

Tagungsbericht: Auftakttagung zum Projekt „KI-GesKom“

Das Projekt „KI-GesKom“ (KI-gestützte Gesundheitskommunikation in Einfacher Sprache) startete im Mai 2024. Es ist ein Projekt zwischen der Forschungsstelle Leichte Sprache (Universität Hildesheim), der Apotheken Umschau (Wort & Bild Verlag) und SUMM AI GmbH. Am 30. August fand die hybride Auftakttagung in den Räumlichkeiten des Wort & Bild Verlags statt. Sarah Ahrens (Universität Hildesheim) und Dr. Julian Hörner (Wort & Bild Verlag) führten durch den Tag.

Die Auftakttagung bestand aus sechs Vortragsslots und einer abschließenden Diskussionsrunde. Der vorliegende Tagungsbericht dient der Tagungsdokumentation und enthält die Aussagen aus den Impulsvorträgen, Fragerunden und der Diskussionsrunde, die im Lichte des Projekts KI-GesKom besonders wichtig sind.

KI-GesKom – KI-gestützte Gesundheitskommunikation in Einfacher Sprache (Ekaterina Lapshinova-Koltunski, Universität Hildesheim, Julian Hörner, Wort & Bild Verlag, Vanessa Theel, SUMM AI GmbH)

Die Tagung begann mit einer Projektvorstellung durch Prof. Ekaterina Lapshinova-Koltunski (Universität Hildesheim), Dr. Julian Hörner (Wort & Bild Verlag) und Vanessa Theel (SUMM AI GmbH). Gesundheitskompetenz ist die Fähigkeit, gesundheitsbezogene Informationen zu finden, zu verstehen, zu beurteilen und anzuwenden. Die Gesundheitskompetenz von großen Teilen der deutschen Bevölkerung ist eingeschränkt (Hurrelmann et al. 2020) – deshalb schlägt der Nationale Aktionsplan Gesundheitskompetenz (Schaeffer et al. 2018) z. B. Einfache Sprache vor, um Gesundheitsinformationen leichter zugänglich zu machen. Allerdings sind für den Bedarf an einfachen Gesundheitsinformationen derzeit nicht genug Übersetzer*innen verfügbar. Hier setzen der Einsatz von KI und auch das Projekt „KI-GesKom“ an. Es geht folgenden Fragen nach:

- Kann die intralinguale Übersetzung so wie die interlinguale Übersetzung von KI profitieren?
- Sind die Anforderungen im Rahmen der Leichten und Einfachen Sprache zu individuell, als dass ein Tool sie lösen könnte?
- Ist KI ein Tool für professionelle intralinguale Übersetzer*innen, für Content-Provider oder für Endnutzer*innen?
- Welche Kompetenzen müssen Nutzer*innen haben, um mit KI zu übersetzen?
- Wie performen existente KI-Tools?

Dazu wird im Projekt das Tool SUMM AI genutzt, das erste maschinelle Übersetzungstool für intralinguale Übersetzung in Leichte und Einfache Sprache. Ein Testdatensatz von 30 Standardtexten von www.apotheken-umschau.de und deren Humanübersetzungen in Einfacher Sprache, der „human gold standard“, wurden mit KI-generierten Übersetzungen verglichen. Die KI-generierten Texte wurden mit je drei SUMM-AI-Modellen generiert: dem „Baseline Model“, dem ein Large Language Model (LLM) ohne Fine-Tuning zugrunde liegt, und Model 1 und 2, denen unterschiedliche LLMs mit Fine-Tuning zugrunde liegen. Die Texte wurden in ihrer Korrektheit, Lesbarkeit, Satzkomplexität und des erreichten [SARI-Scores](#) verglichen. SARI

vergleicht die Qualität der Wörter, die das maschinelle Vereinfachungssystem hinzufügt, löscht oder beibehält.

