KI-Tools und Leichte Sprache.
Eine Untersuchung KI-generierter Texte anhand der DIN SPEC 33429.

Franziska Ludewig
Emily Reich

Erstellt im Rahmen eines Praktikums bei VERSO Dresden gGmbH.

Dresden, März 2025

KI-Tools und Leichte Sprache. Eine Untersuchung KI-generierter Texte anhand der DIN SPEC 33429

Abstract

Wie gut kann Künstliche Intelligenz (KI) standardsprachliche Texte in Leichte Sprache übersetzen? Die vorliegende Arbeit analysiert, inwiefern kostenlose KI-Tools für Leichte und Einfache Sprache die Anforderungen der DIN SPEC 33429 erfüllen. Basierend auf Mannings (2024) Analyse von KI-Tools für Einfache Sprache wurde eine angepasste Tool-Auswahl getestet und die Verständlichkeit der generierten Texte quantitativ sowie qualitativ bewertet. Die Ergebnisse zeigen, dass auf Leichte Sprache spezialisierte KI-Modelle geringere Abweichungen zur Norm aufweisen als Tools für Einfache Sprache. Da die Analyse vorrangig formale Aspekte betrachtet und individuelle Leseerfahrungen unberücksichtigt bleiben, sind weitere empirische Untersuchungen zur funktionalen Verständlichkeit erforderlich.

Schlagworte: KI, Leichte Sprache, DIN SPEC 33429, Übersetzungen

Einleitung

Leichte Sprache folgt strengen grammatikalischen, syntaktischen und formalen Regeln, die sicherstellen, dass Informationen nicht nur sprachlich reduziert, sondern barrierefrei zugänglich werden. Gerade dieser regelbasierte Ansatz macht Leichte Sprache für eine automatisierte Verarbeitung durch KI besonders interessant, bietet Leichte Sprache doch durch die ihr zugrundeliegenden standardisierten Strukturen eine recht klare Operationalisierbarkeit. Dabei bleibt jedoch die Frage offen, ob eine KI nur Oberflächenmerkmale wie Wortlängen oder Satzstrukturen erkennt oder ob sie auch den eigentlichen Kommunikationszweck der Leichten Sprache – eine barrierefreie Informationsvermittlung für spezifische Zielgruppen, ausgelegt auf deren individuelle Bedürfnisse – erfüllt. Inwiefern gelingt es ihr, Verständlichkeit nicht nur formal, sondern auch funktional herzustellen?

Auch Mannings (2024) Beitrag über „KI-Tools für Einfache Sprache“, in dem sie zehn KI-Modelle hinsichtlich ihrer Leistung für Einfache Sprache analysierte, kommt dieser Thematik nahe. Darin bewertete sie die Verständlichkeit verschiedener KI-generierter Übersetzungen anhand des Hohenheimer-Verständlichkeitsindexes (HIX) und nutzte fünf Textarten als Materialgrundlage für ihre Tests. Die vorliegende Untersuchung will dieses Konzept aufgreifen, unterscheidet sich jedoch in wesentlichen Punkten: Während die fünf Textarten weiterhin als Basis dienen, wurde die Auswahl der KI-Tools angepasst – einige kostenpflichtige Modelle wurden ausgeschlossen, weitere kostenfreie hinzugezogen. Die Grundlage für die Bewertung der KI-generierten Texte bildet zudem nicht der HIX, sondern die vor kurzem veröffentlichte DIN SPEC 33429, das zentrale Empfehlungswerk für Leichte Sprache (DIN e.V. 2025).

Ausgehend davon widmet sich der erste Teil der Arbeit der Bestandsaufnahme kostenloser KI-Tools für Leichte Sprache. Hier werden die verwendeten Systeme vorgestellt und nach Manning (2024) typologisiert. Der zweite Teil untersucht die Verständlichkeit der von diesen Tools generierten Texte sowohl quantitativ und qualitativ, vergleicht gewonnene Erkenntnisse mit den Anforderungen der DIN SPEC 33429 und erläutert diese anhand konkreter Anwendungsbeispiele. Dabei werden auch die eingesetzten Prompts und Modi der KI-Tools dokumentiert sowie die aus der DIN SPEC 33429 abgeleiteten Bewertungskategorien erläutert.

A. Welche kostenlose KI-Tools gibt es für Leichte Sprache?

Es gibt verschiedene KI-Tools, die Texte in Leichte Sprache übersetzen können. Eine Typologie zu diesen KI-Tools hat Manning aufgestellt. Sie differenziert dabei zwischen Allzweck-, Analyse- und Spezial-KI-Tools (vgl. Manning 2024).

Unter **Allzweck-KI-Tools** fallen große Sprachmodelle, bei denen Prompts für die Generierung eines neuen Textes notwendig sind (vgl. ebd.). Dazu gehören zum Beispiel *ChatGPT*, *Copilot* von Microsoft, *Gemini* von Google sowie *Hey_Die KI von Bild* und *WELTgo!* – beide von Nachrichtenagenturen stammend. Diese Tools sind zwar mehr für die Erfragung des Nachrichtenstandes ausgelegt, können aber durch die Verwendung von Prompts zur Übersetzung eines Textes angeleitet werden.

Der zweite Typ umfasst **Analyse-KI-Tools**. Das sind all solche, bei denen ein Prozess der Textbearbeitung stattfindet. Verbunden ist diese Analyse mit der Ausgabe von Änderungsvorschlägen für den Ausgangtext. Zusätzlich ist es bei diesen Tools möglich, verschiedene Niveaustufen auszuwählen (vgl. ebd.). Zu diesem Typ zählen zum Beispiel *Capito.ai* und *DeepLWrite*.

Die **Spezial-KI-Tools**, der dritte Typ, sind Tools, die speziell und ausschließlich auf die Textübersetzung eingestellt sind und den Ausgangstext in das gewünschte Sprachniveau übertragen (vgl. ebd.). Dazu gehören *ChatGPT-Bot „Klar und Verständlich“*, *Klartext*, *Language Tool*, *Leichte Sprache Übersetzer*, *Optimeil-Leichte-Sprache KI*, *einfachesprache.xyz*, und *Wortliga*. Letzteres wird hier den Spezial-KI-Tools zugeordnet und nicht, wie bei Manning, den Analyse-KI-Tools, da in diesem Fall die kostenlose Version „Text umschreiben“ verwendet wurde (vgl. ebd.).

Typen	KI-Tools
Allzweck-KI-Tools	ChatGPT, Copilot, Gemini, Hey_Die KI von Bild, WELTgo!
Analyse-KI-Tools	Capito.ai, DeepLWrite
Spezial-KI-Tools	ChatGPT-Bot „Klar und Verständlich“, Klartext, Language Tool, Leichte Sprache Übersetzer, Optimeil-Leichte- Sprache KI, St. Pauli bzw. einfachesprache.xyz, Wortliga

Tabelle 1: Übersicht der KI-Tools unterteilt in die drei Typen

In dieser Analyse wurden auch KI-Tools verwendet, die nicht explizit Texte in Leichte Sprache übertragen, sondern dafür ausgelegt sind, in Einfache Sprache zu übersetzen. Dies betrifft die folgenden Tools: *DeepLWrite*, *ChatGPT-Bot „Klar und Verständlich“*, *Klartext*, *Language Tool* und *einfachesprache.xyz*. Obwohl diese KI-Tools nur eine Übersetzung in Einfache Sprache versprechen, wurden sie für die Analyse verwendet, um zu prüfen, ob und inwieweit sie die Standards der Leichten Sprache erfüllen und für eine Verwendung geeignet sind.

B. Wie verständlich sind die Übersetzungen der KIs? Inwieweit weichen diese Übersetzungen von der DIN SPEC 33429 ab?

1. Ressourcen, Datenerhebung und Korpuserstellung

Ressourcen

In die KI-Tools wurden jeweils fünf Ausgangstexte für die Übersetzung in Leichte bzw. Einfache Sprache eingegeben. Als Ausgangstexte fungierten folgende Texte, die auch Manning in ihrer Analyse verwendet hat (vgl. Manning 2024):

- Käuferschutz bei PayPal (Auszug aus den Nutzungsbedingungen)
- Koloniale Raubkunst im Humboldt Forum (Auszug aus einem Artikel)
- Robert-Koch-Institut (RKI) (Auszug aus der Internetseite des RKI)
- Special Olympics 2023 (Rückblick des Organisationskomitees)
- Volksentscheid Berlin 2030 klimaneutral (Aufruf)

Weiterhin wurden basierend auf der DIN SPEC 33429 potenzielle Analysekategorien für die quantitative sowie qualitative Auswertung der KI-generierten Texte abgeleitet. Die DIN SPEC 33429 ist ein Empfehlungswerk für Deutsche Leichte Sprache. Sie enthält sprachliche Empfehlungen „für das Übersetzen […] [und] das Verfassen von Texten in Leichter Sprache“ (DIN SPEC 33429, 10), die der besseren „Verständlichkeit und Lesbarkeit“ (ebd.) jener Texte für Menschen mit kognitiven Beeinträchtigungen dienen; bestimmt so Anforderungen zur Nutzung Leichter Sprache (vgl. ebd., 9).

In dieser Arbeit wird – basierend auf eben diesen Empfehlungen – folglich untersucht, ob und inwieweit in den KI-Übersetzungen

- ein zentraler Wortschatz,
- kurze Wörter,
- die Zerlegung von Wörtern in Sinneinheiten,
- die Erläuterung schwieriger Wörter,
- keine Synonyme,
- keine Metaphern,
- kurze Sätze,
- einfache Sätze,
- Sinneinheiten pro Zeile,
- Verbalstil,
- Aktiv,
- Indikativ,
- wenig verschiedene Zeitformen (bevorzugt Präsens und Perfekt)
- und eine Textgliederung

verwendet werden. Eine weitere Bewertungskategorie ist, ob die KI den Sinn des Ausgangstextes beibehält und nicht entstellt.

Für die systematische Erfassung dieser Kategorien diente eine manuell in Excel erstelle Datenbank als Datengrundlage, die jedes getestete KI-Tool dem jeweiligem Übersetzungstext gegenüberstellt. Zusätzlich wurden die entsprechenden Ausgangstexte sowie deren thematische Ausrichtung – als Textart deklariert – hinterlegt.

Prompt und Modus

Die Übersetzung der Ausgangstexte wurde entweder durch einen Prompt oder durch die Wahl eines Modus bzw. Sprachniveaus angeleitet. Die Allzweck-KI-Tools erhielten für die Übersetzung folgenden Prompt:

> „Schreibe mir den Text in leichte Sprache um. Nutze kurze Sätze und Wörter, schreibe im Aktiv und Indikativ, vermeide Fachwörter und Fremdwörter, verwende den zentralen Wortschatz, nutze die persönliche Anrede im Sie."

Dieser Prompt wurde nach der Übersetzung des RKI- und Special Olympics-Textes zu folgendem Prompt modifiziert:

> „Schreibe mir den Text in leichte Sprache um. Nutze kurze Sätze und Wörter, schreibe im Aktiv und Indikativ, vermeide Fachwörter und Fremdwörter – wenn du diese verwendest, erkläre sie –, verwende den zentralen Wortschatz, nutze die persönliche Anrede im Sie."

Die Erklärung wurde ergänzt, da die KI-Tools nur Wörter erläuterten, wenn diese auch im Ausgangstext nachfolgend umschrieben wurden. Die Änderung diente der Prüfung, ob die KI-Tools selbstständig schwer verständliche Wörter identifizieren und diese in Leichter Sprache wiedergeben können, insofern diese auch dazu aufgefordert wurden.

