

## Kapitel 6

# LLMs: ChatGPT und weitere KI-basierte Hilfsmittel für Übersetzende

Silvana Deilen<sup>a</sup>, Ekaterina Lapshinova-Koltunski<sup>a</sup> & Hellmut Riediger<sup>b</sup>

<sup>a</sup>Stiftung Universität Hildesheim <sup>b</sup>Civica Scuola Traduttori e Interpreti Altiero Spinelli Milano

In dem Beitrag wird die Anwendung von Large Language Models (LLMs) wie ChatGPT und Google Gemini in der Übersetzungspraxis beleuchtet. Dabei erfolgt zunächst eine allgemeine Einführung gefolgt von einer Erläuterung des theoretischen Hintergrunds der Funktionsweise von LLMs. Anschließend werden aktuelle Forschungsergebnisse anhand verschiedener Studien präsentiert, u.a. Evaluierungen von LLMs im Vergleich zu „klassischenmaschinellen Übersetzungssystemen wie DeepL oder Google Translate. Der Schwerpunkt liegt hierbei auf der Anwendung von LLMs in der Übersetzungspraxis und der didaktischen Nutzung in translationswissenschaftlichen Studiengängen. Wesentlich ist dabei die Vermittlung von Techniken zur Erstellung sinnvoller Prompts (Prompt-Engineering), anhand derer die von LLMs generierten Antworten verbessert werden können. Es wird untersucht, wie LLMs die Übersetzenden in verschiedenen Anwendungsbereichen, wie Pre-Editing, Post-Editing, Terminologearbeit, Qualitätskontrolle, Recherche, Integration von KI und LLMs in Übersetzungsumgebungen wie z. B. Trados unterstützen können. Dabei werden Praxiserfahrungen geschildert und Herausforderungen wie Gender Bias und die Gefahr von Halluzinationen bei der Verwendung von LLMs als Recherchewerkzeug aufgezeigt. Außerdem werden ethische und datenschutzrechtliche Aspekte beleuchtet und es wird die Frage diskutiert, ob die KI eine Gefährdung für den Übersetzerberuf darstellt. Dabei wird betont, dass trotz Veränderungen im Berufsbild, menschliche Übersetzer\*innen, obwohl sich ihre Rolle wandelt, als „expert-in-the-loop in sicherheitskritischen und spezialisierten Domänen weiterhin unverzichtbar sind.



## **1 Allgemeine Einführung**

Dieses Kapitel widmet sich der Anwendung von großen Sprachmodellen, d.h. Large Language Models (LLMs) im Bereich des Übersetzens. LLMs, die ursprünglich für Textgenerierung genutzt wurden, werden mittlerweile für verschiedene Aufgaben benutzt. Der Übersetzungsbereich, in dem LLMs nicht nur für automatische Übersetzung von Texten oder Sätzen, sondern auch für Terminologie- und Informationsrecherche eingesetzt wird, ist dabei keine Ausnahme. Allerdings ist das Potenzial von LLMs in diesem Bereich noch nicht vollständig erforscht. In diesem Kapitel liegt der Fokus auf der Qualität des Outputs, möglichen Halluzinationen sowie auf Möglichkeiten, den Output durch Abfragen (sog. Prompts) zu verbessern.

Neben allgemeinen Informationen über LLMs, ihren Grundlagen und ihrer Architektur wird in diesem Kapitel die Funktionsweise von LLMs dargestellt und auch allgemein auf das Thema Sprachmodellierung eingegangen. Darüber hinaus werden die aktuellen Forschungsarbeiten in diesem Bereich skizziert. Verschiedene Forschungsfragen und ihre Ergebnisse werden anhand zahlreicher Studien illustriert, wie die Evaluierungen von LLMs (z. B. ChatGPT) im Vergleich zu „klassischen“ maschinellen Übersetzungssystemen (z. B. DeepL oder Google Translate). Der Schwerpunkt liegt hierbei nicht auf den technischen Aspekten der Sprachmodellierung, sondern auf der Anwendung von LLMs in der Übersetzungspraxis. Außerdem wird auch die didaktische Nutzung in translationswissenschaftlichen Studiengängen angesprochen und diskutiert. Wesentlich ist dabei die Vermittlung von Techniken zur Erstellung sinnvoller Prompts (Prompt-Engineering). Diese helfen dabei, die von LLMs generierten Antworten zu verbessern. Es wird untersucht, wie LLMs die Übersetzenden in verschiedenen Anwendungsbereichen, wie Pre-Editing, Post-Editing, Terminologiarbeit, Qualitätskontrolle, Recherche sowie Integration von KI und LLMs in Übersetzungsumgebungen wie Trados, unterstützen können.

Zudem werden in diesem Kapitel auch Praxiserfahrungen dargestellt. Dabei werden u.a. Herausforderungen bei der Verwendung von LLMs als Rechercheinstrument, wie etwa Gender Bias oder Halluzinationen, illustriert. Im Anschluss werden auch ethische und datenschutzrechtliche Aspekte beleuchtet sowie die Frage nach der Gefährdung des Übersetzerberufs durch KI diskutiert. Es wird betont, dass trotz Veränderungen im Berufsbild, menschliche Übersetzer\*innen, obwohl sich ihre Rolle wandelt, als „expert-in-the-loop“ in sicherheitskritischen und spezialisierten Domänen weiterhin unverzichtbar sind.

## 2 Grundlagen der LLMs

### 2.1 Funktionsweise von LLMs

Den Begriff *Large Language Models (LLMs)* wurde im Bereich der maschinellen Sprachverarbeitung (*Natural language processing*, NLP) geprägt, um vortrainierte Sprachmodelle von erheblicher Größe und mit vielen Parametern von anderen vortrainierten Sprachmodellen zu unterscheiden. Technisch gesehen können Sprachmodelle die Sprachintelligenz von Maschinen und somit die künstliche Intelligenz vorantreiben. Ein Sprachmodell enthält die Informationen über die Wahrscheinlichkeiten von Wörtern und Wortkombinationen. Diese helfen dabei, die Wahrscheinlichkeiten folgender (oder fehlender) Wörter und Wortkombinationen vorherzusagen (Zhao u. a. 2023). Die Modelle basieren auf der Technik des Vortrainings (Zhou u. a. 2023). Ihr Ziel ist ein allgemeines Modell mit großen Datenmengen und Aufgaben zu trainieren, das in verschiedenen darauffolgenden Anwendungen leicht feinabgestimmt (*fine-tuned*) werden kann.

