahr für Leib und Leben

Bedienungsanleitung einer CNC-Werkzeugmaschine

wahrscheinlich				50	50
möglich		10	50		
entfernt vorstellbar					
unwahrscheinlich					
unmöglich	1				
	Interne Kommunikation erschwert oder unmöglich	Imageschaden	Bedienung erschwert oder unmöglich	Gefahr von Sachschäden/ Vermögensschäden	Gefahr für Leib und Leben

- Text-in-Situation mit Risikoeinschätzung für alle Beteiligten und daraus folgende Prozessgestaltung (Canfora und Ottmann 2015; hier Abb. 7 und 9)
- auch davon inspirierte Nutzen-Risiken-Berechnungsmatrix der Digitalen Akademie der Wissenschaften¹ (Steller 2024)

1 <https://doi.org/10.5281/zenodo.13753399>



Risikomodell der Translation (2)

Categories	Title	
A - Legal documents		
A1	EU legal acts: treaties, regulation, directives, decisions, recommendations, opinions, international agreements	Full revision. Best possible combination of translator Even in situations with extreme work pressure, legal exception to this rule. For other legal documents exc standardized parts of legal acts or procurement docu can be better done by the Legal service or the reque Under extreme circumstances, the appropriate comb level of risk mitigation for revision combined with re For inter-institutional acts, if parts of a legal act has r
A2	Documents in administrative or legal proceedings and inquiries, such as infringements, mergers, anti-trust, state aid and anti-dumping cases (Procedural docs for CoJ) General Court and Civil Service Tribunal, national courts (mémoire)	
A3	Documents for procurement or funding programmes, tenders, grants applications, contracts	
A4	Recruitment notices, EPSO competition notices and test documents	
B - Policy and administrative documents		
B1	Accompanying documents not formally part of legal acts: staff working papers, COM working documents, (summaries of) impact assessments, as well as non-binding parts of legal acts such as financial statements or explanatory memoranda	Revision combined with review or just full review, pr matches very well with the text at hand. Exception: full revision. In situations with extreme work pressure • revise only key parts of the documents (ir • replace revision with review (if the compo
B2	White papers, green papers	
B3	Other official administrative documents flowing	

Generaldirektion Übersetzung der Europäischen Kommission: Kombination von Qualitäts- und Risikokriterien, Dokumentklassen und Prozessbeschreibungen mit Variablen wie Zeitdruck





Was kann ich nun also mit MÜ tun?

- KI-gestützte Übersetzung eignet sich
 - für die Integration in professionelle Prozesse
 - bei entsprechender Beurteilungskompetenz für die Findung und/oder Verifikation eigener Übersetzungsentscheidungen
 - ... in risikoarmen Kommunikationssituationen und im Notfall als Besser-als-nichts-Option, mit entsprechendem Bewusstsein für Fallstricke
- beim Einsatz von Technik:
 - ggf. mit Rückübersetzung arbeiten, um sich bzgl. Adäquatheit der Übersetzung abzusichern
 - ggf. auf Verkehrssprache Englisch ausweichen: die meisten Systeme arbeiten dominant mit englischen Daten, d.h. in Übersetzungsrichtungen „Englisch-von/nach-X“ ist die Qualität der MÜ tendenziell besser
 - bei der Abfrage einzelner Begriffe Wörterbücher und/oder Korpora nutzen: die MÜ bietet i.d.R. nur die als am wahrscheinlichsten errechnete Übersetzung als Lösung an, Wörterbücher und Korpora bieten i.d.R. mehrere mögliche Übersetzungen