Welcher Modus bzw. welches Sprachniveau für die Übersetzung mittels der jeweiligen Analyse- und Spezial-KI-Tools verwendet wurde, lässt sich folgender Übersicht entnehmen:

KI-Tool	Modus und Sprachniveau
Capito.ai	Mit KI-Vereinfachen: sehr leicht (A1-A2)
DeepLWrite	Stil: Einfach
ChatGPT-Bot „Klar und Verständlich"	Auswahl des vorhandenen Prompts: „Gib meinen Text in einfacher Sprache wieder."
einfachesprache.xyz	Modus: Von Alltagssprache in Einfache Sprache übersetzen
Klartext	Modus: Text vereinfachen für allgemeine Öffentlichkeit
Language Tool	Modus: Einfach
Leichte Sprache Übersetzer	Modus: Leicht
Optimeil-Leichte Sprache KI	Modus: „Bitte geben Sie mir einen Text, der in Leichte Sprache übertragen werden soll."
Wortliga	Modus: Leicht

Tabelle 2: Übersicht der verwendeten Modi und Sprachniveaus aufgeschlüsselt nach jeweiligem KI-Tool

2. Quantitativ-statistische Untersuchung

Anhand einer quantitativen Analyse sollen messbare sprachliche Merkmale erhoben werden, die mögliche Muster in den KI-generierten Übersetzungen aufzeigen. Dabei soll insbesondere betrachtet werden, inwiefern sich die maschinellen Übersetzungen formal an die DIN-Vorgaben der Leichten Sprache annähern und in welchen Bereichen Unterschiede zwischen den getesteten KI-Modellen bestehen. Die DIN SPEC 33429 bietet hierzu ein standardisiertes Bewertungsraster, das als Maßstab für die Untersuchung dient.

Methodik und Vorgehen

Inwieweit die KI-Texte von den Standards der DIN SPEC 33429 abweichen, wurde innerhalb der quantitativen Analyse an folgenden Aspekten untersucht:

a) **Wortlängen** – Gibt es Unterschiede hinsichtlich der durchschnittlichen Länge der Wörter? Neigen bestimmte Modelle dazu, kürzere und weniger komplexe Wörter zu verwenden?
b) **Satzlängen** – Wie variieren die Längen der Sätze zwischen den einzelnen KI-Modellen? Werden kürzere, einfachere Sätze bevorzugt?
c) **Lexikalische Variation** – Orientieren sich bestimmte KI-Modelle stärker an einem zentralen Wortschatz? Werden Fremd- und Fachwörter gleichermaßen vermieden?
d) **Nominal- und Verbalstil** – In welchem Maße reduzieren KI-Modelle die Verwendung von Substantivierungen zugunsten des bevorzugten Nominalstils?
e) **Passivkonstruktionen** – Reduzieren bestimmte Modelle das Passiv zugunsten aktiver Formulierungen?
f) **Konjunktiv** – Inwiefern halten sich die KI-Modelle an die Empfehlung, Konjunktiv in Leichter Sprache möglichst zu vermeiden?

Vorgehen und Analyse

Unter Nutzung verschiedener Libraries und Module erwies sich Python für den Analyseprozess als äußerst nützlich. Aber auch mit Sketch Engine konnten relevante Ergebnisse erzielt werden. Dabei wurden folgende Arbeitsschritte durchlaufen:

Datenimport und Vorbereitung

Die Excel-Datenbank (vgl. Ressourcen) wird mittels der Bibliothek *pandas* in ein DataFrame geladen. Um eine einheitliche Verarbeitung und standardisierte Datenbasis zu gewährleisten, werden alle Texte zunächst in Kleinbuchstaben konvertiert und von Sonderzeichen sowie Zahlen bereinigt.

Für die Analyse mit Sketch Engine wurde die Excel-Datenbank in eine xml-Datei überführt, die alle KI-Übersetzungen nach jeweiligem KI-Tool und Textart taggt. Anschließend wurden Subkorpora nach einzelnen KI-Tools erstellt.

Textverarbeitung

Für die linguistische Annotation wird das Python-Framwork *spaCy* verwendet, das speziell für die Verarbeitung natürlicher Sprache (NLP) optimiert ist. Das darin integrierte deutsche Sprachmodell *de_core_news_sm* dient nicht nur der Tokenisierung – der Zerlegung des Textes in einzelne Worteinheiten (Tokens) –, sondern auch der Lemmatisierung, bei der jedes Token seiner Grundform (Lemma) zugeordnet wird. Dieses Verfahren reduziert die sprachliche Varianz – flektierte Wörter werden dieserart z.B. in ihrer Grundform zusammengefasst – und ermöglicht so den Vergleich formaler Aspekte, ohne Verzerrungen zu riskieren. Auch ein *Part-of-Speech (POS)*-Tagging wird vorgenommen, wodurch die Texte zusätzlich mit linguistischem Kontext versehen werden, der für weiterführende textanalytische Untersuchungen von entscheidender Bedeutung ist. Bei Sketch Engine läuft die Annotation bei Korpuserstellung größtenteils automatisch ab.

Erfassung quantitativer Merkmale

a) Wortlänge

Für jedes Token wird die Anzahl der Buchstaben ermittelt. Anschließend wird die mittlere Wortlänge (Median als Zentralwert) gruppiert nach KI-Tool (und optional nach Textart) berechnet.

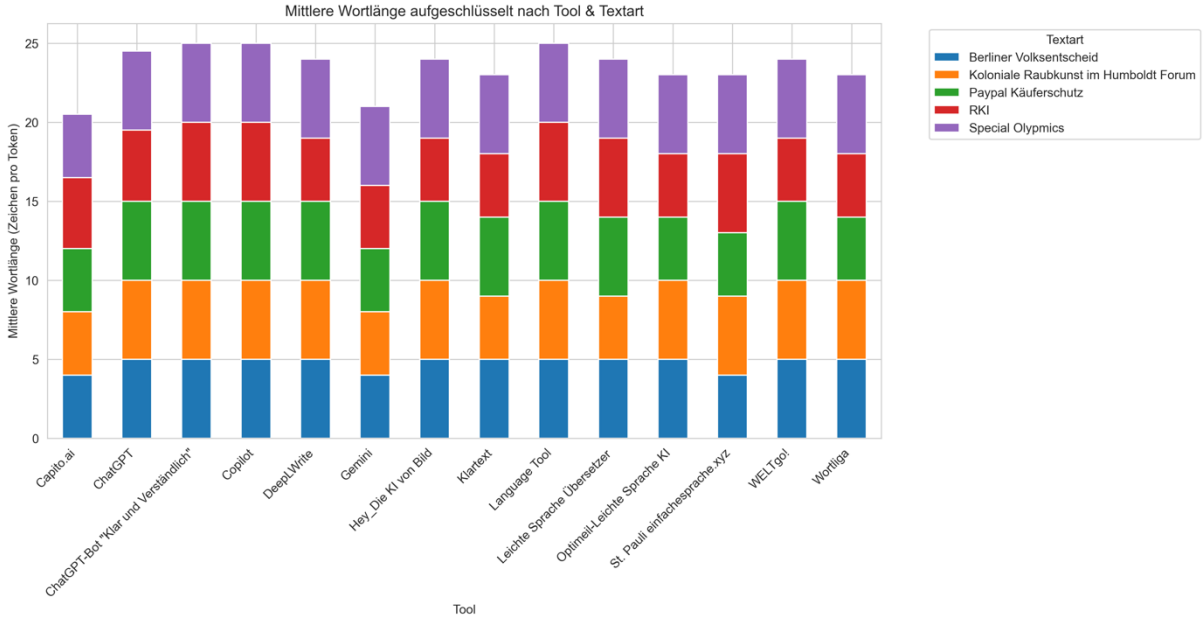

Abbildung 1: Mittlere Wortlänge aufgeschlüsselt nach Tool und Textart

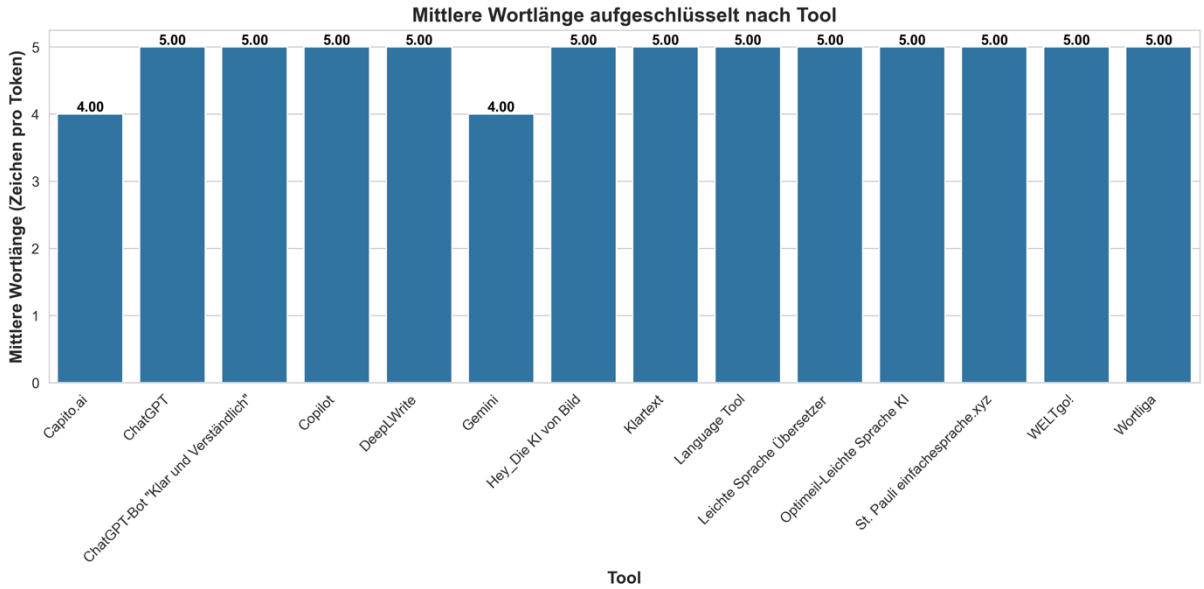

Abbildung 2: Mittlere Wortlänge aufgeschlüsselt nach Tool

Ergebnisse

Die meisten der getesteten KI-Modelle weisen sehr ähnliche Werte auf. Mit einer mittleren Wortlänge von vier bis fünf Zeichen pro Token unterscheiden sich die Mehrheit der Systeme nur minimal voneinander. Nur zwei Modelle finden sich bei einem Median von vier ein: *Capito.ai* und *Gemini*.

b) Satzlänge

Jeder Text wird in seine einzelnen Sätze unterteilt, die Sätze wiederum nach ihren Tokens aufgeschlüsselt, um die Anzahl der Wörter pro Satz zu ermitteln. Anschließend werden die Satzlängen nach entsprechenden Tools (und optional nach Textart) gruppiert und die mittlere Satzlängen berechnet.