Für die vortrainierten Sprachmodelle werden riesige Korpora verwendet, um die Wahrscheinlichkeiten der Wörter und Wortkombinationen zu lernen. Die Korpora können dabei beliebig groß sein. Durch die Entwicklung dieser Technologie sind in den letzten Jahren mehrere vortrainierte Sprachmodelle entstanden, wie z. B. GPT, Gemini, Gemma, Llama, Claude, etc. Frühere Sprachmodelle wurden hauptsächlich für Textgenerierung eingesetzt, während neuere Sprachmodelle (z. B. GPT-4) für die Lösung komplexerer Aufgaben verwendet werden. Fast alle dieser Modelle basieren auf der sogenannten Transformerarchitektur (Vaswani u. a. 2023). Das ist eine neuronale Architektur mit den sogenannten Attention-Mechanismen (Aufmerksamkeitsmechanismen), die das Gewicht verwandter oder zusammenhängender Wörter hervorheben können und somit bessere Vorhersagen über die Wörter und Wortkombinationen machen können. Darüber hinaus kann eine Art von Eingabe in eine andere Art von Ausgabe automatisch umgewandelt werden.

Bei der Anwendung von vortrainierten Sprachmodellen für spezifische Aufgaben, wie zum Beispiel maschinelle Übersetzung, wird noch eine Reihe von Techniken eingesetzt, wie etwa Feinabstimmung (Finetuning) mit spezifischen Daten. Zum Beispiel wird für die Feinabstimmung für maschinelle Übersetzung (MÜ) ein zweisprachiges Parallelkorpus verwendet, z. B. Englisch-Deutsch für die automatische Übersetzung aus dem Englischen ins Deutsche. Die Trainingsdaten von LLMs sind größtenteils englischsprachig und enthalten nur einen geringen Anteil an Daten in anderen Sprachen. Deshalb sind Forschende bemüht, Möglichkeiten zu finden, die zuvor aufgezeigten Probleme zu lösen und die Leistung von LLMs für Übersetzungsaufgaben zu verbessern.

Beim Einsatz von LLMs unterscheidet man zwischen verschiedenen Prompting-Strategien, wie Zero-Shot-Prompting und Few-Shot-Prompting. Bei Zero-Shot-Prompting wird einem LLM eine Aufforderung gestellt, die in den Trainingsdaten des Modells nicht vorkommt. Solche Prompts enthalten üblicherweise spezifische Aufgabenanweisungen zusammen mit der Hauptabfrage. Zum Beispiel kann der Zero-Shot-Prompt für die Übersetzung aus dem Englischen ins Deutsche wie folgt aussehen: *„Übersetze den folgenden englischen Text ins Deutsche:..."*. Da LLMs gegenüber der Struktur und dem Inhalt von Prompts sehr empfindlich sind, sollten die Prompts bedacht formuliert werden. So könnte die Übersetzungsaufgabe oben alternativ auch anders formuliert werden: *„Du bist ein Übersetzer. Übersetze den folgenden englischen Text ins Deutsche. Ausgangstext:... Zieltext: " , usw.* Prompt-Engineering ist ein wichtiger Forschungsbereich, in dem es primär darum geht, Prompts zu verbessern.

Bei einem Few-Shot-Prompting werden zu erwartende Ausgabebeispiele direkt in die Aufforderung an die LLMs integriert. Diese Technik wird auch als kontextbezogenes Lernen bezeichnet. Die Beispiele, die dem Modell zur Verfügung gestellt werden, steuern und instruieren das Modell, die besten Antworten zu geben. Dies kann zur Verbesserung der Outputs von LLMs führen, wie es von Bawden & Yvon (2023) für maschinelle Übersetzung gezeigt wird. Außerdem zeigt der Beitrag, dass beim Zero-Shot-Prompting oft Übergenerierung oder Generierung in der falschen Sprache beobachtet wird. Few-Shot-Prompting verbessert die Ergebnisse für eine Reihe von Sprachpaaren. Prompt-Engineering und die Einstellung verschiedener Parameter für bessere Ergebnisse im Bereich der maschinellen Übersetzung werden derzeit aktiv untersucht (siehe z. B. Vilar u. a. 2023, Zhang u. a. 2023).

## 2.2 Abgrenzungen und Konvergenzen zwischen der regulären NMT und LLMs

Neural Machine Translation (NMT) und LLMs basieren beide auf großen neuronalen Netzwerken, unterscheiden sich jedoch in ihrer Trainingsmethodik. NMT-Modelle werden auf zweisprachigen Übersetzungskorpora trainiert, d.h. direkt auf das Übersetzen gedrillt, jedoch ohne Berücksichtigung des sprachlichen Kontextes. Neuronale maschinelle Übersetzungssysteme sind daher besonders geeignet für spezialisierte Anwendungsgebiete, in denen Genauigkeit, Konsistenz und Datensicherheit entscheidend sind. Dazu gehören Fachtexte, Unternehmensdokumente, datenschutzsensitive Inhalte sowie weitere komplexe Dokumente, also ausgangssprachlich orientierte Übersetzungen. LLMs dagegen werden zunächst auf Millionen von mono- und multilingualen, aber nicht unbedingt parallelen

(alignierten) Texten vortrainiert. Gewissermaßen lernen sie zuerst die Sprachen, zwischen denen sie dann auch übersetzen können. LLMs können also auch übersetzen, obwohl sie nicht direkt für die Übersetzung trainiert wurden. Darüber hinaus können sie aber auch Texte in den verschiedensten Genres und Stilen verfassen, analysieren, zusammenfassen, korrigieren und auf Konsistenz prüfen oder hinsichtlich Sprachregister oder Textgattung umformulieren. Sie können Terminologie extrahieren und anwenden sowie Stilrichtlinien befolgen. Dies ermöglicht LLMs flüssigere, kreativere und kontextuell reichhaltigere Texte zu erzeugen, jedoch oft mit weniger Genauigkeit bei spezialisierten Inhalten. Die Wahl zwischen NMT und LLM hängt stark von den spezifischen Anforderungen der Übersetzungsaufgabe ab.