Unter den KI-Modellen erzielt Model 2 nach den meisten Kriterien die besten Ergebnisse. Die Übersetzung mit allen drei Modellen birgt die Gefahr von Fehlinformationen. Wegen der eingeschränkten Gesundheitskompetenz in großen Teilen der Bevölkerung sind korrekte Informationen notwendig. KI-generierte Texte können von Endnutzer*innen also nicht sicher verwendet werden und es ist ein Post-Editing nötig. Die KI-Tools sind vielmehr Übersetzungstools für professionelle Übersetzer*innen und Content-Provider mit Fachkenntnissen in Einfacher Sprache.

Die Ergebnisse sind nachzulesen in Deilen et al. (2024).

Grußwort (Dennis Ballwieser, Wort & Bild Verlag)

Dr. Dennis Ballwieser, Geschäftsführer der Wort & Bild Verlagsgruppe, unterstrich die Relevanz verständlicher Informationen im medizinischen Bereich. Das Projekt KI-GesKom ist für den journalistischen Ansatz relevant, nach dem die schnelle Erstellung von korrekten und verständlichen Texten im Zentrum steht.

Gesundheitskompetenz als zentraler Prädiktor für die eigene Gesundheit (Kai Kolpatzik, Wort & Bild Verlag)

Prof. Kai Kolpatzik ging im ersten Impulsvortrag auf die Effekte geringer Gesundheitskompetenz auf die Gesundheit von Individuen ein. So haben Menschen mit geringer Gesundheitskompetenz ein höheres Risiko, früher zu sterben als Menschen mit hoher Gesundheitskompetenz. Sie erhalten ihre Diagnosen später, melden sich öfter und länger krank, bewegen sich weniger und essen weniger Obst und Gemüse.

Insbesondere das Beurteilen von Gesundheitsinformationen fällt den meisten Menschen schwer (Hurrelmann et al. 2020). Im Internet gibt es viele verschiedene Gesundheitsinformationen, deren Qualität für Lai*innen nur schwer zu beurteilen ist. Hier setzt die App „VIER400“ an, die den Einstieg in evidenzbasierte Informationen erleichtert. Die App hat das Ziel, die digitale Gesundheitskompetenz von Unternehmensmitarbeiter*innen zu stärken. Neben der Gesundheitskompetenz sind bei der digitalen Gesundheitskompetenz z. B. auch die Medienkompetenz und der Umgang mit personenbezogenen Daten wichtig. VIER400 stellt KI-gestützt evidenzbasierte Informationen zur Verfügung, wobei die Lebensumstände und Interessen der Nutzer*innen berücksichtigt werden. Die Nutzer*in gibt ein Stichwort ein, z. B. „Diabetes“, und erhält allgemeine Informationen dazu sowie vier Quellenvorschläge, die je in 400 Zeichen zusammengefasst sind.

Verständlichkeitsoptimierte Gesundheitskommunikation: Sind Zusammenfassungen klinischer Studien für Laien einfach genug? (Giulia Pedrini, Università degli Studi di Trieste)

Dr. Giulia Pedrini stellte ihre Forschung zu „Plain Language Summaries“, also den Zusammenfassungen von Ergebnissen klinischer Studien für Lai*innen, vor. Sie verglich drei Korpora:

englische, italienische und deutsche Zusammenfassungen. Die Zusammenfassungen werden auf Englisch verfasst und in die Sprachen der Mitgliedsstaaten übersetzt, die an der Durchführung der Studien beteiligt waren. Die Zusammenfassungen sind die einzige medizinische Einfache-Sprache-Textsorte, die eine Rechtsgrundlage *und* spezifische Leitfäden haben.

Die Forschungsfragen lauteten im Kern wie folgt:

- Können die Zusammenfassungen als Beispiele für Einfache Sprache gelten?
- Welche sind die Hauptmerkmale der Zusammenfassungen?
- Werden die geltende Rechtsgrundlagen und Leitfäden bei der Erstellung der Zusammenfassungen beachtet?
- Könnten die existierenden Richtlinien verbessert werden?