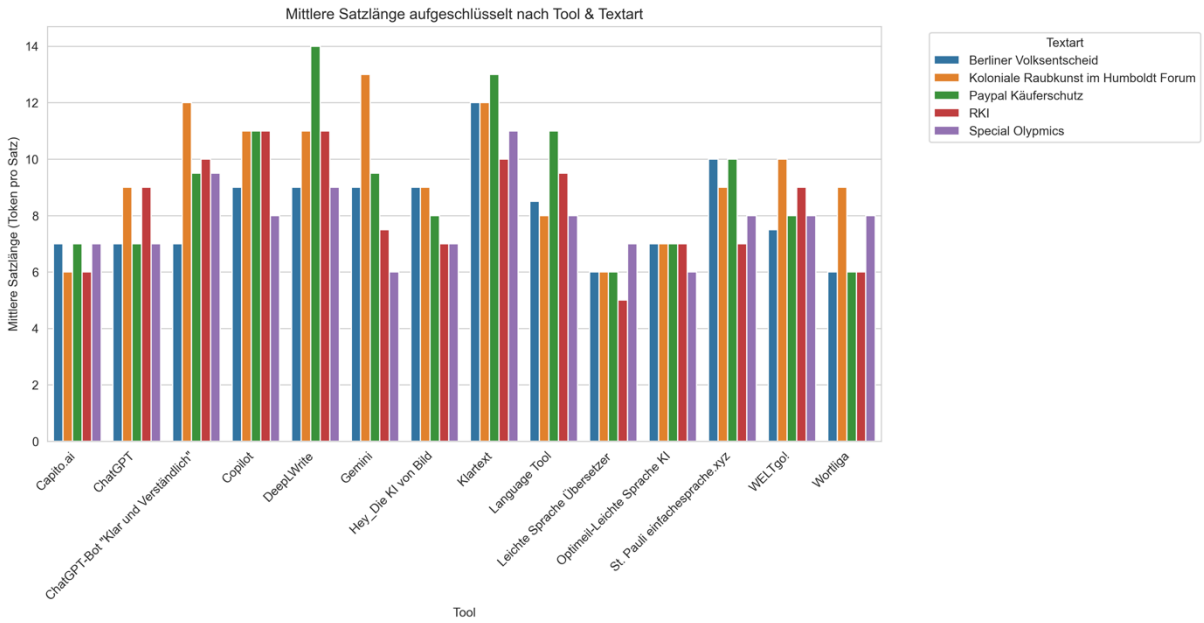

Abbildung 3: Mittlere Satzlängen aufgeschlüsselt nach Tool und Textart

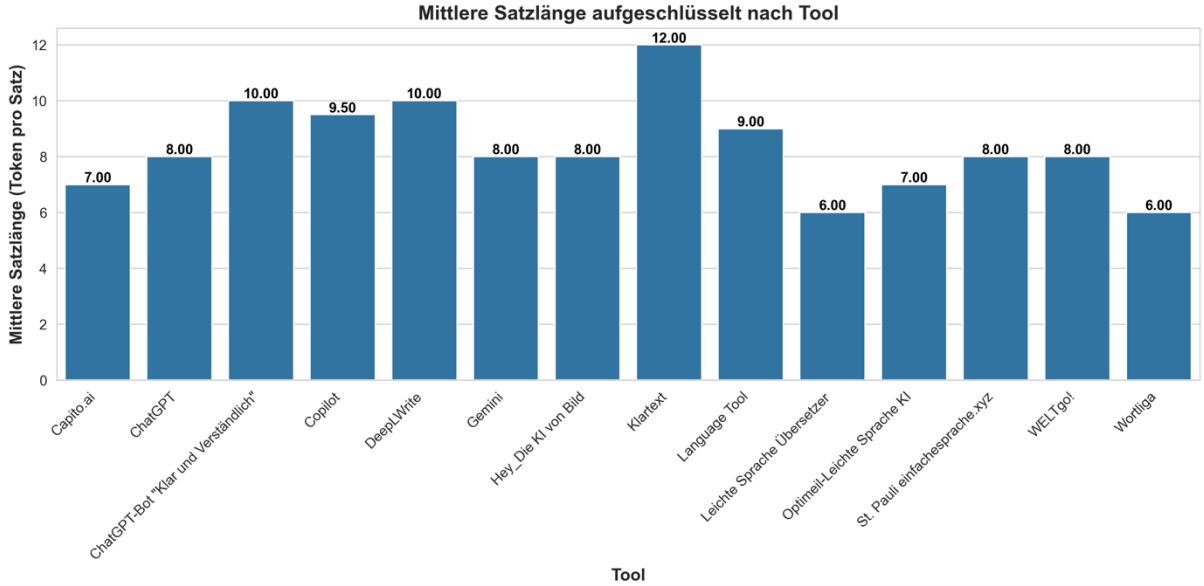

Abbildung 4: Mittlere Satzlängen aufgeschlüsselt nach Tool

Ergebnisse

Während einige Tools konsequent kurze Sätze erzeugen, variieren andere je nach Textart erheblich in ihrer Satzlänge. Die Modelle *Leichte Sprache Übersetzer* und *Wortliga* halten sich über verschiedene Textarten hinweg an eine reduzierte Satzlänge von ca. sechs Token pro Satz. Gegensätzlich dazu generieren Tools wie *DeepLWrite*, *Klartext* und *ChatGPT-Bot „Klar und Verständlich“* längere Sätze von zehn bis zwölf Token pro Satz. Von einer Überschreitung der

empfohlenen DIN-Satzlänge von acht bis zehn Token pro Satz kann hier dennoch nicht gesprochen werden. Auffällig zeigen sich jedoch Texte zum Thema „Paypal Käuferschutz“ oder „Koloniale Raubkunst im Humboldt Forum“, deren mittlere Werte im jeweiligen Toolbereich mehrheitlich stärker ausgeprägt sind als bei anderen Textarten.

c) Lexikalische Variation

Was ist der Kern unseres Wortschatzes? Dass er als „kommunikatives Wortschatzminimum“ verstanden werden kann – so Rosengren –, scheint recht einleuchtend, muss er doch die wichtigsten Wörter enthalten, um überhaupt eine Kommunikationssituation zu erzeugen (vgl. Rosengren, 330). Sicher kann dieser Aspekt noch aus vielen weiteren Perspektiven beleuchtet werden. An dieser Stelle bleibt es jedoch vorerst dabei, festzustellen, dass sich die Definition von „zentralem Wortschatz“ je nach Quelle und Methodik unterschiedlich bestimmen lässt.

Für unsere Zwecke nutzen wir die von gut1.de gut recherchierte, durch mehrere Quellen belegte und kompakt zusammengestellte Liste der 500 häufigsten Wörter der deutschen Sprache. Fach- und Fremdwörter sind nach eigener Sichtung darin nicht zu finden. Alle darin enthaltenen Wörter werden als Set in Python überführt und ebenfalls in Kleinbuchstaben konvertiert, um Validität zu wahren. Sind die lemmatisierten Wörter der KI-Übersetzungen im Grundwortschatz (GWS)-Set enthalten, wird für jedes Tool (und optional jede Textart) der Anteil der vorkommenden Grundwortschatzwörter berechnet.

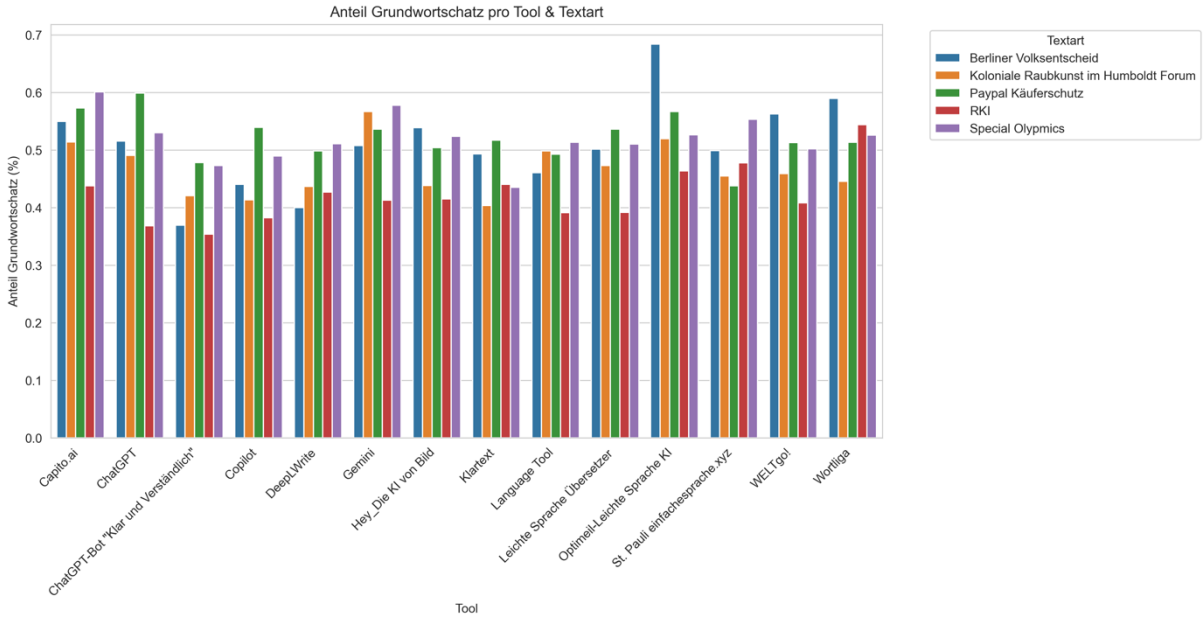

Abbildung 5: Anteil Grundwortschatz pro Tool und Textart

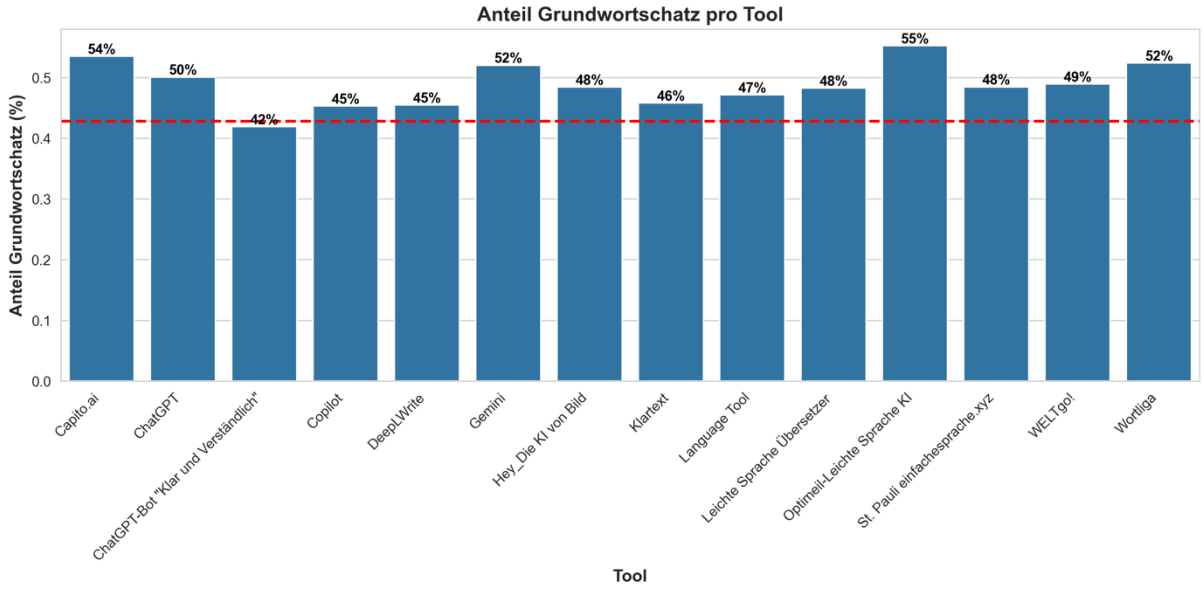

Abbildung 6: Anteil Grundwortschatz pro Tool

Ergebnisse

Sowohl zwischen den KI-Modellen als auch zwischen den Textarten sind deutliche Unterschiede hinsichtlich der jeweiligen GWS-Anteile festzustellen. Dabei weisen Tools wie *Capito.ai*, *Gemini*, *Wortliga* und *Optimeil-Leichte Sprache KI* einen höheren GWS-Anteil auf, während *ChatGPT-Bot „Klar und Verständlich“*, *Copilot* und *DeepLWrite* kaum über oder sogar unter dem durchschnittlichen GWS-Anteil der Ausgangstexte (rote Linie) liegen. Auch die Themen der Texte beeinflussen den GWS-Anteil: Einige Themenbereiche – insbesondere „RKI“ – enthalten weniger Wörter aus dem GWS, während andere – so z.B. „Paypal Käuferschutz“ – tendenziell höhere Anteile aufweisen. Dennoch ist nicht von der Hand zu weisen, dass sich alle getesteten KI-Tools in gewissem Maße an den ursprünglichen Wortschatz anlehnen und keine radikalen Umformungen unternehmen.

d) Nominal- und Verbalstil

Um zu untersuchen, ob die KI-Modelle eher Nominalstil oder Verbalstil verwenden, werden die Part-of-Speech-Tags (POS) der generierten Texte analysiert. Die Einteilung in Nominal- und Verbalstil erfolgte dabei anhand sprachtypischer Merkmale, die sich anhand der getaggten Tokens in zwei Gruppen einteilen lassen:

- Nominalstil: umfasst Nomen (NOUN), Eigennamen (PROPN) und Determinierer (DET). Diese Wortarten sind typisch für einen abstrakten, formellen Stil, der viele Sachverhalte durch Substantivierungen ausdrückt.
- Verbalstil: umfasst Verben (VERB), Adverbien (ADV) und Hilfsverben (AUX). Diese Gruppe steht für eine aktivere, dynamischere Sprache, die eher Handlung und Prozess betont. Verbalstil wird in der Leichten Sprache bevorzugt, da er direkter und verständlicher ist.

