

Kann ChatGPT komplexe wissenschaftsjournalistische Texte verständlich machen?

Fehler bei der automatischen Übersetzung mit ChatGPT in einfache Sprache

Clarissa Elisabeth
Hohenwalde
Department für
Wissenschaftskommunikation
Karlsruher Institut für Technologie
Karlsruhe, Deutschland
clarissa.staudt@kit.edu

Michael Wahl
Überwachungsstelle des Bundes für
Barrierefreiheit von
Informationstechnik
Berlin, Deutschland
michael.wahl@
bfit-bund.de

Markus Lehmkuhl
Department für
Wissenschaftskommunikation
Karlsruher Institut für Technologie
Karlsruhe, Deutschland
markus.lehmkuhl@kit.edu

Abstract

Diese Studie untersucht die Fähigkeit von ChatGPT-4-Turbo, wissenschaftsjournalistische Texte in einfache Sprache zu übersetzen und bewertet die Fehleranfälligkeit dieser Übersetzungen. 20 Artikel aus „Spektrum der Wissenschaft“ wurden automatisch in einfache Sprache übersetzt.

Die durchschnittliche Artikellänge reduzierte sich dabei von 3.530 auf 2.140 Zeichen. Die Lesbarkeit der Original- und vereinfachten Texte wurde mittels der Wiener Sachtextformel verglichen, wobei die vereinfachten Texte deutlich verständlicher waren. Eine inhaltsanalytische Untersuchung identifizierte im Mittel 8,65 Fehler pro Text – hauptsächlich semantische sowie einige syntaktische und morphologische Fehler. Typische Übersetzungsfehler umfassten Bedeutungsverschiebungen, Auslassungen wichtiger Informationen, Falschinformationen und fehlerhafte Umformungen direkter oder indirekter Rede. Empfehlungen der DIN 8581-1-Norm für einfache Sprache wurden weitgehend eingehalten, Schwierigkeiten traten nur bei der Vermeidung von Konjunktiv, bildhafter Sprache und der durchgängigen Verwendung gleicher Begriffe auf.

Die Ergebnisse zeigen, dass ChatGPT Texte vereinfachen kann, jedoch besteht Verbesserungspotential bei der semantischen Präzision. Zukünftige Untersuchungen sollen klären, wie die Übersetzungsfehler das Textverständnis beeinflussen.

Keywords

Einfache Sprache, Digitale Barrierefreiheit, Verständlichkeit, Wissenschaftsjournalismus, Large Language Models, ChatGPT

1 Einleitung

Ziel des Austauschforums Digitale Barrierefreiheit auf der MuC 2024 ist es, die Verbindung zwischen Barrierefreiheit und Künstlicher Intelligenz zu explorieren. Dieser Beitrag analysiert, ob ChatGPT komplexe wissenschaftsjournalistische Texte verständlich machen kann und welche Fehler bei der Übersetzung in einfache Sprache auftreten.

2 Hintergrund

Wissenschaftsjournalistische Inhalte sind für die Meinungsbildung und Entscheidungsfindung im Alltag wichtig. Für viele Menschen ist wissenschaftliches Wissen jedoch schwer zugänglich, weshalb gute Wissenschaftskommunikation essentiell ist. Entsprechend lautet eine häufig formulierte normative Anforderung, dass Wissenschaftskommunikation „die Breite der Gesellschaft und die gesamte Bildungskette ab[decken soll]“ (BMBF, 2021).

In der Praxis werden jedoch viele Gruppen systematisch ausgeschlossen: Repräsentative Studien wie das Wissenschaftsbarometer (2023) zeigen, dass Personen mit geringerem Bildungsniveau weniger über Wissenschaftsthemen informiert sind und sich oft von Forschenden unzureichend aufgeklärt fühlen. Hinzu kommt, dass über die Hälfte der Erwachsenen in Deutschland eine geringe Lesekompetenz besitzt und längere, komplexe Texte nicht versteht (OECD, 2013). Einfache Sprache, die an den Bedürfnissen der Lesenden orientiert ist, kann helfen, diese Menschen zu erreichen (Baumert, 2019). Textvarianten in unterschiedlichen Schwierigkeitsgraden werden jedoch in der journalistischen Praxis nur selten angeboten.