Der Trend geht jedoch in Richtung der Integration von LLMs in andere Übersetzungstechnologien. So wie schon NMT-Systeme, wird es allmählich möglich sein, auch LLMs in CAT-Tools wie RWS Trados einzubinden, sodass die menschlich überarbeiteten Übersetzungen für weitere Projekte in Translation Memories gespeichert werden können.

Unterdessen sind auch Übersetzungs-LLMs wie Unbabel oder Tower auf den Markt gebracht worden, sowie Tools wie Wordscope, eine Übersetzungssoftware, welche die integrierte Nutzung von verschiedenen NMT-Systemen, ChatGPT, eigenen Translation Memories und Termbases ermöglicht.

### 3 Aktueller Forschungsstand

Im Folgenden wird eine Übersicht über den aktuellen Forschungsstand gegeben. Da die Anzahl an Veröffentlichungen zum Thema LLMs in der Translationswissenschaft aufgrund der aktuellen Relevanz enorm ist und rapide steigt, wird nur auf die Studien eingegangen, die sich konkret mit der Integration von LLMs in die übersetzerische Praxis befassen. Einen umfassenden Überblick über die Performance von LLMs als MÜ-Systeme bietet Krüger (2023).

#### 3.1 Anwendungsbereiche

Krüger (2023) beschäftigt sich mit der praktischen Anwendbarkeit von KI-Technologien in der Übersetzungsbranche und zeigt auf, wie die Arbeitsabläufe durch LLMs automatisiert und optimiert werden können. Er veranschaulicht, dass LLMs nicht nur als MÜ-Tools, sondern auch zur automatischen Evaluation der MÜ-Qualität eingesetzt werden können (Krüger 2023: 35–36). So veranschaulicht er, dass ChatGPT zur Berechnung verschiedener Scores wie der *translation*

*edit rate* (TER) und folglich auch zur *machine translation quality evaluation* (MTQE) verwendet werden kann. Zudem können LLMs problematische Stellen im Text kennzeichnen, die dann entweder von professionellen Übersetzer\*innen oder aber vom Modell selbst korrigiert werden können. Ebenso können LLMs einen Text nach kundenspezifischen Vorgaben und Styleguides anfertigen oder anpassen sowie zur automatischen Terminologiekxtraktion und Terminologearbeit genutzt werden. Außerdem könnten autonome LLMs laut Krüger sogar Managementaufgaben übernehmen, wie Termine oder Übersetzungsprojekte koordinieren. Hier ist es denkbar, dass in Zukunft alle mit einem Übersetzungsprojekt einhergehenden Aufgaben und eingesetzten Tools in einem zentralen LLM zusammengefügt werden. Übersetzende werden dabei vermehrt die Rolle von „experts-in-the-loop“ (Krüger 2023: 39) einnehmen und die Performance des LLMs überwachen, im Falle von Defiziten eingreifen sowie für die Kommunikation zwischen Kunden und LLM zuständig sein. Im Hinblick auf die benötigten Kompetenzen für die Nutzung von KI führt Krüger (2023) darüber hinaus den Ausdruck *artificial intelligence literacy* (KI Kompetenz) ein und zieht Parallelen zu den Konzepten *professional MT literacy* (professionelle MÜ-Kompetenz) und dem daraus entwickelten *MT-oriented data literacy framework* (MÜ-orientierter Datenkompetenzrahmen).

So vielversprechend die potentiellen Anwendungsbereiche auch klingen mögen, so ernüchternd ist teilweise die Realität. Reineke (2023) ist der Frage nachgegangen, für welche Tätigkeiten KI-gestützte Chatbots im Bereich der Terminologearbeit eingesetzt werden können. Er zeigt auf, dass ChatGPT zwar prinzipiell Definitionen und Quellen für Fachtermini ausgeben kann, die vom Tool angegebenen Quellen jedoch häufig auf nicht existierende Webseiten verweisen. Ebenso wie bei Anfragen zu Definitionen und Benennungen sind die angezeigten Ergebnisse auch bei der Frage nach Begriffssystemen nicht stabil. Er schlussfolgert, dass ChatGPT „für eine erste Annäherung an die Begriffssystematik eines Fachgebietes sicherlich ein nützliches Werkzeug [ist]“ (Reineke 2023: 27), aber selbst vermeintlich brauchbare Ergebnisse mit Vorsicht zu genießen sind. Die Tatsache, dass die Antworten auf identische Fragen erheblich voneinander abweichen und angegebene Quellen nicht vorhanden sind, widerspricht den Grundsätzen der Terminologearbeit, die sich durch Eineindeutigkeit und Konsistenz auszeichnet.

### 3.2 Performance von LLMs

In vielen der bislang durchgeführten Studien wurde die Performance von LLMs als MÜ-Tool evaluiert und mit der Performance von konventionellen MÜ-Systemen verglichen. So haben Hendy u. a. (2023) für 18 Sprachrichtungen

die Performance von drei GPT-Modellen evaluiert und mit state-of-the-art NMT-Systemen verglichen. Die Evaluation, die sowohl aus einer Humanevaluation als auch aus automatischen Evaluationsmetriken bestand, ergab, dass die Performance für *High-Resource Languages* mit der Performance von NMT-Systemen vergleichbar ist und durch Few-Shot-Prompting weiter verbessert werden kann. Bei *Low-Resource Languages* lag die Qualität jedoch unter der Qualität von NMT-Systemen.

Dieses Ergebnis wurde in mehreren Studien repliziert: So haben Moslem u. a. (2023) ebenfalls gezeigt, dass die Performance von LLMs für *High-Resource Languages* sehr vielversprechend ist, wohingegen dies bei *Low-Resource Languages* nicht der Fall ist. Darüber hinaus untersuchten sie, inwiefern der MÜ-Output von LLMs (insbesondere GPT-3.5) durch In-Context-Learning verbessert werden kann. Ihre Untersuchung ergab, dass das Few-Shot-Prompting mit Fuzzy Matches (Übersetzungsbeispiele, bei denen Segmente des Ausgangstextes eine hohe Ähnlichkeit zu den Segmenten des zu übersetzenden Zieltext aufweisen) zu besseren Ergebnissen führte als das Few-Shot-Prompting mit zufälligen Beispielen (d.h. Beispiele, die keine Ähnlichkeit zum Zieltext aufweisen). Übersetzungen ohne Prompting (Zero-Shot-Translation) schnitten am schlechtesten ab (Moslem u. a. 2023: 229). Zudem zeigten sie auf, dass beim Inkludieren von Glossareinträgen im Prompt die Terminologie häufiger erfolgreich in den Text übernommen wurde als es bei Zero-Shot- und Few-Shot-Übersetzungen ohne das Bereitstellen von Terminologie der Fall war (Moslem u. a. 2023: 232). Der Vergleich von GPT-3.5 und GPT-4 hat zudem gezeigt, dass mit GPT-4 zwar eine bessere Zero-Shot-Übersetzungsqualität erzielt wurde, die Qualität von Few-Shot-Übersetzungen jedoch mit beiden Modellen ähnlich war (Moslem u. a. 2023: 233).