Es folgt die Berechnung des Nominal- und Verbalstilanteils relativ zur Gesamtzahl der Tokens. Da andere Tags wie Konjunktionen (CONJ) oder Adjektive (ADJ) dabei für die Berechnung nicht relevant sind, werden die Anteile von Nominal- und Verbalstil auf 100 Prozent normiert. Beide Kategorien können dieserart leichter miteinander verglichen werden.

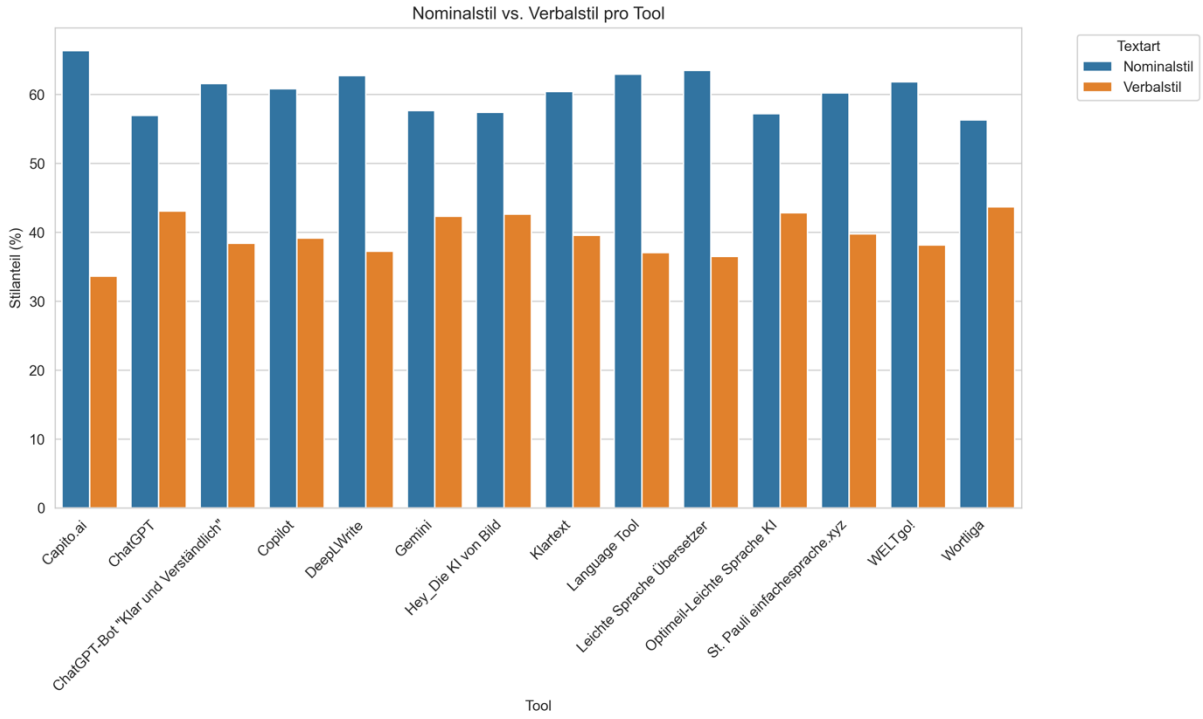

Abbildung 7: Stilanteil pro Tool

Ergebnisse

Alle getesteten KI-Modelle weisen entgegen den Empfehlungen der DIN SPEC 33429 einen höheren Anteil an nominalen Satzkonstruktionen auf. *Capito.ai* zeigt dabei sowohl den höchsten nominalen Anteil von über 65 Prozent als auch den niedrigsten verbalen Anteil von ca. 35 Prozent an. *Wortliga* hingegen setzt mit ca. 43 Prozent verbalem und 57 Prozent nominalem Anteil auf einen recht ausgeglichenen Ausdruck.

e) Passivkonstruktionen

Auf Basis der Subkorpora werden die Passivkonstruktionen anhand von festgelegten Schlagworten ausgewertet. Mit der Concordance-Funktion sucht eine CQL-Abfrage im Korpus gezielt nach „wurde, wurden, worden“ als Prädiktoren für Passivkonstruktionen.

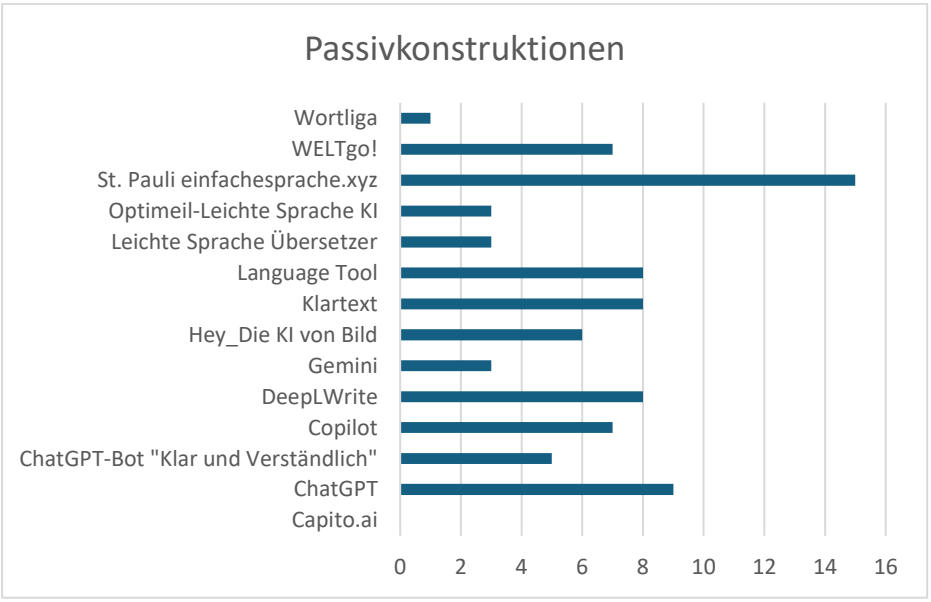

Abbildung 8: Passivkonstruktionen pro Tool

Ergebnisse

Alle untersuchten KI-Tools weisen trotz der Vorgaben der DIN SPEC 33429 Passivkonstruktionen auf. Bei den Übersetzungen des Tools *einfachesprache.xyz* lagen die meisten Passivkonstruktionen – mit insgesamt 15 – vor. Die übrigen KI-Tools weisen durchschnittlich drei bis acht Passivkonstruktionen auf. Dahingegen verwendete das KI-Modell *WELTgo!* nur eine Passivkonstruktion in den gesamten fünf Übersetzungen.

f) Konjunktiv

Um zu untersuchen, ob die KI-Modelle in ihren Übersetzungen den Konjunktiv verwenden, wird eine vordefinierte, manuell erstellte Liste mit der Concordance-Funktion in Sketch Engine abgefragt. Diese Liste enthält: Hilfsverben und Modalverben, jeweils in den Formen des Konjunktiv I und Konjunktiv II.

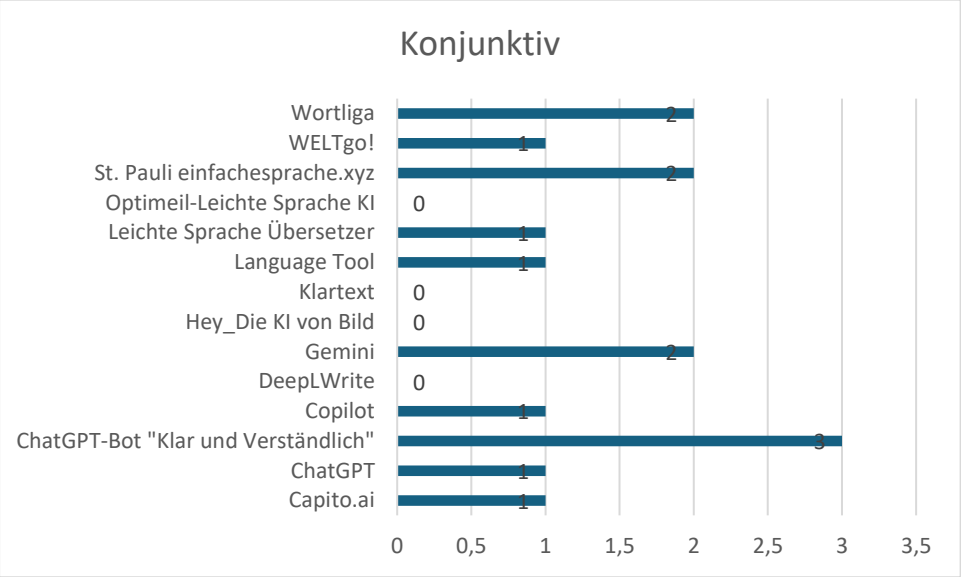

Abbildung 9: Konjunktivanteil pro Tool

Ergebnisse

Trotz der Empfehlung der DIN SPEC 33429 nutzten einige Tools einen Konjunktiv, wenn auch im geringen Maß von einer bis drei Konjunktivkonstruktionen. In den Texten von *ChatGPT-Bot „Klar und Verständlich“* lag am häufigsten ein Konjunktiv vor. Dahingegen wurden die Texte der Tools *Optimeil-Leichte Sprache KI*, *Klartext*, *Hey_Die KI von Bild*, *DeepLWrite* vollständig im Indikativ formuliert.

3. Qualitative Bewertung

Bei der qualitativen Auswertung wurde jeder KI-generierte Text einzeln auf die Vorgaben der DIN SPEC 33429 hin analysiert. Es wurde geprüft, ob und inwieweit die oben aufgeführten Aspekte erfüllt worden sind (vgl. Ressourcen, Datenerhebung und Korpuserstellung). Die Einhaltung wurde zusätzlich zur quantitativ-statistischen Analyse einzelner Aspekte kategorial erfasst: vollständige Einhaltung, teilweise Einhaltung, keine Einhaltung. Diese dreistufige Kategorisierung diente zur systematischen Bewertung der Texte nach der DIN SPEC 33429. Für einen textübergreifenden Vergleich der qualitativen Ergebnisse wurden die drei Kategorien mit numerischen Werten versehen: vollständige Einhaltung = 1, teilweise Einhaltung = 0.5, keine

Einhaltung = 0.[1] Dieses Vorgehen ermöglichte – insbesondere bei der großen Datenmenge – eine zusammenfassende Bewertung der KI-Tools, indem die Werte der fünf Texte für jeden Aspekt der DIN SPEC 33429 summiert und zusammengetragen wurden. So steht ein abschließender Wert von 5 für eine vollständige Einhaltung der Norm, während 0 keine Einhaltung darstellt. Diese Einhaltungen und Abweichungen von der DIN SPEC 33429 werden im Folgenden durch Beispiele verdeutlicht und mit den quantitativen Ergebnissen aus dem vorherigen Kapitel abgeglichen.