Der Fortschritt bei großen Sprachmodellen wie ChatGPT bietet nun neue Möglichkeiten, große Textmengen schnell und kostengünstig zu verarbeiten. Wenn ChatGPT komplexe wissenschaftsjournalistische Texte verständlich machen kann, könnten Medienorganisationen künftig automatische zielgruppenangemessene Übersetzungen in einfache Sprache auf Knopfdruck bereitstellen.

3 Material und Methodik

Um zu evaluieren, ob ChatGPT wissenschaftsjournalistische Texte in einfache Sprache übersetzen kann, wurden 20 Texte aus der populärwissenschaftlichen Monatszeitschrift Spektrum der Wissenschaft ausgewählt. Die Artikel, die zwischen 2019 und 2024 in der Kategorie „News“ erschienen und online frei zugänglich sind, wurden mittels des Chat-GPT-4-Turbo-Modells in einfache Sprache übersetzt. Das mehrstufige Verfahren

orientiert sich an der DIN 8581-1-Norm (2024) zur einfachen Sprache und umfasste im ersten Schritt die Erfassung der Autor*innenziele und der für den Lesenden relevanten Informationen. Basierend auf diesen Daten wurde ChatGPT aufgefordert, auf den Originaltext zunächst auf Text-, dann Satz- und schließlich auf Wortebene die wichtigsten Regeln einfacher Sprache laut der DIN-Norm anzuwenden (siehe Prompts im Anhang). Zuletzt sollte der Chatbot finale Korrekturen durchführen und auf einen journalistischen Stil achten. Während die Originaltexte im Mittel 3.530 Zeichen enthielten, umfassten die Übersetzungen in einfacher Sprache im Schnitt 2.140 Zeichen.

Die Lesbarkeit der Originaltexte und der vereinfachten Texte wurde mit der ersten Wiener Sachtextformel verglichen. Anschließend erfolgte eine inhaltsanalytische Untersuchung durch zwei Codierer*innen, die gemeinsam alle Texte sichteten und Auffälligkeiten sowie Fehler herausarbeiteten. Das Kategoriensystem entstand durch eine Kombination aus deduktivem und induktivem Vorgehen, wobei die Hauptkategorien der Fehler (Semantik, Syntax und Morphologie) sowie die Kategorie „Verstoß gegen die Empfehlungen für einfache Sprache“ vorab definiert wurden.

4 Ergebnisse

ChatGPT konnte die Lesbarkeit der Texte stark verbessern, nämlich von einem mittleren Wert von 2,2 entsprechend der ersten Wiener Sachtextformel zu 1,1 bei den vereinfachten Texten. Im Mittel wurden bei der Inhaltsanalyse 8,65 Fehler pro Text festgestellt. Beispiele typischer Fehler finden sich im Anhang.

Semantische Fehler: Im Mittelwert traten pro Text 7,4 semantische Fehler auf. Diese Fehler beinhalteten u.a. die Verschiebung der Bedeutung durch die Wahl unpassender Synonyme oder Phrasen. Dies führte meist zu einer zu starken Verallgemeinerung von Aussagen. Ferner fehlten in den vereinfachten Texten häufig wichtige Inhalte, es wurden Falschaussagen gemacht, es wurde auf im vereinfachten Text weggelassene Teile referenziert, es gab Kausalitätsfehler und unpassende bildhafte Vergleiche.

Syntaktische Fehler: Es wurden im Mittel 0,65 Fehler pro Text festgestellt. Hauptsächlich betrafen diese den Umgang mit direkter und indirekter Rede, die oft nicht korrekt gekennzeichnet wurde. Zudem gab es vereinzelt Probleme mit der Satzstellung und der Kommasetzung.

Morphologiefehler: Morphologiefehler traten im Mittel 0,6 Mal pro Artikel auf. Diese Fehler betrafen Bereiche wie die Groß- und Kleinschreibung, Genus, Kasus, Tempus, Modus, Numerus oder die Bildung von Komposita.

Verstoß gegen die Empfehlungen für einfache Sprache: Verstöße gegen die in der DIN 8581-1-Norm (2024) ausgesprochenen Empfehlungen zur einfachen Sprache traten im Schnitt 1,75 Mal pro Artikel auf. Häufig gelang es ChatGPT nicht, den Konjunktiv und bildhafte Sprache zu vermeiden,

durchgängig gleiche Begriffe zu verwenden oder geläufigere bzw. kürzere Wörter zu finden.