Raunak u. a. (2023) untersuchten, ob GPT-4 für automatisches Post-Editing von MÜ-Output eingesetzt werden kann. Hierfür forderten sie das Tool auf, Verbesserungsvorschläge für eine vorgefertigte Übersetzung zu liefern und anschließend den post-editierten Zieltext zu produzieren. Sie kamen zu dem Ergebnis, dass die post-editierte Version von GPT-4 deutlich besser war als die nicht post-editierte Version. Dies galt sowohl für Übersetzungen ins Englische (DE-EN, ZH-EN) als auch für Übersetzungen in andere Zielsprachen (EN-DE, EN-ZH). Darüber hinaus untersuchten sie, wie viele der Verbesserungsvorschläge tatsächlich vom Tool in die finale Zieltextversion eingearbeitet wurden. Sie zeigten, dass GPT-3.5 turbo deutlich weniger Vorschläge auch tatsächlich implementiert als GPT-4 und schlussfolgern, dass GPT-4 in dieser Hinsicht vertrauenswürdiger ist. Auch Jiao u. a. (2023) kamen zu dem Schluss, dass die Übersetzungsqualität durch den Einsatz von GPT-4 deutlich verbessert werden kann, und das auch für

Sprachen aus unterschiedlichen Sprachfamilien. Dass GPT-4 mit anderen state-of-the-art NMT-Systemen mithalten und teilweise sogar besseren Output liefern kann, zeigten Manakhimova u. a. (2023) auch für die Sprachenpaare DE-EN und EN-DE.

Interessante Ergebnisse zum Einsatz von LLMs lieferten auch die Studien von Popovic u. a. (2024) und Castilho u. a. (2023). Popovic u. a. (2024) untersuchten die Auswirkungen von fehlerhaften Ausgangstexten (z. B. Texte mit Rechtschreib-, Grammatik-, oder Ausdrucksfehlern) auf die Qualität von maschinellen Übersetzungen. Hierbei gingen sie der Frage nach, ob die Fehler im Zieltext übernommen oder korrigiert werden. Bei den Ausgangstexten handelte es sich um Amazon-Rezensionen, die aus dem Englischen ins Kroatische, Russische und Finnische übersetzt wurden, und zwar von konventionellen MÜ-Systemen und ChatGPT. Die Studie ergab, dass die Fehler von ChatGPT deutlich häufiger korrigiert wurden als von konventionellen MT-Systemen (Popovic u. a. 2024: 22). Dennoch bereiten insbesondere Rechtschreibfehler und informelle Konstruktionen sowohl ChatGPT als auch MÜ-Systemen Probleme und führen häufig zu Übersetzungsfehlern (z. B. falsche Übersetzung, Auslassung etc.). Insgesamt kamen sie zu dem Ergebnis, dass ChatGPT in Hinblick auf die Tolerierung von Fehlern im Ausgangstext „robuster“ (Popovic u. a. 2024: 24) ist als konventionelle MÜ-Systeme.

Castilho u. a. (2023) untersuchten hingegen, wie MÜ-Systeme mit kontextbezogenen Übersetzungsproblemen wie lexikalischer Ambiguität, grammatischem Geschlecht, Referenzen, Ellipsen und Terminologie umgehen, wenn die Lösung für dieses Problem in einer längeren Kontextspanne gegeben wird. Die Studie wurde mit Sprachen aus unterschiedlichen Sprachfamilien durchgeführt, und zwar mit deutschen, portugiesischen, irischen und chinesischen Texten. Der Vergleich von konventionellen NMT-Systemen (Google Translate, DeepL und Microsoft Bing) und ChatGPT-3 zeigte, dass ChatGPT zwar insgesamt besser abschnitt; die Performance für das Irische jedoch unabhängig von der auftretenden Übersetzungsherausforderung häufig schlechter war als die Performance der NMT-Systeme. Angesichts dessen, dass Irisch zu den *Low-Resource Languages* zählt, stimmt dieses Ergebnis ebenfalls wieder mit dem Ergebnis von Hendy u. a. (2023) überein. Zudem zeigte die Studie, dass die Position des Kontextes (vor oder nach dem problematischen Wort) keinen Einfluss auf die Ergebnisse hatte.

Darüber hinaus wurde auch bereits untersucht, inwiefern ChatGPT auch im Bereich der intralingualen Übersetzung eingesetzt werden kann. Deilen u. a. (2023) ließen 30 juristisch-administrative Texte von ChatGPT aus dem Standarddeutschen in Leichte Sprache übersetzen. Die Analyse des Outputs, in welcher Korrektheit des Inhalts, Verständlichkeit und syntaktische Komplexität evaluiert



wurden, ergab allerdings, dass die von ChatGPT erstellten Texte zwar leichter als die Standardtexte, jedoch nachwievor nicht leicht genug waren und zudem die gängigen Leichte-Sprache-Regeln nicht eingehalten wurden. Zudem wurde der Inhalt nicht immer korrekt wiedergegeben.