Untersucht wurden in einem Korpus von 180 Zusammenfassungen z. B. die Worthäufigkeit und Anzahl an Nebensätzen sowie die Existenz von Fachtermini mit griechischen oder lateinischen Wurzeln. Die Analyseergebnisse zeigen, dass die Zusammenfassungen im Allgemeinen zu komplex sind, um von Lai*innen verstanden zu werden. Sie sind der Standardsprache näher als der Einfachen Sprache.

Dr. Pedrini verweist auf ihr neu erschienenes Buch „[Medical communication between Plain Language and Einfache Sprache](#)“.

Wie funktioniert KI für Textvereinfachung? (Miriam Anschütz, Technische Universität München)

Miriam Anschütz nutzte ihren Impulsvortrag, um die Funktionsweise KI-Modellen für die Textvereinfachung zusammenzufassen. Texte werden generiert, indem das je nächstwahrscheinliche Wort berechnet wird, wobei der Kontext die Auswahl der Worte beeinflusst:

Ich hab dich → Ich hab dich lieb

Die maschinelle Übersetzung in Leichte und Einfache Sprache hat Elemente von Zusammenfassung und der Übersetzung – zwei unterschiedlichen Aufgaben für KI. Der Übersetzungsprozess umfasst die Reduktion, aber auch die Addition von Informationen – nur die für die Zielgruppe relevanten Informationen bleiben im Text und Erklärungen werden hinzugefügt. Die Sprache der Texte wird angepasst, indem eine Komplexitätsreduktion stattfindet. Dies geschieht durch Encoder-Decoder-Modelle. Der Encoder-Teil entnimmt dem Ausgangstext die Kernaussagen, also den Inhalt. Der Decoder-Teil formuliert diese Kernaussagen neu und in anderen Worten.

Probleme können bei der Interpretation der Kernaussagen auftreten, z. B. bei der Auswahl der tatsächlich relevanten Informationen. Aber auch bei der Generierung der vereinfachten Texte und bei der Erklärung von Konzepten können Schwierigkeiten auftreten. Derzeit werden KIs mit standardsprachlichen Texten trainiert und deshalb „kennt“ die KI die Einfache Sprache nicht gut. Sie generiert sie deshalb nicht fließend und bewertet sie nicht vorurteilsfrei.

Post-Editing: Potenziale, Richtlinien, Risiken und Kompetenzen (Jean Nitzke, Universität i Agder)

Prof. Jean Nitzke stellte die Prinzipien des interlingualen Post-Editings vor. Da die automatische Übersetzung jeglicher Art nicht fehlerfrei erfolgt, müssen professionelle Übersetzer*innen den maschinenübersetzten Text nachbearbeiten. Dabei werden unterschiedliche Qualitätsstandards angesetzt, z. B. einen Text zu erstellen, dessen Bedeutung man entnehmen kann, oder einen natürlich wirkenden Zieltext zu erstellen. Diese Nachbearbeitung nach bestimmten Qualitätsstandards heißt „Post-Editing“. Die maschinelle Übersetzung und das Post-Editing sollen den Übersetzungsprozess schneller und kostengünstiger machen. Dabei soll die Qualität des Zieltextes nicht sinken (nur innerhalb vorgegebener Standards). Es dürfen aber auch keine sensiblen oder personenbezogenen Daten in die Maschine gegeben werden, und es besteht die Gefahr des Kontrollverlusts. Ungeklärt ist auch die Frage, wer bei Rechtskonflikten haftet, wenn ein Text maschinell übersetzt und post-editiert wurde. Der Übertrag der Post-Editing-Prinzipien auf die intralinguale Übersetzung in Leichte und Einfache Sprache erfolgten in der Diskussionsrunde.

Poster-Präsentationen

Im letzten Vortragsslot wurden drei Poster vorgestellt:

1. Vorstellung der bisherigen Ergebnisse aus KI-GesKom (Silvana Deilen, Sergio Hernández Garrido, Universität Hildesheim)
2. Redaktioneller Live-Betrieb mit KI bei Apotheken Umschau (Janina Kröger, Wort & Bild Verlag)
3. KI-Nutzung in der App VIER400 (Silke Heller-Jung, Berlin Health Media)

Die Ergebnisse aus KI-GesKom sind Abschnitt 1 „Projektvorstellung“ oder Deilen et al. (2024) zu entnehmen.