Tools und ihre Ergebnisse

Die Analyse zeigt, dass die Texte der Spezial-KI-Tools *Optimeil-Leichte-Sprache KI*, *Leichte Sprache Übersetzer* und *einfachesprache.xyz* die geringste Abweichung von der DIN SPEC 33429 aufweisen. Danach folgen im Ranking die Allzweck-KI-Tools *ChatGPT* und *Gemini*, aber auch *Hey_Die KI von Bild* weist eine vergleichsweise geringe Abweichung auf.

Interessanterweise befinden sich im unteren Bereich des Rankings auch einige Spezial-KI-Tools wie der *ChatGPT-Bot „Klar und Verständlich“*, *Klartext* und *Language Tool*. Diese Modelle sind jedoch auf die Einfache Sprache ausgerichtet, was die größere Abweichung von der DIN SPEC 33429 erklären kann. Ähnliches kann auch für das Analyse-Tool *DeepLWrite* gelten, dessen Texte auch von der DIN SPEC 33429 abweichen. Auch die Allzweck KI *Copilot* erreichte nur mäßige Ergebnisse.

Wortebene:

Zentraler Wortschatz

Die DIN SPEC 33429 sieht vor, dass für Leichte Sprache ein zentraler Wortschatz verwendet wird. Dieser wird jedoch im Empfehlungswerk nicht definiert. Die DIN SPEC 33429 verweist darauf, dass der zentrale Wortschatz nach Nutzenden variieren kann (vgl. DIN e.V. 2025, 13). Die fehlende Definition, welche Wörter dem zentralen Wortschatz angehören, stellt für Auswertung dieserart – wie schon oben thematisiert – ein klares Problem dar. Um dennoch brauchbare Ergebnisse zu erhalten, wurden die Texte zunächst quantitativ mit einer vordefinierten GWS-Sammlung – bereitgestellt durch gut1.de – abgeglichen, ehe sie qualitativ nach eigenem Ermessen danach bewertet wurden, ob sie alltagsnahe Wörter sowie keine Fach- und Fremdwörter beinhalten.

Die KI-Tools *Optimeil-Leichte-Sprache KI*, *Wortliga* und *Gemini* verwendeten diesbezüglich besonders häufig einen zentralen Wortschatz bei ihrer Übersetzung – in fast gleichem Umfang auch das auf Einfache Sprache spezialisierte KI-Tool *einfachesprache.xyz*. Die übersetzten Texte der Allzweck KI-Tools *ChatGPT*, *Hey_Die KI von Bild* und *WELTgo!* besaßen teilweise einen zentralen Wortschatz. Dahingegen war bei den Übersetzungen von *Klartext*, *DeepLWrite*, *Language Tool*, *ChatGPT-Bot „Klar und Verständlich“*, *Copilot* und *Capito* kaum ein zentraler Wortschatz erkennbar.

Trotzt dessen die quantitative Analyse eine vordefinierte GWS-Sammlung nutzte, stimmt sie in wesentlichen Punkten, insbesondere hinsichtlich der Tools, die einen hohen oder niedrigen Anteil an GWS verwenden, mit der qualitativen Analyse überein. Dennoch zeigen sich Abweichungen, etwa bei *Capito.ai*: Während der KI in der quantitativen Analyse ein hoher GWS-Anteil zugeschrieben wird, weisen die Übersetzungen nach qualitativer Betrachtung kaum zentralen Wortschatz auf. Diese Diskrepanzen lassen sich auf mehrere Ursachen

[1] Die Abstufung „teilweise Einhaltung“ entfiel bei den folgenden Aspekten: Zerlegung langer Wörter, Erläuterung schwer verständlicher Wörter, keine Synonyme, keine Metapher, Sinneinheiten pro Zeile, Textgliederung und Sinn. Entweder lag der Aspekt vor oder war nicht vorhanden.

zurückführen: So misst die quantitative Analyse lediglich den Anteil an Wörtern aus einer vordefinierten GWS-Liste, ohne deren Kontext oder Verständlichkeitsfunktion zu bewerten. Die qualitative Analyse hingegen berücksichtigt den sprachlichen Gesamteindruck, wobei subjektive Einschätzungen und die Definition dessen, was als „zentraler Wortschatz" gilt, eine Rolle spielen. Dieserart ergibt sich auch aus dem Einfluss des thematischen Kontextes auf den GWS-Anteil die Notwendigkeit, die reine Wortschatzanalyse mit einer qualitativen Bewertung wie obenstehend zu ergänzen, lässt sich doch so erst verstehen, inwiefern Textvarianten tatsächlich zu besserer Verständlichkeit beitragen oder gegebenenfalls an Präzision verlieren.

Kurze Wörter

Unter kurzen Wörtern sind nach der DIN SPEC 33429 all die Wörter zu verstehen, die nicht mehr als drei Silben besitzen (vgl. DIN e.V. 2025, 15).[2] Die DIN SPEC 33429 empfiehlt, lange Wörter zu meiden, da diese von Menschen mit kognitiven Beeinträchtigungen schwieriger erfasst werden (vgl. ebd.).

Auch bei der Verwendung von kurzen Wörtern schneiden die KI-*Tools Optimeil-Leichte-Sprache KI* und *Wortliga* nach qualitativer Betrachtung sehr gut ab. Ebenso benutzte auch das auf Einfache Sprache spezialisierte KI-Tool *einfachesprache.xyz* kurze Wörter. Dahingegen verwendeten *DeepLWrite* und *Language Tool* in ihren Texten kaum kurze Wörter; ebenso – und trotz Prompt – das Allzweck KI-Tool *Copilot*. Jedoch nutzten die restlich untersuchten Allzweck KI-Tools teilweise kurze Wörter.

So zeigt auch die quantitative Analyse: Benannte KI-Übersetzungen finden sich bei einer geringeren mittleren Wortlänge ein. Dabei lassen sich allerdings kaum Unterschiede zwischen den einzelnen Modellen ausmachen. Dies findet seine Ursache möglicherweise in dem Umstand, dass zwar eine mittlere Wortlänge berechnet werden kann, diese aber nicht spezifisch aussagt, ob längere, komplexe Wörter tatsächlich reduziert werden. Entscheidend bleibt, ob die KI-Tools nicht nur kurze Wörter nutzen, sondern diese auch gezielt einsetzen, um den Text verständlicher zu machen.

Dahingehend belegt folgender Beispielsatz zunächst die Beobachtung, dass im Ausgangstext „RKI" kein zentraler Wortschatz und lange Wörter vorhanden sind. Deutlich wird dies durch Wörter wie „Krankheitsüberwachung und -prävention", „anwendungs- und maßnahmenorientierten" sowie „biomedizinisch":

> *Das RKI ist die zentrale Einrichtung der Bundesregierung auf dem Gebiet der Krankheitsüberwachung und -prävention und damit auch die zentrale Einrichtung des Bundes auf dem Gebiet der anwendungs- und maßnahmenorientierten biomedizinischen Forschung.*

Gemäß der DIN SPEC 33429 löste *einfachesprache.xyz* diese schwierigen Wörter durch einfache Verben wie „kümmern" und „schauen" sowie durch das Substantiv „Krankheiten" auf. Der Verbalstil sorgt dabei für eine bessere Verständlichkeit des Satzes.

> einfachesprache.xyz:
>
> *Das Gesundheits-Ministerium ist ein Teil von der Regierung.*
> *Das Robert Koch-Institut kümmert sich um Krankheiten.*
> *Es schaut, welche Krankheiten es gibt.*

Dagegen blieben bei der Übersetzung von *Language Tool* die langen und schwer verständlichen Wörter weiter erhalten. Zudem wurde der zweite Teilsatz von der KI weggelassen, ohne

[2] Operationalisiert wurde dieser Aspekt in der quantitativen Analyse mittels der Berechnung der mittleren Wortlänge. Dabei wird der Wert von maximal sechs Zeichen pro Token als DIN-Wert angenommen.

dessen Inhalt an anderer Stelle umformuliert wiederzugeben. Eine Reduktion, die durchaus zu kritisieren ist.

> Language Tool:
>
> *Das RKI ist die zentrale Einrichtung der Bundesregierung auf dem Gebiet der Krankheitüberwachung und -prävention.*

Zerlegung von langen Wörtern und Komposita

Für längere Wörter und Komposita empfiehlt die DIN SPEC 33429, die Wörter – wenn möglich – in ihre einzelnen Sinneinheiten zu zerlegen. So wird das Wort „Blumengeschäft“ in Leichter Sprache zu „Blumen-Geschäft“ (vgl. DIN e.V. 2025, 15).

Die KI-Tools *Optimeil-Leichte Sprache KI*, *Leichte Sprache Übersetzer*, *einfachesprache.xyz* und *ChatGPT* zerlegten lange Wörter in einzelne Sinneinheiten. Die zuletzt genannten Tools trennten lange Wörter nur teilweise und das auch nicht durchgängig im Text. Die anderen Tools zerlegten lange Wörter gar nicht.

Ein Beispielauszug aus dem KI-Tool *Leichte Sprache Übersetzer* veranschaulicht, wie das lange Wort „Kolonialzeit“ im übersetzten Text einmal zerlegt wurde, um an anderer Stelle wieder zusammengeschrieben zu werden. Diese fehlende Kontinuität entspricht nicht den Standards der DIN SPEC 33429 und bleibt zu kritisieren.

> Leichte Sprache Übersetzer:
>
> *Es zeigt Kunst aus der Kolonialzeit. […]*
> *Erinnerung an die Kolonial-Zeit.*

Erläuterung von schwierigen Wörtern

Die DIN SPEC 33429 empfiehlt Erläuterungen in Form von „Umschreibungen, Erklärungen und Beispielen“ (DIN SPEC 33429, 13).

Insbesondere die ChatBots *Gemini*, *Optimeil-Leichte-Sprache KI* und *ChatGPT* erklärten schwierige Wörter ausführlich. Im Gegensatz dazu ließen sich kaum erklärende Erläuterungen in den Übersetzungen der Tools für Einfache Sprache finden. Das betrifft *DeepLWrite*, *Language Tool* und *Klartext*.

Im folgenden Beispiel ist zu erkennen, dass im Ausgangstext „Raubkunst“ die schwer verständlichen Begriffe „Exponate“ und „Kolonialzeit“ nicht erläutert wurden.

> *Das Berliner Humboldt Forum stand in der Kritik, als es noch gar nicht eröffnet war. Neben den explodierten Baukosten steht das Museum im Mittelpunkt der Debatte über Exponate aus der Kolonialzeit, die es zeigt. Der Deutsche Kulturrat schlägt für das Haus eine Generalintendanz vor, um dem Haus endlich „Leben einzuhauchen“.*

Das KI-Tool *Gemini* ersetzte gemäß der DIN SPEC 33429 zum einen das Wort „Exponate“ durch das im Grundwortschatz enthaltende Wort „Sachen“. Zum anderen integrierte das Tool eine Erläuterung zum Begriff „Kolonialzeit“ in Leichter Sprache.

> Gemini:
>
> ***Das Humboldt Forum: Ein Haus voller Geschichte(n)[3]***
>
> *Das Humboldt Forum in Berlin ist ein neues Museum. Aber es gab schon Streit, bevor es überhaupt aufgemacht hat. Warum? Es geht um die Sachen, die dort ausgestellt sind. Viele davon stammen aus der Kolonialzeit. Das bedeutet, sie wurden aus Ländern geholt, die früher von anderen Ländern beherrscht wurden.*

[3] Die Hervorhebung generierte das KI-Tool selbst. Auch alle weiteren Hervorhebungen in den übersetzten Texten stammen von den KI-Tools.

Was ist Kolonialzeit?

Die Kolonialzeit war eine lange Zeit, in der mächtige Länder andere Länder besetzt und ausgebeutet haben. Das ist oft sehr ungerecht gewesen.

Dahingegen verwendete *Klartext* weiterhin diese schwer verständlichen Wörter und erläuterte diese auch nicht.