### 3.3 Auswirkungen auf das Berufsbild und die übersetzerische Ausbildung

Lee (2024) beschäftigt sich mit der Frage, welche Auswirkungen die technologischen Entwicklungen und damit einhergehend der „Paradigmenwechsel“ (Lee 2024: 13) auf das Berufsbild von Übersetzer\*innen haben werden. Es besteht Konsens darüber, dass LLMs Übersetzer\*innen (noch) nicht ersetzen können, da sie nachwievor fehleranfällig und nicht zuverlässig genug sind und darüber hinaus keine „ethische Verantwortung“ (Lee 2024: 12) übernehmen können. Beim Einsatz von MÜ-Tools wie DeepL nimmt der Übersetzende heutzutage vermehrt die Rolle des Post-Editors ein. ChatGPT hingegen kann – aufgrund der Möglichkeit des Promptings – nicht nur als Übersetzer, sondern zugleich auch als Post-Editor fungieren. Das wiederum hat zur Folge, dass Humanübersetzer\*innen nicht von Übersetzer\*innen zu Post-Editor\*innen werden, sondern eher zu „Post-Post-Editor\*innen“ (Lee 2024: 13). Eine wichtige Kompetenz von Übersetzer\*innen ist somit neben der Post-Editing- auch die Prompt-Engineering-Kompetenz. Der immer weiter voranschreitende technologische Wandel und Paradigmenwechsel kann schließlich dazu führen, dass Übersetzer\*innen alles machen außer übersetzen (Lee 2024: 13). Lee (2024) geht davon aus, dass Übersetzer\*innen vermehrt die Rolle einer zweisprachigen Berater\*in oder Projektmanager\*in einnehmen werden. Übersetzen wird somit vermehrt zu einer *cross-cultural-communication* bzw. einem *interpersonal and interactional event* und somit Teil der interkulturellen Kommunikation. Übersetzende benötigen hierfür ein komplexes, umfangreiches Repertoire an Fähigkeiten, u. a. logistische, verwaltungstechnische und zwischenmenschliche Fähigkeiten. Aufgrund dessen stellt Lee (2024) auch in Frage, ob die derzeit recht statischen Kriterien zur Bewertung von Übersetzungsklausuren in der Übersetzerausbildung überhaupt noch zeitgemäß sind. Er plädiert neben der Anpassung von Prüfungsformen auch dafür, die neu erforderlichen Kompetenzen in die Studieninhalte und somit in das „Toolkit“ (Lee 2024: 17) der Übersetzer\*innen zu integrieren. Als Kernkompetenzen nennt er Recherchekompetenzen, also das Wissen darüber, wie man die Zuverlässigkeit von Informationen und Informationsquellen bewertet, technologische Kompetenzen, d.h. die Arbeit mit digitalen Texten und deren Empfang, Bearbeitung und Übermittlung an Kunden, sowie Dienstleistungskompetenzen, d.h. die Abwicklung

eines Übersetzungsauftrags.

Die Potentiale von LLMs für die Übersetzungsbranche werden auch von Siu (2023) untersucht. Als Vorteile von ChatGPT gegenüber konventionellen MÜ-Systemen nennen sie die Fähigkeit, die Bedeutung eines Ausdrucks in unterschiedlichen Kontexten zu erklären, was bei polysemen Wörtern, die je nach Kontext eine andere Bedeutung haben, relevant ist. Darüber hinaus ist es möglich, sich kulturspezifische Ausdrücke und fachspezifische Termini erklären zu lassen. Im Bereich des Fachtextübersetzens kann es zudem hilfreich sein, sich eine vereinfachte Textzusammenfassung ausgeben zu lassen. Im Unterschied zu konventionellen MÜ-Tools ist es auch möglich, sich mehrere Übersetzungsvorschläge generieren zu lassen, so dass der Text nicht nur als *draft* fungiert, sondern dem Übersetzenden zugleich auch als Inspiration dient. Weitere Anwendungsmöglichkeiten sind Fehlererkennung, Grammatikprüfung und Revision. Diese Vorteile von ChatGPT gegenüber anderen MÜ-Tools sind insbesondere auf die Fähigkeit der interaktiven Übersetzung zurückzuführen. Die Nutzer\*innen arbeiten nicht mit einem finalen vom Tool generierten Output, sondern können sich durch den Dialogmodus den Output erneut vom Tool überarbeiten und je nach spezifischen Vorgaben anpassen lassen. Mit diesen einzelnen Instruktionen kann das Produkt somit *step by step* verbessert werden.

## 4 Anwendungen und Erfahrungen in Didaktik

Im folgenden Kapitel wird näher auf Erfahrungen im Umgang mit LLMs im Übersetzungsunterricht und der Übersetzungspraxis eingegangen. Dies wird an einem praktischen Beispiel aus dem Unterricht illustriert. Zwischen Herbst 2023 und Frühling 2024 wurden am Mailänder Institut für Dolmetschen und Übersetzen Altiero Spinelli mit knapp 30 Studierenden aus Masterstudiengängen und 10 Dozierenden mit unterschiedlichen Sprachkombinationen Workshops zur Nutzung von LLMs als Hilfsmittel beim Übersetzen organisiert.

Zunächst erhielten die Teilnehmer\*innen (Studierende und Dozierende) eine allgemeine theoretische Einführung und es wurden die Grundregeln des Prompt-Engineering vermittelt:

1. Klare Definition der Aufgabe.
2. Komplexe Aufgaben in einfachere Unteraufgaben unterteilen.
3. Sich nicht mit dem ersten Ergebnis zufrieden geben und die Unterschiede ständig überprüfen.

4. Dem Chatbot so viel Kontext wie möglich und Zeit zum „Denken“ geben.
5. Sich vom Chatbot helfen lassen, das Prompt-Engineering zu verbessern.

Sie wurden aufgefordert sich mit den Anleitungen von OpenAI<sup>1</sup> und von RWS Group Limited<sup>2</sup> vertraut zu machen. Danach sollten sich die verschiedenen Gruppen innerhalb ihrer Kurse einer oder mehrerer der folgenden Aufgaben widmen:

- Entwicklung eines Repertoires von Beispielen wie LLMs zur Verbesserung des Übersetzungsprozesses genutzt werden können.
- Erstellung einer Prompt-Bibliothek mit allgemein nützlichen Prompts.
- Entwicklung von Übungs-, Unterrichts- und Prüfungsmaterialien.
- Erstellung eines Leitfadens zur Nutzung von LLMs im Übersetzungsprozess.
- Beurteilung des Nutzens von LLMs im Übersetzungsprozess.