Der redaktionelle Live-Betrieb läuft neben SUMM AI auch mit Mistral AI. Die zu übersetzenden Gesundheitstexte von www.apotheken-umschau.de werden mit Mistral AI vorbereitet. Mistral AI strukturiert den Text nach einer vorgegebenen Makrostruktur. Die Vorstrukturierung wird von einer Redakteur*in geprüft, bevor der vorstrukturierte Text in SUMM AI eingegeben und in Einfache Sprache übersetzt wird. Im Post-Editing wird der KI-generierte Einfache-Sprache-Text sprachlich und inhaltlich geprüft. Der Text und seine Meta-Daten werden ins Content-Management-System eingepflegt und zuletzt wird der Text freigegeben. So werden 25.000 Zeichen in 4 Stunden übersetzt.

Die KI-Nutzung für die App VIER400 wurde in zwei Schritten analysiert. Ziel war es, zu prüfen, ob KI effizient und zuverlässig zu einem barrierefreien Angebot beiträgt. Die Outputs zweier KI-Modelle wurden anhand der Lesbarkeitsindizes, des Sprachniveaus nach dem Gemeinsamen Europäischen Referenzrahmen und der inhaltlichen Komplexität der Outputs verglichen. Beide KI-Modelle verringern die Textkomplexität, aber eines der Modelle erreichte sogar das Sprachniveau A2.

In einem zweiten Schritt ging es um die Entscheidung für die Nutzung von Leichter oder von Einfacher Sprache in der VIER400-App. Eines der KI-Modelle übersetzte die Texte zuverlässig

in Einfache Sprache (gleichgesetzt mit Niveau B1) und z. T. in Leichte Sprache Plus (A2). Auch die redaktionelle Nachbearbeitung bringt nicht alle Texte auf ein Leichte-Sprache-Plus-Niveau. Die Einfache Sprache ist deshalb die gewählte Varietät für VIER400.

Diskussionsrunde

Im Laufe des Tages ergaben sich aus den Fragerunden nach den Impulsvorträgen drei Schwerpunkte für die Diskussionsrunde. Das wichtige Thema „gendergerechte Sprache“ wurde ausgeklammert, da die Schnittstelle zwischen KI und Leichte Sprache im Zentrum des Interesses stand. Die drei Schwerpunkte waren folgende:

1. Was unterscheidet Leichte von Einfacher Sprache?
2. Was unterscheidet die in KI-GesKom verwendeten KI-Modelle voneinander?
3. Welche Post-Editing-Kompetenzen braucht es in der KI-gestützten intralingualen Übersetzung?

1 Was unterscheidet Leichte von Einfacher Sprache?

Die Unterscheidung zwischen Leichter und Einfacher Sprache ist aus wissenschaftlicher Sicht facettenreich, lässt sich aber in ihre Grundzüge herunterbrechen. Beides sind Varietäten des Deutschen, die sich u. a. in der intendierten Adressatenschaft unterscheiden. Das bedeutet, dass sie sich auch im Ausmaß der Verständlichkeitsoptimierung unterscheiden.

Die Adressat*innen der Leichten Sprache sind in erster Linie Menschen mit intellektueller Beeinträchtigung. Für Leichte Sprache gibt es feste Regeln, für Wortwahl, Grammatik, Syntax und Textgestaltung. Man erkennt Leichte-Sprache-Texte meist sofort, was ein Stigmatisierungspotenzial birgt – die Texte macht die Leser*innen als „schwache Leser*innen“ erkennbar. Wegen des Stigmatisierungspotenzials werden Leichte-Sprache-Texte nicht an Standardleser*innen herangetragen, sondern meist direkt an die Adressat*innen, z. B. in Wohneinrichtungen. Gesundheitstexte in Leichter Sprache gibt es z. B. auf www.patienten-information.de und von Aktion Mensch oder Special Olympics.