Klartext:

Das Humboldt Forum in Berlin wurde kritisiert, noch bevor es eröffnet wurde. Die Kritik liegt nicht nur an den gestiegenen Baukosten, sondern auch an der Darstellung von Exponaten aus der Kolonialzeit. Der Deutsche Kulturrat schlägt vor, eine Generalintendanz für das Museum einzurichten, um es zu beleben.

Keine Synonyme

Die DIN SPEC 33429 sieht vor, keine Synonyme in einem Text zu verwenden, da dies das Textverständnis beeinträchtigen kann (vgl. DIN e.V. 2025, 13). Stattdessen sollten gleiche Begriffe kontinuierlich gleich genutzt und benannt werden.

Keine bzw. kaum Synonyme benutzten *Optimeil-Leichte-Sprache KI*, *Hey_Die KI von Bild* und *Klartext*, wohingegen bei den Übersetzungen von *ChatGPT-Bot „Klar und Verständlich“*, *DeepLWrite* und *Language Tool* häufiger Synonyme vorhanden waren. Der eben bereits verwendete Ausschnitt aus dem Text „Raubkunst“ lässt auch Rückschlüsse auf die Verwendung von Synonymen zu, er zeigt beispielsweise die synonyme Nutzung der Wörter „Museum“ und „Haus“ im Ausgangstext. Diese Synonyme wurden auch von *Language Tool* in dessen Übersetzung übernommen.

Language Tool:

Das Humboldt Forum wurde kritisiert, als es noch nicht eröffnet wurde. Das Museum steht im Mittelpunkt, weil es viele Exponate aus der Kolonialzeit zeigt. Der Deutsche Kulturrat schlägt vor, dass das Haus eine Generalintendanz bekommt.

Dahingegen verwendete das Tool *Hey_Die KI von Bild* nach den Standards der DIN SPEC 33429 durchgängig das Wort „Museum“.

Hey_Die KI von Bild:

Das Humboldt Forum in Berlin wurde schon kritisiert, bevor es eröffnet wurde. Die Baukosten waren sehr hoch. Das Museum zeigt Kunstwerke aus der Kolonialzeit. Der Deutsche Kulturrat möchte, dass das Museum eine neue Leitung bekommt, damit es besser wird.

Keine Metaphern

Nach der DIN SPEC 33429 sollte eine bildhafte Sprache vermieden werden, weil dies eine Übertragungsleistung voraussetzt und damit das Textverständnis kompliziert (vgl. DIN e.V. 2025, 14).

Einige Tools integrierten Metaphern in ihren Übersetzungen, insbesondere dann, wenn auch im Ausgangstext ein metaphorischer Schreibstil verwendet wurde. Dies trifft auf die Tools *einfachesprache.xyz*, *Capito* und *Language Tool* zu. In den Texten von *Leichte Sprache Übersetzer* und *Wortliga* waren durchgängig keine Metaphern vorhanden.
Deutlich wird dies ebenfalls im Textausschnitt zum Thema „Raubkunst“. Dieser enthält die Metapher „um dem Haus endlich ‚Leben einzuhauchen‘“. *Capito.ai* übernahm diese in ähnlicher Form, während das Tool *Wortliga* die Metapher durch die Formulierung „mehr Bedeutung zu geben“ ersetzte.

Capito:

Der Kulturrat hat eine Idee.
Er möchte die Leitung für das Haus.
Das soll das Haus lebendig machen.

Wortliga:

Der Deutsche Kulturrat fordert eine neue Leitung, um dem Museum mehr Bedeutung zu geben.

Satzebene:

Kurze Sätze und einfache Satzstruktur

Als kurze Sätze gelten nach der DIN SPEC 33429 all die Sätze, die nicht mehr als acht bis zehn Wörter aufweisen. Einfache Satzstrukturen sind nach der DIN SPEC 33429 Sätze ohne mehrere Teil- und Nebensätze, ohne Relativsätze mit Präposition am Anfang und ohne Nominalphrasen (vgl. DIN e.V. 2025, 18).

Besonders KI-Tools, die auf Leichte Sprache spezialisiert sind oder über leichte Niveaustufen verfügen, gaben Texte mit kurzen Sätzen und einer einfachen Satzstruktur aus. Dazu gehören *Leichte Sprache Übersetzer*, *Optimeil-Leichte Sprache KI* und *Capito*. Die Texte der Allzweck KI-Tools *ChatGPT*, *Hey_Die KI von Bild* und *WELTgo!* wiesen teilweise kurze Sätze und eine einfache Satzstruktur auf. Lange und schwer verständliche Sätze waren häufiger bei den Tools für Einfache Sprache vorhanden wie bei *ChatGPT-Bot „Klar und Verständlich"* und *Language Tool*, aber auch bei dem Allzweck KI-Tool *Copilot*.

Während *Capito* sowohl quantitativ als auch qualitativ als Tool mit kurzen und einfachen Sätzen eingestuft wird, weichen die Einschätzungen bei anderen Modellen erheblich voneinander ab. Insbesondere *Klartext* wird in der quantitativen Analyse als Tool mit den längsten Sätzen identifiziert, während es in der qualitativen Betrachtung nicht als auffällig komplex beschrieben wird. Diese Abweichung findet Ursache darin, dass die quantitative Analyse ausschließlich die Satzlänge in Token misst, während die qualitative Analyse auch Satzstrukturen wie Nebensätze, Relativsätze oder Nominalphrasen berücksichtigt. Ein Satz kann formal kurz sein, aber durch syntaktische Komplexität dennoch schwer verständlich bleiben. Entscheidend ist, ob die Satzstruktur einfach und linear aufgebaut ist oder ob grammatische Konstruktionen enthalten sind, die das Erfassen erschweren.

Zum Beispiel enthält ein Satz des Textes „PayPal-Käuferschutz" 27 Wörter und komplexe Nominalphrasen.

Bei Vorliegen der entsprechenden Voraussetzungen haben Sie nach dem PayPal-Käuferschutz Anspruch auf die Erstattung des vollen Kaufpreises für den Artikel zuzüglich der von Ihnen gezahlten ursprünglichen Versandkosten.

Language Tool löste diesen langen Satz zwar in zwei Sätze auf, jedoch besteht einer davon weiterhin aus 17 Wörtern. Zudem bleibt die Satzstruktur durch den Nominalstil im Hauptsatz und den vorhandenen Nebensatz komplex.

Language Tool:

Wenn Sie die Voraussetzungen erfüllen, bekommen Sie nach dem PayPal-Käuferschutz das volle Geld für den Artikel zurück. Dazu gehören auch die Versandkosten von Ihnen.

Das Tool *Capito* löste nach den Standards der DIN SPEC 33429 den ursprünglichen Satz in mehrere, kürzere Sätze mit maximal zehn Wörter pro Satz auf. Die Sätze wurden zusätzlich im Verbalstil formuliert.

Capito:

Wenn die Voraussetzungen stimmen, bekommen Sie Geld von PayPal zurück.
Sie bekommen den vollen Kaufpreis zurück.

Sie bekommen auch die Versandkosten zurück.

Sinneinheiten pro Zeile

Die DIN SPEC 33429 empfiehlt, eine Sinneinheit, d.h. ein Satz pro Zeile (vgl. DIN e.V. 2025, 19).

Die Texte der KI-Tools *Optimeil-Leichte-Sprache KI*, *Leichte Sprache Übersetzer*, *Capito* und *einfachesprache.xyz* wiesen Sinneinheiten pro Zeile auf. Diese Textstrukturierung erleichtere die Verständlichkeit der Texte. Eine solche Strukturierung fehlte bspw. sowohl bei dem Ausgangstext „PayPal-Käuferschutz" als auch bei dessen Übersetzung durch *DeepLWrite*.

Ausgangstext:

Heute beginnt die PIN AG mit der Zustellung der Abstimmungsbenachrichtigungen an die Berliner Stimmberechtigten für den Volksentscheid Berlin 2030 klimaneutral. Spätestens am Samstag, dem 4. März 2023, haben dann alle stimmberechtigten Berlinerinnen und Berliner ihre Abstimmungsbenachrichtigung im Briefkasten.

DeepLWrite:

Heute beginnt die PIN AG mit der Zustellung der Abstimmungsbenachrichtigungen an die Berliner Stimmberechtigten für den Volksentscheid Berlin 2030 klimaneutral. Spätestens am Samstag, dem 4. März 2023, haben alle Berlinerinnen und Berliner ihre Abstimmungsbenachrichtigung im Briefkasten.

Dahingegen strukturierte das Tool *Leichte Sprache Übersetzer* den generierten Text nach Sinneinheiten pro Zeile. Das Tool fügte zusätzlich noch Zwischenüberschriften ein.

Leichte Sprache Übersetzer:

Informationen zum Volks-Entscheid
Heute beginnt die PIN AG.
Sie schickt Briefe an Berliner.
Es geht um den Volks-Entscheid.
Der Volks-Entscheid ist: Berlin soll 2030 klimaneutral sein.

Wichtige Daten
Bis Samstag, 4. März 2023, bekommen alle Berliner ihren Brief.
Im Brief steht, wie sie abstimmen können.

Verbalstil

Für das bessere Textverständnis sieht die DIN SPEC 33429 vor, Texte im Verbalstil zu verfassen und Nominalphrasen in einzelne Sätze zu zerlegen (vgl. DIN e.V. 2025, 19). Besonders die auf Leichte Sprache spezialisierten KI-Tools wie *Optimeil-Leichte-Sprache KI* und *Leichte Sprache Übersetzer* wendeten einen Verbalstil in ihren Texten an. Aber auch das auf Einfache Sprache spezialisierte KI-Tool *einfachesprache.xyz* gab Texte im Verbalstil aus. Die Texte der KI-Tools *DeepLWrite*, *Klartext* und *Copilot* wiesen dahingegen vorrangig einen Nominalstil auf.

Während die qualitative Analyse bestimmte KI-Tools also als besonders verbalstilorientiert beschreibt, zeigt die quantitative Analyse durch die POS-Tag-Auswertung, dass alle getesteten Modelle – einschließlich dieser Tools – einen höheren nominalen Anteil aufweisen. Diese Diskrepanz legt nahe, dass die Kategorisierung in Nominal- und Verbalstil auf rein statistischer Ebene möglicherweise zu abstrakt und somit wenig zielführend ist, wenn sie nicht alle syntaktischen und semantischen Relationen berücksichtigt. Ein hoher Anteil an Verben führt nicht automatisch zu mehr Verständlichkeit, ebenso wie ein hoher Anteil an Nomen nicht zwangsläufig zu komplexen Strukturen führt. Es bleibt demnach fraglich, ob die POS-Tag-basierte Einteilung allein ausreicht, um den tatsächlichen Stil eines Textes korrekt zu erfassen.

Das folgende Beispiel zeigt, inwieweit im Ausgangstext „RKI" und in dessen Übersetzung durch *Copilot* ein Nominalstil durch die Reihung von Substantiven zu erkennen ist:

Ausgangstext:

Die Kernaufgaben des RKI sind die Erkennung, Verhütung und Bekämpfung von Krankheiten, insbesondere der Infektionskrankheiten.

Copilot:

Es kümmert sich um die Überwachung und Vorbeugung von Krankheiten, besonders um ansteckende Krankheiten.

Im Gegensatz dazu löste *Optimeil-Leichte-Sprache KI* diese komplexe Nominalphrase in mehrere Sätze auf und lässt jenen Verbalstil erkennen, der den Empfehlungen der DIN SPEC 33429 entspricht.

Optimeil-Leichte-Sprache KI:

Das RKI hat eine wichtige Aufgabe:
Es beobachtet Krankheiten.
Es hilft, Krankheiten zu verhindern.
Es hilft, Krankheiten zu bekämpfen.
Besonders wichtig sind ***Infektions-Krankheiten****.*

Aktiv

Die Verwendung des Aktivs wird von der DIN SPEC 33429 bevorzugt – insbesondere mit Nennung des Handlungsträgers (vgl. DIN e.V. 2025, 21).