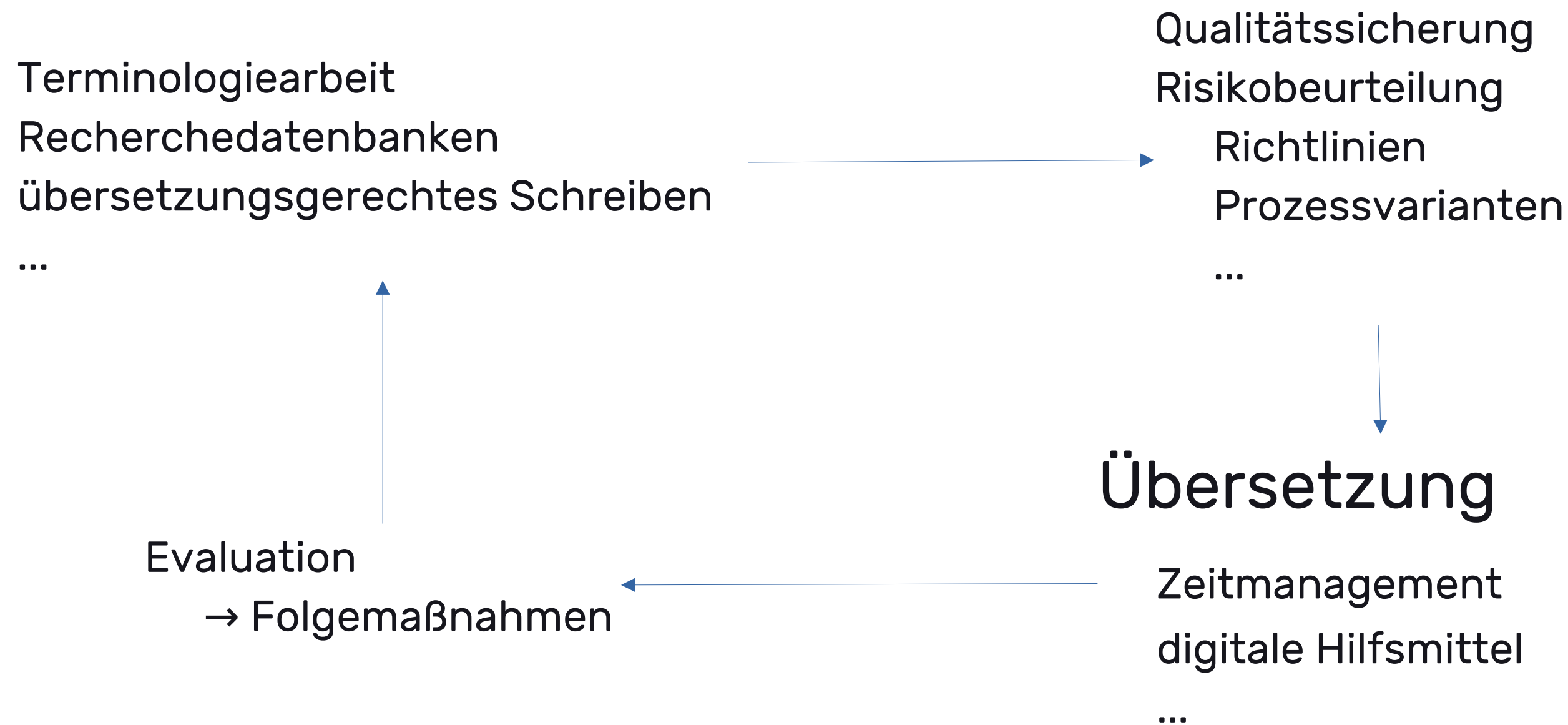
Erkenntnis 3

**Investitionen in
eigene Kompetenzen und
Ressourcen sind der beste
Garant für erfolgreiche
Mehrsprachigkeit**



Und das war noch nicht alles zum Thema!

- weitere Aspekte von Translationsprozessen nicht außer Acht lassen, hier ausschnittsweise beispielhaft zum Übersetzungsprozess:





Man muss nicht bei null anfangen :) Also ...

- erprobte Verfahren gibt es an verschiedenen Stellen von Verwaltung bereits: Hochschulen, oberste Bundesbehörden etc.
- eigene, souveräne Ressourcen
 - sind oft schon unerkannt vorhanden
 - lassen sich teils vergleichsweise kostengünstig aufbauen
 - unterstützen Wissensmanagement und Entwicklung digitaler Lösungen

→ Lösungsansätze sind vorhanden, sie müssen nur koordiniert und umgesetzt werden!



- Abbas, Muhammad, Farooq Ahmed Jam, und Tariq Iqbal Khan. 2024. „Is It Harmful or Helpful? Examining the Causes and Consequences of Generative AI Usage among University Students“. *International Journal of Educational Technology in Higher Education* 21 (1). <https://doi.org/10.1186/s41239-024-00444-7>.
- Canfora, Carmen, und Angelika Ottmann. 2015. „Risikomanagement für Übersetzungen“. *trans-kom* 8 (2): 314–46.
https://www.trans-kom.eu/bd08nr02/transkom_08_02_02_Canfora_Ottmann_Risiko.20151211.pdf
- Macken, Lieve, Daniel Prou, und Arda Tezcan. 2020. „Quantifying the Effect of Machine Translation in a High-Quality Human Translation Production Process“. *Informatics* 7 (2): 1–19.
<https://doi.org/10.3390/informatics7020012>.
- Papadimitriou, Isabel, Kezia Lopez, und Dan Jurafsky. 2023. „Multilingual BERT Has an Accent: Evaluating English Influences on Fluency in Multilingual Models“. In *Findings of the Association for Computational Linguistics: EACL 2023*, 1194–1200. Dubrovnik, Croatia: Association for Computational Linguistics.
<https://doi.org/10.18653/v1/2023.findings-eacl.89>.
- Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, und Ross Anderson. 2023. „The curse of recursion: Training on generated data makes models forget“. <https://doi.org/10.48550/arXiv.2305.17493>
- Steller, Jonatan Jalle. 2024. „KI × DH: Zum Umgang mit ML, Transformern und LLMs in der Digitalen Akademie“. <https://doi.org/10.5281/zenodo.13831622>



Fragen?

Kontakt

Oliver Czulo

iftra@mailbox.org

institut-translatologie.de



Oliver Czulo ist Professor in der Translatologie mit Spezialisierung in datenbasierter Semantik. Er studierte Computerlinguistik in Saarbrücken und promovierte dort in der Übersetzungswissenschaft. Der akademische Weg führte ihn auch durch Prag, Germersheim, Berkeley und Leipzig. Translatologie ist für ihn eine Wissenschaft der Verständigung.