Die Teilnehmer\*innen sollten die frei zugänglichen LLMs wie ChatGPT, Gemini oder Perplexity AI befragen, um herauszufinden, wie diese sie in ihrer Tätigkeit unterstützen können (als Übersetzungsprofis, -studierende bzw. -dozierende). Auf die Frage „Wie kannst du Übersetzende in ihrer Tätigkeit unterstützen?“ gab ChatGPT z. B. folgende Antworten:

1. Automatisierte Übersetzungen: Schnelle Vorübersetzungen.
2. Terminologie und Konsistenz: Terminologie-Extraktion und Erstellung von Glossaren.
3. Qualitätssicherung: Korrekturlesen und stilistische Anpassungen.
4. Kulturelle und kontextuelle Anpassungen: Kulturelle Anpassungen und Kontextanalysen.
5. Recherche und Informationsbeschaffung: Hintergrundrecherche und Synonymvorschläge.
6. Projektmanagement: Zeiteffizienz und Textanalysen.
7. Lernen und Weiterbildung: Sprachpraxis und Musterbeispiele.

---

<sup>1</sup><https://platform.openai.com/docs/guides/prompt-engineering>

<sup>2</sup>Introduction to Linguistic Prompt Design Training for Everyone - Linguistic Prompt Design Training for Everyone (rws.com)

Studierende und Dozierende wurden aufgefordert mit den in ihren Kursen behandelten Texten einige der genannten Kompetenzen im Übersetzungsprozess zu testen. In einem Abschlusstreffen sollten die Teilnehmenden den Nutzen der LLMs für verschiedene Zwecke spontan evaluieren.

Zusammenfassend wurden LLMs als reine Übersetzungstools ähnlich wie NMT-Systeme betrachtet, obwohl sie tendenziell mehr „Halluzinationen“ aufweisen. Besonders geschätzt wurden die LLMs jedoch als Hilfsmittel zur Textanalyse vor der Übersetzung, als lexikalisch-terminologische Verständnishilfe (Wörterbuchfunktion), zur Erstellung von Zusammenfassungen, als Korrektur- bzw. Post-Editing-Tool, zur Anpassung von Sprachregistern oder Terminologie sowie zur Vereinfachung von Texten. Sie wurden auch als nützlich für die Erstellung von Taxonomien und als erste Anlaufstelle für Hintergrundinformationen (z. B. als Wikipedia-Ersatz) angesehen. Eine Schwachstelle ist die geringe Zuverlässigkeit bzw. Transparenz der Quellen, auf die sich die Outputs stützen. Einige Teilnehmende betonten jedoch, dass Perplexity AI in dieser Hinsicht überlegen sei. Die Dozierenden fanden LLMs auch nützlich als Hilfsmittel für die Erstellung von Präsentationen, zur Einholung von Tipps zur Kursgestaltung, zur Erstellung von Übungen und Texten zur Fokussierung auf bestimmte Merkmale bzw. Probleme und als Hilfsmittel bei der Korrektur.

Was die Leistungsnachweise betrifft, besteht Konsens darüber, dass traditionelle Übersetzungsaufgaben weitgehend überholt sind. Der Trend verlagert sich von der Beurteilung des Endprodukts hin zur kritischen Beurteilung des Übersetzungsprozesses, bzw. den verschiedenen Prozessschritten.

Tabelle 1 illustriert die Beispiele von Prompts, die in verschiedenen Phasen des Übersetzungsprozesses (wie z. B. Textanalyse, Vorgabe von Terminologie, usw.) integriert werden können.

In Beispiel (1) wird auch ein konkretes Beispiel für die Verbesserung einer Übersetzung durch detaillierte Prompts gegeben. Im ersten Prompt wird das LLM aufgefordert, den Ausgangstext zu übersetzen. Im zweiten Prompt werden dem Sprachmodell noch weitere Restriktionen und Anweisungen gegeben. Aus den in den Outputs illustrierten Texten wird deutlich, dass das detaillierte Prompting zu einer verbesserten Übersetzung führt. Das deutet darauf hin, dass die Informationen in Prompts eine große Rolle spielen und dass die Ausformulierung von geeigneten Prompts für den Übersetzungsprozess sehr wichtig ist.

(1) Verbesserung einer Übersetzung mittels Prompts

**Ausgangstext** *Tradition, die man schmeckt. Seit 1885 werden im Weingut Hallenbach, nun schon in der fünften Generation, edle Weine geschaffen. Durch diese lange kellerwirtschaftliche Tradition erfuhr das Weingut*

Tabelle 1: Prompt-Beispiele nach verschiedenen Phasen des Übersetzungsprozesses

Phasen des Übersetzungsprozesses	Mögliche Prompts
1. Textanalyse	Was sind die sprachlichen, inhaltlichen und kommunikativen Merkmale? Was ist der Zweck des Textes? Welche Probleme könnten bei der Übersetzung auftreten?
2. Allgemeine Übersetzung	Übersetze den Text [Ausgangstext] in [Zielsprache].
3. Adaption der Übersetzung	Übersetze [Quelltext] in [Zielsprache]. Der zu übersetzende Text ist [Textart]. Das Zielpublikum der Übersetzung sind [Zielgruppe]. Der Tonfall der Übersetzung sollte [Tonfall/Sprache] sein. Er erscheint in [Medium].
4. Revision	Führe eine Revision der Übersetzung durch, um sicherzustellen, dass sie den ursprünglichen Text genau widerspiegelt. Ausgangstext: [...] Übersetzung: [...]
5. Bewertung	Bewerte die folgende Übersetzung, die als Leistungsnachweis in einem Übersetzungskurs erstellt wurde, nach den folgenden Kriterien: 1) Bedeutung der Übersetzung 2) Grammatik 3) Stil und Flüssigkeit 4) Wortschatz 5) Kohärenz, Kohäsion und innere Logik Ausgangstext: [...] Zu bewertende Übersetzung: [...]
6. Erkennung von Übersetzungss Stolpersteinen	Erstelle eine Tabelle der Wörter und Ausdrücke im folgenden Text, die schwer zu übersetzen sind. Gib die folgenden Informationen an: Begriff, Wortart, Satz aus dem Text [...]
7. Vorgabe von Terminologie/Lexik	Verwende diese Wortpaare, um die Übersetzungsaufgaben zu optimieren: [Wort Ausgangssprache] = [Wort Zielsprache] [Wort Ausgangssprache] = [Wort Zielsprache]

*überregionale Anerkennung. Wir bewirtschaften ausschließlich schieferdurchsetzte Steilhänge in den besten Lagen von Zell. Das Weingut umfasst insgesamt 4,5 Hektar Weinberge, in denen hauptsächlich Riesling, jedoch auch Spätburgunder und Dornfelder angebaut werden. Der Riesling, die edelste Weißweinsorte der Welt, bildet den Grundstock dieser feinen Weine.*