Die KI-Tools *Capito* und *Wortliga* formulierten Texte überwiegend im Aktiv. Dahingegen enthielten die Texte von *ChatGPT*, *DeepLWrite*, *Klartext*, *Language Tool* und *einfachesprache.xyz* häufiger Passivkonstruktionen. Somit unterstützt die qualitative Analyse die Ergebnisse der quantitativen Untersuchung – auch wenn es sich bei der quantitativen Analyse um keinen „echten" Passiv-Nachweis handelt. Eine bloße Auszählung der eigens festgelegten Prädiktor-Schlagworte bezeugt nur, dass die Wortformen vorkommen, belegt jedoch nicht, ob sie auch tatsächlich in einer Passivkonstruktion stehen. Auch die Mehrdeutigkeit von „werden" – einerseits ist es Passiv-Hilfsverb, andererseits bildet es auch Futur-Formen und fungiert als Vollverb – bleibt zu hinterfragen. Andere Passivformen, so beispielsweise Konstruktionen mit „wird", sind zudem nicht Teil der Abfrage.

Verdeutlicht wird die Nutzung von Passivkonstruktionen in den KI-Übersetzungen am Beispiel des Tools *einfachesprache.xyz*, das weitgehend die Passivkonstruktionen des Ausgangstextes „Special Olympics" übernahm.

Ausgangstext:

Es wurde gejubelt und gefeiert, auch mal geweint, wenn ein Spiel verloren gegangen war, es wurde getröstet und weitergefeiert, es wurden weltweite Freundschaften geknüpft.

Einfachesprache.xyz:

Es wurde viel gejubelt und gefeiert.
Manchmal wurde auch geweint, wenn ein Spiel verloren war.
Aber dann wurde getröstet und weiter gefeiert.
Die Sportler haben Freundschaften aus der ganzen Welt geschlossen.

Im Vergleich dazu formte *Capito* diese Passivkonstruktionen ins Aktiv um, indem das Tool „Menschen", die bei den Special Olympics anwesend waren, zu den Subjekten der Sätze machte.

Capito:

Die Menschen freuten sich und feierten.
Manchmal weinten sie, wenn ein Spiel verloren war.
Sie trösteten sich und feierten weiter.
Sie fanden Freunde aus der Welt.

Indikativ

Für die leichte Erfassung des Textes sieht die DIN SPEC 33429 Formulierungen im Indikativ vor (vgl. DIN e.V. 2025, 21f.).

Nur vereinzelt haben die KI-Tools in ihren Übersetzungen den Konjunktiv benutzt. Aber diese verwendeten ihn auch teilweise dann, wenn im Ausgangstext kein Konjunktiv vorhanden war, so zum Beispiel die KI-Tools *Gemini*, *einfachesprache.xyz*, und der *ChatGPT-Bot „Klar und Verständlich"*. Auch hier unterstützt die qualitative Analyse die Ergebnisse der quantitativen Untersuchung. Dennoch bleibt zu sagen, dass die CQL-Abfrage aus der quantitativen Analyse ausschließlich die vorkommenden Konjunktivformen auf Wortebene zählt. Sie erfasst dabei lediglich Stellen, an denen die hinterlegten Wortformen im Text auftreten – unabhängig davon, ob sie tatsächlich als Konjunktiv verwendet werden oder in einem anderen grammatischen Zusammenhang stehen.

Im Beispiel wird erkenntlich, dass das Tool *ChatGPT-Bot „Klar und Verständlich"* einen Konjunktiv durch die Formulierung „das wäre ein Erfolg" in den Satz einfügt, obwohl ursprünglich im Ausgangtext „Special Olympics" keiner vorhanden war.

> Ausgangstext:
>
> *Die wahrlich erfolgreichen, stimmungsvollen und beeindruckenden Weltspiele in Berlin sind ein ideales Instrument, in der Gesellschaft etwas zu bewegen und um Special Olympics und der gesamten Inklusion in Deutschland einen notwendigen Schub zu verleihen*.
>
> ChatGPT-Bot „Klar und Verständlich":
>
> *Die Weltspiele waren ein starkes Zeichen. Doch jetzt kommt es darauf an,* ***dass sich wirklich etwas ändert. Mehr inklusive Sportangebote, mehr Begegnungen, mehr Möglichkeiten für alle*** *– das wäre ein echter Erfolg.*

Zeitformen

Die DIN SPEC 33429 empfiehlt die Zeitformen, „die in der mündlichen Sprache verwendet werden" (DIN SPEC 33429, 22). Das sind vor allem Präsens, Perfekt und das Präteritum von Hilfs- und Modelverben (vgl. DIN e.V. 2025, 22f.).

In den übersetzten Texten lagen verschiedene Zeitformen vor, auch meist bedingt durch die variierenden Zeitformen im Ausgangstext. Dennoch gab es bei den Texten der Allzweck KI-Tools *Gemini*, *ChatGPT* und *Copilot* sowie der auf Einfache Sprache spezialisierten KI-Tools *ChatGPT-Bot „Klar und Verständlich"*, *Klartext* und *Language Tool* die meisten Variationen.

Folgender Textausschnitt bietet eine hohe Variation an Tempusformen wie Präsens, Präteritum und Futur I:

> *Der Einfluss des Großevents auf die Gesellschaft wird sich erst in den kommenden Monaten und Jahren zeigen, aber der Grundstein wurde zweifellos gelegt. Jetzt gilt es, den Schwung der Weltspiele in den Alltag zu tragen. Die Chancen für nachhaltige Veränderungen stehen gut.*

Das Tool *ChatGPT* tauschte zwar das Präteritum gegen eine Perfekt-Konstruktion ein, aber es reduzierte damit nicht die Anzahl verschiedener Zeitformen.

> ChatGPT:
>
> ***Die Zukunft sieht gut aus***
> *Es wird noch etwas dauern, bis man sieht, was die Spiele wirklich gebracht haben. Aber der Anfang ist gemacht. Jetzt müssen alle dafür sorgen, dass es so weitergeht. Die Chancen für Veränderungen sind gut.*

Dahingegen vermied *Optimeil-Leichte-Sprache KI* das Futur I im generierten Text, wodurch lediglich die Zeitformen Präsens und das Präteritum für das Hilfsverb „sein" vorlagen.

Optimeil-Leichte-Sprache KI:

Die Weltspiele waren ein großer Erfolg.
Sie helfen, dass mehr Menschen über Inklusion nachdenken.
Jetzt ist es wichtig:
Der Schwung der Spiele muss bleiben.
Dann kann sich wirklich etwas verändern.

Textgliederung

Die DIN SPEC 33429 misst der Textgliederung für die Informationserfassung eine besondere Bedeutung bei (vgl. DIN e.V. 2025, 23, 25).
Die meisten getesteten KI-Tools strukturierten ihre Übersetzungen mit Absätzen und Hervorhebungen der wichtigsten Informationen. Lediglich bei den Übersetzungen von *Klartext*, *DeepLWrite* und *Language Tool* lag keine Textgliederung vor. Dies ist bspw. bei der *Klartext*-Übersetzung des „Berliner Volksentscheides" erkennbar, jedoch war der Ausgangstext in Absätze gegliedert.

Ausgangstext:

Stimmberechtigte, die am Abstimmungstag verhindert sind, können vorab per Brief abstimmen. Hierzu kann ab heute ein Abstimmungsschein beantragt werden. Der Antrag kann online über die Website des Landeswahlleiters gestellt werden. Wer die Zustellung der Abstimmungsbenachrichtigung abwartet, kann die Abstimmungsbenachrichtigung via Scan des dort auf der Rückseite eingedruckten QR-Codes vornehmen. Durch das Scannen des QR-Codes mit einer geeigneten App wird der Abstimmungsscheinantrag im Internet aufgerufen und die verschlüsselten personenbezogenen Daten werden automatisch eingetragen. Der Antrag muss dann nur noch bestätigt werden. Die Abstimmungsunterlagen werden vom Bezirkswahlamt nach Hause oder an die angegebene Wunschadresse geschickt.
Neben der Onlinebeantragung besteht auch die Möglichkeit durch Rücksenden der ausgefüllten Abstimmungsbenachrichtigung oder schriftlich – auch per E-Mail – die Briefabstimmungsunterlagen anzufordern.

Klartext:

Wer am Abstimmungstag nicht kann, hat die Möglichkeit, per Brief zu stimmen. Dafür kann man ab heute einen Abstimmungsschein online beantragen. Alternativ kann man die Briefabstimmungsunterlagen per E-Mail anfordern oder in einer Briefabstimmungsstelle abstimmen.

Der *ChatGPT-Bot „Klar und Verständlich"* strukturierte gemäß der DIN SPEC 33429 auch die einzelnen Absätze in ihre zentralen Punkte. Zusätzlich hob dieses KI-Tool auch die wichtigsten Informationen hervor.

ChatGPT-Bot „Klar und Verständlich":

Briefabstimmung möglich

Wer am Abstimmungstag (26. März 2023) nicht ins Wahllokal gehen kann, darf vorher per Brief abstimmen. Dafür muss ein ***Abstimmungsschein*** *beantragt werden*

So funktioniert die Briefabstimmung:

*- **Online-Antrag:** Ab sofort kann der Abstimmungsschein auf der Website des Landeswahlleiters beantragt werden.*
*- **Per QR-Code:** Wer die Abstimmungsbenachrichtigung abwartet, kann den dort aufgedruckten QR-Code scannen und den Antrag direkt online ausfüllen.*
*- **Per Post oder E-Mail:** Man kann den Antrag auch schriftlich oder per E-Mail stellen.*

Sinn

In der qualitativen Untersuchung wurde auch geprüft, ob bei den übersetzten Texten der Sinn des Ausgangstextes weiterhin vorhanden ist. Dieser war häufig bei den Übersetzungen

folgender Tools entstellt (in aufsteigender Häufigkeit): *WELTgo!*, *Copilot*, *Language Tool*, *Leichte Sprache Übersetzer*, *Hey_Die KI von Bild*, *Wortliga* und *Capito*.

Zum Beispiel reduzierte zwar das Tool *Wortliga* die schwer verständlichen Wörter und Satzstrukturen des Ausgangstextes „Berliner Volksentscheid“ gemäß den Standards der DIN SPEC 33429. Allerdings geht bei dieser Übersetzung der ursprüngliche Sinn verloren, dass entweder die Abstimmung vorher digital/schriftlich oder vor Ort in einer Abstimmungsstelle erfolgen konnte. Das KI-Tool verallgemeinerte den Text so, dass all diese Schritte zu machen seien, was zu einem verzerrten Verständnis des Textes führt.

Ausgangstext:

Neben der Onlinebeantragung besteht auch die Möglichkeit durch Rücksenden der ausgefüllten Abstimmungsbenachrichtigung oder schriftlich – auch per E-Mail – die Briefabstimmungsunterlagen anzufordern.
Zudem besteht die Möglichkeit, in einer Briefabstimmungsstelle ihres Bezirkes abzustimmen. Anschriften und Öffnungszeiten der Briefabstimmungsstellen finden Sie auf der Abstimmungsbenachrichtigung.

Wortliga:

Schicke den Brief zurück. Schreibe einen Brief oder eine E-Mail. Dann bekommst du die Unterlagen.
Gehe zur Abstimmungsstelle. Die Adresse steht im Brief. Die Zeiten stehen auch dort.