5 Diskussion

Die Ergebnisse zeigen, dass ChatGPT Texte insgesamt verständlicher machen kann. Ebenso gelingt es weitgehend, die Empfehlungen für einfache Sprache umzusetzen. Jedoch gilt: „Die einfache Sprache gibt es nicht“ (Baumert, 2019). Sind den Lesenden Metaphern bekannt, so kann deren gezielter Einsatz in journalistischen Texten das Textverständnis sogar fördern statt behindern.

Weiterhin ist festzustellen, dass ChatGPT überwiegend morphologisch und syntaktisch korrekte Texte produziert, aber noch Verbesserungsbedarf im Umgang mit direkter und indirekter Rede aufweist. Auf semantischer Ebene treten jedoch häufig Fehler auf, die zu einer Unschärfe in der Bedeutung führen. Dies ist vermutlich darauf zurückzuführen, dass große Sprachmodelle mit Wortvektoren arbeiten, bei denen Synonyme nah beieinander liegen, obwohl sie feine Bedeutungsunterschiede aufweisen. Die Ergebnisse verdeutlichen auch, dass noch Arbeit im Bereich des AI-Alignments notwendig ist, insbesondere wenn es darum geht, korrekte Informationen zu liefern und Halluzinationen zu vermeiden.

Obwohl die Anzahl der Fehler auf den ersten Blick hoch erscheinen mag, handelt es sich häufig um sprachliche Feinheiten. Es bleibt zu untersuchen, ob diese Unsauberkeiten tatsächlich zu Verständnisproblemen führen und wie die Zielgruppe die vereinfachten Texte wahrnimmt. Diese Fragen streben wir an, in einem zukünftigen Laborexperiment zu untersuchen.

Referenzen

- [1] Baumert, A. (2019). Mit einfacher Sprache Wissenschaft kommunizieren. Springer Spektrum.
- [2] BMBF, Bundesministerium für Bildung und Forschung (2021). Wissenschaftskommunikation in der Projektförderung. https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/1/668936_Wissenschaftskommunikation_in_der_Projektforderung.pdf?__blob=publicationFile&v=5
- [3] DIN (2024). DIN 8581-1:2024-05: Einfache Sprache - Anwendung für das Deutsche - Teil 1: Sprachspezifische Festlegungen. <https://dx.doi.org/10.31030/3523268>
- [4] OECD, Organisation for Economic Co-operation and Development. (2013). OECD skills outlook (2013): First results from the survey of adult skills. Paris: OECD Publishing.
- [5] WiD, Wissenschaft im Dialog (2023). Wissenschaftsbarometer. https://wissenschaft-im-dialog.de/documents/47/WiD-Wissenschaftsbarometer2023_Broschuere_web.pdf

Anhang

In diesem Abschnitt werden die für ChatGPT verwendeten Prompts sowie Beispiele für häufige Fehlertypen vorgestellt. Um die Transparenz der Arbeit zu erhöhen und eine Reproduktion der Analysen sicherzustellen, wird das Datenmaterial, der Code und die Codierung öffentlich zugänglich gemacht:

<https://github.com/clarissastaudt/Fehler-bei-der-automatischen-Textvereinfachung-mit-ChatGPT.git>

Prompts

Die DIN 8581-1-Norm zur einfachen Sprache arbeitet heraus, dass „die Inhalte eines Textes [...] gebrauchsfreundlich und konsequent auf die Zielgruppe ausgerichtet sein [müssen]“ (DIN, 2024). Neben den Zielen und Erwartungen des Lesenden sind die Ziele der Autor*innen zu berücksichtigen. Folglich wird ChatGPT im ersten Prompt beauftragt, die Autor*innenziele und die für den Lesenden relevante Informationen herauszuarbeiten. Diese Daten sind der Grundstein für alle weiteren Schritte der Übersetzung in einfache Sprache.

Du bist ein herausragender Wissenschaftsjournalist. Welche Ziele verfolgt der Autor mit diesem Text und was sind die für den Leser relevantesten Aussagen? Halte dich kurz.