**Prompt 1** Übersetze den Text ins Englische.

**Output 1** *Tradition You Can Taste. Since 1885, noble wines have been created at the Hallenbach Winery, now in its fifth generation. Through this long tradition of cellar management, the winery has gained recognition beyond the region. We exclusively cultivate slate-infused steep slopes in the best locations of Zell. The winery comprises a total of 4.5 hectares of vineyards, primarily growing Riesling, as well as Spätburgunder and Dornfelder. Riesling, the noblest white wine variety in the world, forms the foundation of these fine wines.*

**Prompt 2** Übersetze denselben Text ins Englische unter

Berücksichtigung der folgenden Aspekte:

- Es handelt sich um einen Marketingtext für eine Website.
- Lokalisere die englische Übersetzung für nicht-deutsche Leser.
- „Weingut Hallenbach“ ist der Name des Unternehmens: übersetze ihn nicht ins Englische.
- Sei kreativ in Bezug auf Wortwahl und Satzstruktur.
- Du musst der deutschen Syntax nicht strikt folgen.
- Wichtig ist, dass die Bedeutung des ursprünglichen Textes genau ins Englische übertragen wird.
- Vermeide nominale Sätze.
- Verwende kürzere Sätze.
- Bevorzuge Verben anstelle von Nomen, wo möglich.
- Berücksichtige SEO<sup>3</sup>.

**Output 2** *Savor Tradition Since 1885. For over 130 years, Weingut Hallenbach has crafted premium wines, passed down through five generations. This deep-rooted tradition has earned us recognition well beyond our region. We focus on cultivating steep, slate-rich slopes in Zell's top vineyards. Our 4.5 hectares are primarily planted with Riesling, the world's most celebrated white grape, alongside select Spätburgunder and Dornfelder vines. At Weingut Hallenbach, Riesling*

---

<sup>3</sup>SEO=Search Engine Optimisation

*stands at the core of our fine wines, delivering unmatched flavor and elegance. Discover the essence of tradition in every sip.*

Die zweite Version zeigt eine deutlich bessere Anpassung an ein internationales, nicht-deutschsprachiges Publikum. Sie ist optimiert für das Web, mit einem flüssigeren und ansprechenderen Stil. Die Satzstruktur ist klarer und weniger verschachtelt. Zudem wurde der Titel kreativ angepasst, um Aufmerksamkeit zu erzeugen. SEO-relevante Schlüsselwörter wie „Riesling“ und „Weingut Hallenbach“ wurden hervorgehoben, während unnötige Füllwörter vermieden wurden.

## 5 Fazit

Im Fazit sollen, basierend auf den obigen Ausführungen, die Auswirkungen für die Übersetzer\*innenausbildung abgeleitet werden.

In vielen Übersetzungsabteilungen wird diskutiert, wie der Umgang mit LLMs in die Lehrpläne integriert werden sollte und ob LLMs in Prüfungen und Abschlussarbeiten erlaubt sein sollten. Darüber hinaus stellt sich die Frage, wie sie richtig eingesetzt werden können, um den Übersetzungsprozess zu unterstützen und die Fähigkeiten von Studierenden zu verbessern. Wie schon nach dem Erscheinen der NMT-Systeme, erzeugen LLMs gemischte Gefühle in der Didaktik und in der Praxis: Verwunderung und Begeisterung, aber auch Skepsis, Desorientierung und sogar Angst. Denn einerseits können Produktivität und Qualität verbessert werden, andererseits besteht aber das Risiko, dass bestimmte menschliche Tätigkeiten, insbesondere viele „klassische Übersetzungsarbeiten“, überflüssig werden. Schon vor dem Durchbruch von LLMs wurde beobachtet, wie sich die Rollen, Aufgaben und Kompetenzen in der Übersetzungsbranche verändern. Anstatt *from scratch* zu übersetzen, gehen die Aufgaben mehr und mehr in Richtung Adaption, Post-Editing, Qualitätskontrolle und Gutachten, Transkreation, Sprachdatenmanagement, Testen von Tools oder Beratung für mehrsprachige Kommunikation (Bernardini & Miličević Petrović 2021, Massey u. a. 2022 und Riediger & Galati 2023) Der Umgang mit LLMs ist mittlerweile zu einer dieser neuen Kompetenzen geworden.

Die Nutzung von LLMs erfordert somit technologische sowie sprachliche, fachliche und kulturelle Kenntnisse. Erfahrungen zeigen, dass der Einsatz von LLMs sowie der von MÜ eine Anleitung der Nutzenden erfordert, um Missbrauch und Fehlanwendungen zu vermeiden. Daher sollten Studierende frühzeitig in einem reflektierten Umgang mit diesen Technologien geschult werden. Der Einsatz von

LLMs sowie von MÜ erfordert somit eine gezielte Anleitung durch Lehrkräfte, um einen effektiven Lerneffekt zu erzielen.

KI und MÜ sollten daher nicht nur „erlaubt oder verboten“ werden, sondern aktiv in die Lehre integriert werden, um Missbrauch und negative Gewohnheiten zu vermeiden. Eine mögliche Herausforderung besteht nämlich darin, dass Übersetzer\*innen sich zu sehr auf LLMs verlassen könnten, was zu einer Abnahme der eigenen Übersetzungsfähigkeiten führen kann.

Studierende sollten bereits auf Bachelor-Niveau und hier nicht nur in Technologiekursen, sondern vielmehr in allen Übersetzungskursen zu einem reflektierten Umgang mit MÜ und KI angeleitet werden.