Einfache Sprache ist dynamischer und weniger regelgeleitet. Es stehen mehr sprachliche Mittel zur Verfügung – die Auswahl richtet sich nach den Kompetenzen der Adressat*innen und nach der intendierten Kommunikationssituation. Dringende Informationen, die in emotionalen Situationen nötig sind (z. B. die Gebrauchsanleitung auf einem Feuerlöscher) müssen leichter formuliert sein als Informationen, für die man viel Zeit und Ruhe hat.

Die Grenzen zwischen, Leichter, Einfacher und Standardsprache sind fließend. Zwischen der Leichten und der Einfachen Sprache ist die „Leichte Sprache Plus“ zu verorten. Für diese Varietät werden die Regeln der Leichten Sprache erweitert, damit keine stigmatisierenden Elemente im Text sind.

Es gibt Ansätze, z. B. von VIER400 oder auch Capito, die Leichte bzw. Einfache Sprache den Sprachniveaus nach dem Gemeinsamen Europäischen Referenzrahmen (GER) einordnen. Häufig wird der Leichten Sprache das Niveau A1 zugeordnet (also das Kompetenzniveau absoluter Sprachlernanfänger*innen), manchmal auch A2. Einfache Sprache ist i. d. R. B1 und Standardsprache B2. Aus wissenschaftlicher Sicht ist hier hinzuzufügen, dass der GER

Kompetenzziele für Fremd- und Zweitsprachler*innen formuliert, aber für ein Niveau keine spezifischen sprachlichen Elemente definiert. Auf den Niveaus A1 und A2 ist der Wortschatz der Lerner*innen darauf ausgelegt, über das alltägliche Umfeld zu sprechen (GER o. J.), aber nicht über Medizin und Gesundheit.

2 Was unterscheidet die in KI-GesKom verwendeten KI-Modelle voneinander?

In KI-GesKom werden drei KI-Modelle von SUMM AI verwendet:

- Baseline-Modell: ein Large Language Model (LLM), das mit einem prototypischen Datensatz vortrainiert wurde
- Modell 1 unterliegt dasselbe LLM wie dem Baseline-Modell. Modell 1 wurde zudem einem Fine-Tuning mit einem Datensatz von 170 Texten von www.apothekenumschau.de unterzogen.
- Modell 2 unterliegt ein anderes LLM und es fand dasselbe Fine-Tuning wie bei Modell 1 statt.

In der jetzigen Projektphase wird auch Output von ChatGPT (GPT-4o) analysiert. Es gibt einen wesentlichen Unterschied zwischen spezialisierten KI-Tools mit Fine-Tuning und allgemeinen KI-Tools, die mit Prompts arbeiten. Mit Fine-Tuning „weiß“ die KI, welche Art von Text sie generieren soll. Einem allgemeinen Tool muss dies in einem Prompt mitgeteilt werden. Der Vorteil von Programmen mit Fine-Tuning ist, dass man nicht vor jeden Vorgang ein oder mehrere Prompts eingeben muss.

KIs müssen lernen, gute Texte zu schreiben. Eine KI, die mit Prompts arbeitet, kann mit einer angehenden Textexpert*in verglichen werden, die gerade eine erste eintägige Leichte-Sprache-Schulung absolviert hat. Die KI kennt die Grundlagen, kann diese aber noch nicht sicher anwenden. Das Fine-Tuning hingegen kann mit der umfassenden Übersetzungspraxis verglichen werden, in der die Grundlagen bekannt sind und die Anwendung funktioniert.

Unsere ersten Ergebnisse aus dem Vergleich zwischen SUMM AI und ChatGPT zeigen, dass Tools mit Fine-Tuning (SUMM AI) bessere Ergebnisse erzielen als solche mit Prompting (ChatGPT). ChatGPT hat in fast allen Kriterien schlechter abgeschnitten als die SUMM-AI-Modelle.

Kann KI einen umfassend barrierefreien Leichte-Sprache-Text erstellen?