{TEXT}

In der DIN 8581-1-Norm (2024) werden verschiedene Ebenen unterschieden. Für das Prompting zur Übersetzung in einfache Sprache wird ChatGPT angewiesen, besonderen Wert auf ausgewählte Regeln zur Text-, Satz- und Wortebene zu legen. Um ein besseres Endergebnis zu erzielen, werden diese Schritte nacheinander ausgeführt, wobei stets die Autor*innenziele und für den Lesenden relevante Informationen beibehalten werden sollten.

Textebene

Du bist ein Experte für einfache Sprache und ein herausragender Wissenschaftsjournalist. Formuliere den Text so um, dass er einfach zu verstehen ist. Behalte den journalistischen Stil bei.

Beachte insbesondere folgende Regeln:

Textebene

1. Füge bei Bedarf zusätzliche Informationen und Beispiele ein.
Schlecht: Mobilität
Besser: Mobilität bedeutet, wie Menschen sich fortbewegen, zum Beispiel zu Fuß, mit dem Skateboard, Auto, Bus, Zug oder Flugzeug.
2. Nutze die direkte Anrede.
Schlecht: Der Antragsteller sollte das Dokument gründlich lesen.
Besser: Bitte lesen Sie das Dokument gründlich.
3. Vermeide Nominalstil.
Schlecht: Von Seiten des Gremiums erfolgte keine Beteiligung an dieser Untersuchung.
Besser: Das Gremium beteiligte sich nicht an dieser Untersuchung.
4. Gliedere den Text in sinnvolle Abschnitte.
5. Verwende kurze Sätze.

6. Verwende für längere Aufzählungen nummerierte Listen.
Schlecht: Ins Freibad sollte man ein Handtuch, einen Badeanzug oder eine Badehose und Sonnencreme mitnehmen.
Besser: Ins Freibad sollte man Folgendes mitnehmen:
– Handtuch;
– Badeanzug oder Badehose;
– Sonnencreme.

Beachte, dass folgende Informationen im finalen Text enthalten sein sollten:

{INFOS}

Text:

{TEXT}

Satzebene

Du bist ein Experte für einfache Sprache und ein herausragender Wissenschaftsjournalist. Formuliere den Text so um, dass er einfach zu verstehen ist. Behalte den journalistischen Stil bei.

Beachte insbesondere folgende Regeln:
Satzebene

1. Nutze möglichst kurze Sätze. Kein Satz darf mehr als 15 Wörtern haben. Variiere die Satzlänge.

2. Benutze eine einfache Satzstruktur aus Subjekt – Prädikat – Objekt. Ein Satz darf maximal einen Nebensatz haben.
Schlecht: Der Zug kommt, wenn er keine Verspätung hat, um 12 Uhr an.
Besser: Der Zug kommt um 12 Uhr an, wenn er keine Verspätung hat.
3. Verbalklammer und Nominalklammer.
Zwischen den Teilen eines Prädikats und einem Artikel oder Pronomen und dem dazugehörige Substantiv sollten nicht zu viele Wörter stehen.
Schlecht: Irina hat Michael auf das Ende der Einreichungsfrist für den Antrag am 30. April hingewiesen.
Besser: Irina hat Michael gesagt, dass er den Antrag bis zum 30. April einreichen muss.

4. Benutze keine Verneinungen.
Schlecht: Sie haben uns die Angaben noch nicht geschickt.
Besser: Bitte schicken Sie uns noch die fehlenden Angaben.

5. Texte sollen im Präsens geschrieben werden. Wenn Vergangenheitsformen notwendig sind, sollte das Perfekt genutzt werden.
Schlecht: Sie schrieben, lasen und beteten.
Besser: Sie haben geschrieben, gelesen und gebetet.

6. Vermeide Konjunktiv. Der Konjunktiv I darf nicht verwendet werden.
Schlecht: Anna sagte, sie wolle morgen kommen. Sie habe sich extra freigegeben.
Besser: Anna sagte, sie will morgen kommen. Sie sagte, dass sie sich extra freigegeben hat.

7. Ersetze Passivkonstruktionen durch Aktiv.
Schlecht: Es wurde eine Untersuchung seitens der Behörde durchgeführt.
Besser: Die Behörde führte eine Untersuchung durch.