Zusammenfassung der qualitativen Gesamtergebnisse

Kategorien DIN SPEC 33429/ Text	zentraler Wortschatz	kurze Wörter	Zerlegung Wörter	Erläuterungen	keine Synonyme	keine Metaphern	kurze Sätze	einfache Satzstruktur	Sinneinheiten	Verbalstil	Aktiv	Indikativ	Präsens	Textgliederung	Sinn	Mittelwert	Summe
Optimell-Leichte-Sprache KI	4	4,5	3,5	3	5	4	5	4,5	5	5	4	5	2	5	5	4,3	64,5
Leichte Sprache Übersetzer	2	2,5	3,5	2,5	4	5	5	4	5	5	4,5	5	2	5	3	3,8666667	58
St. Pauli bzw. einfachesprache.xyz	2	2,5	2	2	4	3	2,5	1,5	3,5	4,5	2,5	4,5	2	5	5	3,1	46,5
ChatGPT	1,5	1,5	1,5	3	4	4	4	2,5	0	3,5	2,5	5	1,5	5	5	2,9666667	44,5
Hey_Die KI von Bild	2	2	0	2	5	4	4	2	0	4,5	3,5	5	2	5	3	2,9333333	44
Wortliga	3,5	3	0	1,5	3	5	2,5	2,5	0	4	4,5	4,5	2	5	2	2,8666667	43
Capito.ai	0,5	1,5	0	2,5	3	2	5	4	4	4,5	5	4,5	2	3	1	2,8333333	42,5
Gemini	2,5	2	0	4,5	4	4	3	1	0	3,5	3	4	1,5	5	5	2,8666667	43
WELTgo!	1,5	1	0	2,5	4	3	3,5	2	0	3,5	3,5	5	2	4	4	2,6333333	39,5
ChatGPT-Bot "Klar und Verständlic	0,5	1	0	2,5	2	4	2	0,5	0	2,5	2,5	3,5	1,5	5	5	2,1666667	32,5
Copilot	0,5	0,5	0	2,5	3	4	1	0,5	0	1	3	5	1,5	5	4	2,1	31,5
Klartext	0,5	1	0	1	5	3	2	1	0	1,5	2	5	2	0	5	1,9333333	29
DeepL Write	0,5	0,5	0	1,5	2	4	1	1	0	1	2,5	5	2	1	5	1,8	27
Language Tool	0,5	0,5	0	1,5	2	2	1,5	0,5	0	2,5	2	4,5	1,5	2	4	1,6666667	25

Tabelle 3: Qualitative Gesamtergebnisse

Anmerkung zur Qualität der KI-generierten Texte

Die Ausgangstexte besitzen jeweils folgende sprachliche Besonderheiten, mit denen die KI-Tools bei der Übersetzung in Leichte Sprache unterschiedlich umgingen:

- Käuferschutz von PayPal: schwierige, lange Wörter; schwierige Satzstruktur; Variation an Zeitformen
- Koloniale Raubkunst im Humboldt Forum: schwierige, lange Wörter; schwierige Satzstruktur; Passivkonstruktionen; Variation an Zeitformen; Zitation von direkter Rede
- Robert-Koch-Institut (RKI): schwierige, lange Wörter und schwierige Satzstruktur
- Special Olympics (2023): metaphorischer Stil; schwierige Satzstruktur; große Variation an Zeitformen; Passivkonstruktionen
- Volksentscheid Berlin 2030 klimaneutral: schwierige, lange Wörter und schwierige Satzstruktur; Passivkonstruktionen

Die Analyse zeigte auch, dass die Qualität der KI-Übersetzungen zu Teilen von der Textstruktur des Ausgangstextes und nicht nur von dessen Inhalt abhängt. So schnitten bspw. die Übersetzungen des RKI-Textes deutlich besser als die des Special-Olympic-Textes ab. Letzterer war – wie oben gelistet – durch einen sehr metaphorischen Stil, eine große Variation an Zeitformen und einige Passivkonstruktionen geprägt. All dies wurde teils auch von den KI-Tools übernommen. Dahingegen war der RKI-Text inhaltlich zwar anspruchsvoller, jedoch grammatikalisch und stilistisch einfacher, was womöglich zur besseren KI-Übersetzung führte. Das könnte daran liegen, dass KIs für Einfache und Leichte Sprache vorrangig darauf trainiert sind, schwierige

Wörter und komplexe Satzkonstruktionen aufzulösen als grammatikalische und stilistische Standards nach der DIN SPEC 33439 einzuhalten.

C. Zusammenfassung der Ergebnisse und Ausblick

Zusammenfassend machte die Analyse sowohl quantitativ als auch qualitativ Unterschiede zwischen den KI-Tools deutlich. Die auf Leichte Sprache spezialisierten KI-Tools wie *Optimeil-Leichte-Sprache KI* und *Leichte Sprache Übersetzer* erzeugten meist Texte, die in hoher Übereinstimmung zur DIN SPEC 33429 stehen. Dies liegt insbesondere an der konsequenten Einhaltung der Standards auf der Wort- und Satzebene sowie in der Grammatik. Die Allzweck-KI-Tools wie *ChatGPT* und *Gemini* lieferten ebenfalls weitgehend normkonforme Texte, jedoch setzten diese Tools trotz Prompt nicht alle Anforderungen vollständig um. Die auf Einfache Sprache spezialisierten KI-Tools wichen mit ihren übersetzten Texten stärker von der DIN SPEC 33429 ab, was sich darauf zurückführen lässt, dass diese eher auf eine einfache Ausdrucksweise statt auf die spezifischen Regeln der Leichten Sprache trainiert sind. Eine Ausnahme bildet *einfachesprache.xyz*, dessen Texte in vielen Aspekten mit der DIN SPEC 33429 übereinstimmten. Somit stellen sich die spezialisierten KI-Tools für Leichte Sprache für die Erstellung verständlicher Übersetzungen in Leichter Sprache als besonders geeignet heraus. Dennoch ist selbst bei diesen Tools eine menschliche Kontrolle des übersetzten Textes notwendig, auch um Sinnentstellungen und Fehler zu erkennen und zu korrigieren.

Denn obwohl die spezialisierten KI-Tools für Leichte Sprache viele formale Anforderungen der DIN SPEC 33429 erfüllen, bleibt die Frage offen, inwiefern sie auch die funktionale Verständlichkeit sicherstellen können. Unsere Analyse konzentrierte sich auf messbare sprachliche Merkmale wie Satzlänge, Wortwahl oder die Vermeidung von Nominalstil; auf Regeln standardisiert nach DIN SPEC 33429. Doch Verständlichkeit ist mehr als die Einhaltung sprachlicher Regeln – sie hängt auch davon ab, ob die Texte tatsächlich von der Zielgruppe verstanden werden. Textgestaltung allein schafft keine Verständlichkeit. Abhängig ist diese vielmehr von den individuellen Bedürfnissen, kognitiven Prozessen und Vorkenntnissen der Rezipierenden. Zukünftig sehen wir es daher als sehr relevant, die Frage der funktionalen Verständlichkeit zusätzlich empirisch zu überprüfen – etwa durch Lesestudien mit der Zielgruppe. Gleichzeitig sollte weiterhin in neue Ansätze zur inklusiven Kommunikation investiert werden, die über die reine Anwendung von Sprachregeln hinausgehen – auch in Hinblick auf KI-gestützte Übersetzungssysteme, die sicherlich das Potential bieten, zukünftig stärker auf individuelle Verständnishorizonte der Lesenden einzugehen.

Quellen und Literatur

Ressourcen:

Der Landeswahlleiter für Berlin (2023): Volksentscheid Berlin 2023 klimaneutral- Briefabstimmungszeitraum beginnt heute. In: Berlin. URL: https://www.berlin.de/wahlen/pressemitteilungen/2023/pressemitteilung.1293868.php#kontakt (zuletzt abgerufen am 17.03.2025).

König, Jürgen (2022): Das Humboldt Forum und die koloniale Raubkunst. In: Deutschlandfunk. URL: https://www.deutschlandfunk.de/koloniale-raubkunst-100.html (zuletzt abgerufen am 17.03.2025).

PayPal (2022): PayPal-Käuferschutz, In: PayPal. URL: https://www.paypal.com/de/legalhub/paypal/buyer-protection (zuletzt abgerufen am 17.03.2025).

Robert Koch-Institut (o.D.): Institut. In: Robert Koch Institut. URL: https://www.rki.de/DE/Institut/institut-node.html zuletzt abgerufen am 17.03.2025).

Special Olympics World Games Berlin 2023 Organizing Committee gGmbH (2023): Erfolgreiche Weltspiele und der Weg in den Alltag. In: Special Olympics World Games Berlin 2023. URL: https://www.berlin2023.org/de/news/erfolg-fuehrt-zu-verantwortung-im-alltag (zuletzt abgerufen am 17.03.2025).

Quellen:

DIN Deutsches Institut für Normung e.V. (2025): DIN SPEC 33429. Empfehlungen für Deutsche Leichte Sprache. Berlin: DIN Media GmbH.

Manning, Sabine (2024): KI-Tools für Einfache Sprache: Leistungen von 10 Tools auf einen Blick. In: Multisprech. URL: https://multisprech.org/2024/05/22/ki-tools-fur-einfache-sprache-leistungen-von-10-tools-auf-einen-blick/?utm_source=chatgpt.com (zuletzt abgerufen am 17.03.2025).

o. A.: Grundwortschatz 500. URL: https://www.gut1.de/grundwortschatz/grundwortschatz-500/ (zuletzt abgerufen am 14.03.2025).

Rosengren, Inger (1976): Der Grundwortschatz als theoretisches und praktisches Problem, in: Probleme der Lexikologie und Lexikographie : Jahrbuch 1975 des Instituts für deutsche Sprache, 313-333. URL: https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/1345/file/Rosengren_Der_Grundwortschatz_als_theoretisches_und_praktisches_Problem_1976.pdf (zuletzt abgerufen am 14.03.2025).

Abbildungs- und Tabellenverzeichnis

Abbildung 1: Mittlere Wortlänge aufgeschlüsselt nach Tool und Textart 6
Abbildung 2: Mittlere Wortlänge aufgeschlüsselt nach Tool 6
Abbildung 3: Mittlere Satzlängen aufgeschlüsselt nach Tool und Textart 7
Abbildung 4: Mittlere Satzlängen aufgeschlüsselt nach Tool 7
Abbildung 5: Anteil Grundwortschatz pro Tool und Textart 8
Abbildung 6: Anteil Grundwortschatz pro Tool 9
Abbildung 7: Stilanteil pro Tool 10
Abbildung 8: Passivkonstruktionen pro Tool 10
Abbildung 9: Konjunktivanteil pro Tool 11

Tabelle 1: Übersicht der KI-Tools unterteilt in die drei Typen 2
Tabelle 2: Übersicht der verwendeten Modi und Sprachniveaus aufgeschlüsselt nach jeweiligem KI-Tool 4
Tabelle 3: Qualitative Gesamtergebnisse 21

Dank

Unser herzlicher Dank gilt der VERSO Dresden gGmbH, insbesondere Juliane Heidelberger, für die Möglichkeit, diesen Artikel im Rahmen unseres Praktikums schreiben zu dürfen sowie für ihre fachliche Begleitung und umfassende Unterstützung während der gesamten Bearbeitung.

Autorinnen

Franziska Ludewig absolvierte ihren Bachelor der Geschichts- und Politikwissenschaft 2023 an der Martin-Luther-Universität Halle-Wittenberg. Aktuell studiert sie an der Universität Bielefeld Geschichtswissenschaft im Master. Während ihres Praktikums bei VERSO konzipierte und leitete sie eine Veranstaltung zur Bundestagswahl 2025 in Leichter Sprache mit entsprechenden Zielgruppen. Zudem analysierte sie in Zusammenarbeit mit Emily Reich verschiedene KI-Tools für Leichte Sprache.

Emily Reich studierte Germanistik an der TU Chemnitz und setzt seit 2023 ihr Studium im Master Digital Humanities an der TU Dresden fort. Im Rahmen ihres Studiums absolvierte sie ein Praktikum bei Verso im Bereich Barrierefreie Kommunikation. Dort beschäftigte sie sich schwerpunktmäßig mit der statistischen Analyse von Leichter und Einfacher Sprache und der Anwendung bürgernaher Sprache in Verwaltungskontexten. Weiterhin untersuchte sie, wie Künstliche Intelligenz und Leichte Sprache zusammenwirken – ein Projekt, das sie gemeinsam mit Franziska Ludewig umsetzte.