## Literatur

- Bawden, Rachel & François Yvon. 2023. Investigating the translation performance of a large multilingual language model: The case of BLOOM. In Mary Nurminen, Judith Brenner, Maarit Koponen, Sirkku Latomaa, Mikhail Mikhailov, Frederike Schierl, Tharindu Ranasinghe, Eva Vanmassenhove, Sergi Alvarez Vidal, Nora Aranberri, Mara Nunziatini, Carla Parra Escartín, Mikel Forcada, Maja Popovic, Carolina Scarton & Helena Moniz (Hrsg.), *Proceedings of the 24th annual Conference of the European Association for Machine Translation*, 157–170. Tampere, Finland: European Association for Machine Translation. <https://aclanthology.org/2023.eamt-1.16>.
- Bernardini, Silvia & Maja Miličević Petrović. 2021. *Toward a new profile for twenty-first century language specialists: Industry, institutional and academic insights. UPSKILLS Intellectual Output 1.2*. Techn. Ber. Zenodo. DOI: 10.5281/zenodo.5030873.
- Castilho, Sheila, Clodagh Quinn Mallon, Rahel Meister & Shengya Yue. 2023. Do online machine translation systems care for context? What about a GPT model? In Mary Nurminen, Judith Brenner, Maarit Koponen, Sirkku Latomaa, Mikhail Mikhailov, Frederike Schierl, Tharindu Ranasinghe, Eva Vanmassenhove, Sergi Alvarez Vidal, Nora Aranberri, Mara Nunziatini, Carla Parra Escartín, Mikel Forcada, Maja Popovic, Carolina Scarton & Helena Moniz (Hrsg.), *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, 393–417. Tampere, Finland: European Association for Machine Translation. <https://aclanthology.org/2023.eamt-1.39>.



- Deilen, Silvana, Sergio Hernández Garrido, Ekaterina Lapshinova-Koltunski & Christiane Maaß. 2023. Using ChatGPT as a CAT tool in Easy Language translation. In Sanja Štajner, Horacio Saggio, Matthew Shardlow & Fernando Alva-Manchego (Hrsg.), *Proceedings of the second workshop on text simplification, accessibility and readability*, 1–10. Varna, Bulgaria: INCOMA Ltd., Shoumen, Bulgaria. <https://aclanthology.org/2023.tsar-1.1>.
- Hendy, Amr, Mohamed Abdelrehim, Amr Sharaf, Vikas Raunak, Mohamed Gabr, Hitokazu Matsushita, Young Jin Kim, Mohamed Afify & Hany Hassan Awadalla. 2023. *How good are GPT models at machine translation? A comprehensive evaluation*. DOI: 10.48550/arXiv.2302.09210.
- Jiao, Wenxiang, Wenxuan Wang, Jen-tse Huang, Xing Wang, Shuming Shi & Zhaopeng Tu. 2023. *Is ChatGPT a good translator? Yes with GPT-4 as the engine*. DOI: 10.48550/arXiv.2301.08745.
- Krüger, Ralph. 2023. Artificial intelligence literacy for the language industry – with particular emphasis on recent large language models such as GPT-4. *Lebende Sprachen* 68(2). 283–330. DOI: 10.1515/les-2023-0024. (27 Mai, 2024).
- Lee, Tong King. 2024. Artificial intelligence and posthumanist translation: ChatGPT versus the translator. *Applied Linguistics Review* 15(6). 2351–2372. DOI: <https://doi.org/10.1515/applirev-2023-0122>.
- Manakhimova, Shushen, Eleftherios Avramidis, Vivien Macketanz, Ekaterina Lapshinova-Koltunski, Sergei Bagdasarov & Sebastian Möller. 2023. Linguistically motivated evaluation of the 2023 state-of-the-art machine translation: Can ChatGPT outperform NMT? In Philipp Koehn, Barry Haddow, Tom Kocmi & Christof Monz (Hrsg.), *Proceedings of the eighth conference on machine translation*, 224–245. Singapore: Association for Computational Linguistics. DOI: 10.18653/v1/2023.wmt-1.23.
- Massey, Gary, Elsa Huertas-Barros & David Katan (Hrsg.). 2022. *The human translator in the 2020s*. London: Routledge. DOI: 10.4324/9781003223344.
- Moslem, Yasmin, Rejwanul Haque, John D Kelleher & Andy Way. 2023. Adaptive machine translation with large language models. *arXiv preprint arXiv:2301.13294*.
- Popovic, Maja, Ekaterina Lapshinova-Koltunski & Maarit Koponen. 2024. Effects of different types of noise in user-generated reviews on human and machine translations including ChatGPT. In Rob van der Goot, JinYeong Bak, Max Müller-Eberstein, Wei Xu, Alan Ritter & Tim Baldwin (Hrsg.), *Proceedings of the Ninth Workshop on Noisy and User-generated Text (w-NUT 2024)*, 17–30. San Ġiljan, Malta: Association for Computational Linguistics. <https://aclanthology.org/2024.wnut-1.3>.

- Raunak, Vikas, Amr Sharaf, Hany Hassan Awadallah & Arul Menezes. 2023. Leveraging GPT-4 for automatic translation post-editing. *arXiv preprint arXiv:2305.14878*.
- Reineke, Detlef. 2023. Terminologearbeit mit ChatGPT & Co. *edition 1*. 25–27.
- Riediger, Hellmut & Gabriele Galati. 2023. La traduzione nell'era dell'IA: Nuovi ruoli, nuove competenze, nuova formazione. *mediAzioni* 39. A35–A54. DOI: 10.6092/issn.1974-4382/18786.
- Siu, Sai Cheong. 2023. ChatGPT and GPT-4 for professional translators: Exploring the potential of large language models in translation. *SSRN*. DOI: 10.2139/ssrn.4448091.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser & Illia Polosukhin. 2023. *Attention Is All You Need*. <https://arxiv.org/abs/1706.03762>.
- Vilar, David, Markus Freitag, Colin Cherry, Jiaming Luo, Viresh Ratnakar & George Foster. 2023. Prompting PaLM for translation: Assessing strategies and performance. In Anna Rogers, Jordan Boyd-Graber & Naoaki Okazaki (Hrsg.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 15406–15427. Toronto, Canada: Association for Computational Linguistics. DOI: 10.18653/v1/2023.acl-long.859.
- Zhang, Biao, Barry Haddow & Alexandra Birch. 2023. *Prompting large language model for machine translation: A case study*. arXiv: 2301.07069 [cs.CL]. <https://arxiv.org/abs/2301.07069>.
- Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie & Ji-Rong Wen. 2023. *A survey of large language models*. <https://arxiv.org/abs/2303.18223>.
- Zhou, Ce, Qian Li, Chen Li, Jun Yu, Yixin Liu, Guangjing Wang, Kai Zhang, Cheng Ji, Qiben Yan, Lifang He, Hao Peng, Jianxin Li, Jia Wu, Ziwei Liu, Pengtao Xie, Caiming Xiong, Jian Pei, Philip S. Yu & Lichao Sun. 2023. *A comprehensive survey on pretrained foundation models: A history from BERT to ChatGPT*. <https://arxiv.org/abs/2302.09419>.