Beachte, dass folgende Informationen im finalen Text enthalten sein sollten:

{INFOS}

Text:

{TEXT}

Wortebene

""

Du bist ein Experte für einfache Sprache und ein herausragender Wissenschaftsjournalist. Formuliere den Text so um, dass er einfach zu verstehen ist. Behalte den journalistischen Stil bei.

Beachte insbesondere folgende Regeln:

Wortebene

1. Benutze immer die gleichen Begriffe. Vermeide Synonyme.
 Schlecht: Heute gehen alle Gläubigen in die Kirche, denn Ostern ist ein wichtiges Fest im Kirchenjahr. Im Gotteshaus singen die frommen Besucher Lieder.
 Besser: Heute gehen alle Gläubigen in die Kirche, denn Ostern ist ein wichtiges Fest im Kirchenjahr. In der Kirche singen die Gläubigen Lieder.

2. Verwende konkrete und eindeutige Wörter.

Schlecht: Werkzeug

Besser: Hammer

3. Benutze kurze und gebräuchliche Wörter.

Schlecht: Wägeinheit

Besser: Waage

4. Vermeide Substantivierungen.

Schlecht: Das Öffnen der Verschlüsse ermöglicht die Wartung des Schmiernippels.

Besser: Öffnen Sie die Verschlüsse, um die Schmiernippel zu warten.

5. Vermeide Funktionsverbgefüge.

Schlecht: zur Sprache bringen

Besser: sagen

6. Vermeide Metaphern und Allegorien.

Schlecht: Eine Flut von Menschen strömt zum Stadion.

Besser: Sehr viele Menschen gehen zum Stadion.

7. Vermeide Fremdwörter. Wenn du trotzdem Fremdwörter benutzt, erkläre sie.

Schlecht: Evaluation

Besser: Auswertung

8. Löse Komposita auf.

Schlecht: Infektionsschutzmaßnahmenverordnung

Besser: Verordnung über Maßnahmen zum Schutz vor Infektionen

Beachte, dass folgende Informationen im finalen Text enthalten sein sollten:

...

{INFOS}

...

Text:

...

{TEXT}

...

""

Nach der automatischen Vereinfachung des Textes erfolgen in einem letzten Schritt Korrekturen, um der situativen Textgattung Rechenschaft zu tragen und den Textfluss zu verbessern.

""

Du bist ein Experte für einfache Sprache und ein herausragender Wissenschaftsjournalist. Formuliere den Text so um, dass er den Regeln der einfachen Sprache folgt und einfach zu verstehen ist. Behalte den journalistischen Stil bei. Beginne nicht mit "Liebe Leserinnen und Leser". Ende nicht mit einer Ansprache an die Leser oder einem Dank an die Leser.

Beachte, dass folgende Informationen im finalen Text enthalten sein sollten:

...

{INFOS}

...

Text:

...

{TEXT}

...

""

Beispiele für häufige Fehlertypen

Im Rahmen der Inhaltsanalyse wurden Morphologie-, Syntax- und Semantikfehler unterschieden.

Morphologiefehler

Die Morphologie befasst sich mit der Struktur und Form von Wörtern. Sie umfasst die Regeln, durch die Wörter gebildet, verändert und angepasst werden, wie etwa die Flexion (Veränderung der Wortform) oder die Wortbildung (Schaffung neuer Wörter durch Zusammensetzung oder Ableitung). Morphologiefehler beziehen sich somit beispielsweise auf die fehlerhafte Bildung von Genus, Kasus, Tempus, Modus, Numerus, Komposita und die Groß- und Kleinschreibung.

Vereinfachter Text	Fehlerbeschreibung	Fehlertyp
Sie bemerken kaum jemand	Korrekt wäre: kaum jemanden	Genus
das Anti-Elektron-Neutrino.	Korrekt wäre: Elektron-Antineutrino	Komposita-bildung
Wissenschaftler haben jetzt eine Hefe erschaffen, die zu einem großen Teil aus künstlichen Erbgut besteht.	Korrekt wäre: künstlichem	Kasus

Syntaxfehler

Die Syntax befasst sich mit dem Satzbau innerhalb einer Sprache, also sprachlichen Strukturen, die über die Wortebene hinausgehen. Syntaktische Fehler betreffen etwa die fehlerhafte Kennzeichnung direkter und indirekter Rede, die Satzstellung oder die Kommasetzung.