Barrierefreiheit hat viele Dimensionen, z. B. Bilder, Alternativtexte, funktionierende und richtig ausgezeichnete Links, multimediale Elemente mit ihren Barrierefreiheitsansprüchen usw. Ein KI-Tool müsste also viele unterschiedliche Aufgaben erfüllen können. Sie könnte z. B. für den Text passende Bilder auswählen, z. B. aus einer Bilddatenbank. Heute ist die Qualität KI-generierter Alternativtexte noch nicht ausreichend.

Ein Programm für alles ist nicht ausgeschlossen, es ist aber davon abzuraten. Stattdessen könnte man mehrere KI-Prozesse verbinden. Barrierefreiheit ist in die jeweilige Kommunikationssituation eingebunden, auf die das Tool individuell reagieren muss. Bspw. beeinflusst die Zielgruppe eines Textes, was unter „barrierefrei“ zu verstehen ist – ob z. B. mit vielen Visualisierungen gearbeitet wird oder lieber ohne.

3 Welche Post-Editing-Kompetenzen braucht es in der KI-gestützten intralingualen Übersetzung?

Für das Post-Editing braucht eine Expert*in Übersetzungskompetenzen, z. B. um alle Fehler im maschinell übersetzten Text zu finden. Zu diesen Übersetzungskompetenzen gehören sprachliche Kompetenzen, aber auch Sprachtransferkompetenzen, Fachwissen, Recherchekompetenzen und Wissen über die Funktionsweise der maschinellen Übersetzung. Im Rahmen der Übersetzungskompetenzen für die intralinguale Übersetzung ist die Kenntnis der Einfachen bzw. Leichten Sprache nötig.

Relevant ist auch die Kompetenz, zwischen zeitlichem Aufwand und Ästhetik abzuwägen und nach den vorgegebenen Qualitätsanforderungen nur die nötigen Änderungen vorzunehmen. Ausschlaggebend für die Qualitätsanforderungen ist, wie und wofür der fertige Text verwendet werden soll.

Darüber hinaus muss die Person, die KI nutzt, sich datenschutzrechtlichen (und im Gesundheitsbereich auch patientenrechtlichen) Aspekten bewusst sein. Die Person muss wissen, wie die KI mit sensiblen und personenbezogenen Daten umgeht, wie die KI mit den Textdaten verfährt und im Ergebnis, ob ein Text in die KI hochgeladen werden darf. Texte mit sensiblen oder personenbezogenen Daten eignen sich ggf. nicht für die KI-gestützte Übersetzung.

Beispielhaft wurde die Vorgehensweise fürs intralinguale Post-Editing im Rahmen der App VIER400 diskutiert: Hier arbeiten Journalist*innen in einem Buddy-System mit Übersetzer*innen für Einfache Sprache zusammen. Die Journalist*innen bekommen einen Workshop, in sie die DIN-Norm Einfache Sprache und das Post-Editing kennenlernen. Sie bearbeiten an den KI-generierten Texten nur die nötigen Fehler.

Kontakt: ahrenss@uni-hildesheim.de | [Projektwebseite](#)

Literatur

Deilen, Silvana et al. (2024): „Evaluation of intralingual machine translation for health communication.“ In: Song, et al. (Hrsg.): *Proceedings of the 25th Annual Conference of the European Association for Machine Translation: Volume 1 (EAMT)*, 467–477.

DIN-Normenausschuss Terminologie (NAT) (2024): *DIN ISO 24495-1:2024-03 (D): Einfache Sprache - Teil 1: Grundsätze und Leitlinien*. Beuth.

GER – Gemeinsamer Europäischer Referenzrahmen für Sprachen (o. J.): <https://www.europaeischer-referenzrahmen.de/sprachniveau.php>

Hurrelmann, Klaus, Klinger, Julia, Schaeffer, Doris (2020). *Gesundheitskompetenz der Bevölkerung in Deutschland. Vergleich der Erhebungen 2014 und 2020*. Bielefeld: Universität Bielefeld.

Schaeffer, Doris, Hurrelmann, Klaus, Bauer, Ulrich, Kolpatzik, Kai (2018): *Nationaler Aktionsplan Gesundheitskompetenz: Die Gesundheitskompetenz in Deutschland stärken*. Berlin: KomPart.