Vereinfachter Text	Fehlerbeschreibung	Fehlertyp
"Die KI lernt viel, indem sie viele Bilder betrachtet. Sie merkt sich, wie eine gesunde Netzhaut auszusehen hat."	Das Originalzitat lautet: »Im Zuge der Analyse von Millionen Bildern lernt das Modell, die typischen Merkmale einer Netzhaut zu erkennen«. Somit findet eine falsche Zitation statt.	falsche oder fehlende Kennzeichnung direkter oder indirekter Rede
Aber man sollte noch nicht zu früh sich freuen.	Korrekt wäre: Aber man sollte sich noch nicht zu früh freuen.	Satzstellung

Semantikfehler

Häufig wird die Semantik als Bedeutungslehre bezeichnet. Entsprechend handelt es sich um semantische Fehler, wenn es zur Verschiebung der Bedeutung durch die Wahl unpassender Synonyme oder Phrasen kommt, wichtige Aussagen ausgelassen werden, Falschaussagen getroffen werden, sich fälschlicherweise auf im vereinfachten Text weggelassene Teile bezogen wird, Kausalitätsfehler auftreten oder unpassende bildhafte Vergleiche verwendet werden.

Vereinfachter Text	Fehlerbeschreibung	Fehlertyp
Das ist bei vielen anderen Computerprogrammen nicht so.	Im Original geht es nur um KI-Anwendungen, nicht Computerprogramme im Allgemeinen. Der vereinfachte Text nimmt eine Übergeneralisierung vor, wodurch sich die Bedeutung verschiebt.	Bedeutungsver-schiebung
Die Netzhaut ist sehr interessant.	Im Original geht es darum, dass eine Untersuchung der Netzhaut interessant ist, um Rückschlüsse über die Gesundheit zu treffen. Es fehlt also eine wichtige Aussage.	wichtige Aussage fehlt
Wir müssen den Boden heilen, damit wir sauberes Wasser und nährstoffreiche Pflanzen haben.	Im Originaltext wird Trinkwasser nicht thematisiert, somit handelt es sich um eine Falschaussage.	Falschaussage
Es geht dabei um viel mehr als um Bier.	Zuvor wurde in dem in einfache Sprache übersetzten Text nur von Hefe im Allgemeinen, jedoch nicht von Bierhefe gesprochen. Somit bleibt der Bezug dieses Satzes unklar.	nicht definierte Referenz

Hilfsmittel

Hiermit weisen wir darauf hin, dass bei der Erstellung dieser Einreichung ChatGPT genutzt wurde, um die Kohärenz des Textes zu erhöhen und einzelne Formulierungen zu verbessern.

Verstoß gegen die Empfehlungen für einfache Sprache

In der DIN 8581-1-Norm (2024) werden eine Reihe von Empfehlungen für die einfache Sprache ausgesprochen. Diese beziehen sich auf verschiedene Ebenen eines Artikels, wie etwa die Text-, Satz- oder Wortebene. An den von ChatGPT erstellten Übersetzungen war auf Satzebene der häufige Einsatz des Konjunktivs auffällig. Auf Wortebene zeigte sich, dass bildhafte Sprache, wie etwa Metaphern und Allegorien nicht durchgängig vermieden wurden. Auch wurden nicht konsequent die gleichen Begriffe für dieselben Konzepte verwendet, sondern teils Synonyme, und schwierige Wörter wurden nicht immer durch kürzere und gebräuchlichere Begriffe ersetzt.

Vereinfachter Text	Fehlerbeschreibung	Fehlertyp
Könnte das an uns gelegen haben?	Besser wäre: Ist es möglich, dass das an uns gelegen hat?	Konjunktiv
Forscher aus Berlin haben das Eis genauer unter die Lupe genommen.	Besser wäre: Forscher aus Berlin haben das Eis genauer untersucht.	bildhafte Sprache
Verwendung der Begriffe Kunststoff und Plastik	Kunststoff und Plastik sind Synonyme. Besser wäre es, einen Begriff konstant zu verwenden	gleiche Begriffe
Spezielle Technik hat die Bewegung von Kopf, Rumpf und Armen aufgenommen.	Im Original wurde der gebräuchlichere Begriff „Oberkörper“ verwendet	überkomplexer Begriff