

Wiki-Modellprojekt

Kurzinformation zu Wikipedia-Artikeln: Test des KI-Tools Wiki-KurzInfo (wiki-kurzinfo.org)

Ergebnisse der Pilotphase (Januar bis März 2025)

[BineMan \(Sabine Manning\)](#) 23. April 2025



KI und Wikipedia

Überblick zum Bericht

1. Einführung

- 1.1 Wie kann das KI-Tool Wiki-KurzInfo den Lesern helfen?
- 1.2 Welche Aufgabe hat das Modellprojekt?
- 1.3 Wer nahm an der Pilotphase teil?
- 1.4 Welche Meinung hatten die Teilnehmer zum Modellprojekt?
- 1.5 Wie kamen die Teilnehmer mit dem Testen zurecht?

2. Bewertung der inhaltlichen Relevanz von KurzInfos

- 2.1 Wie werden die erzeugten KurzInfos inhaltlich und stilistisch bewertet?
- 2.2 Welche Beispiele zeigen unkorrekte Aussagen und Mängel?
- 2.3 Wie schätzen Teilnehmer die KurzInfos allgemein ein?

3. Analyse zur Verständlichkeit von WP-Artikeln und KurzInfos

- 3.1 Wie lesbar sind WP-Artikel auf verschiedenen Fachgebieten?
- 3.2 Wie lesbar sind die KurzInfos verglichen mit den WP-Artikeln?
- 3.3 Welches Sprachniveau erreichen die KurzInfos gegenüber den WP-Artikeln?
- 3.4 Wie lang sind die KurzInfos im Vergleich zu den WP-Artikeln?

4. Fazit

5. Nachwort

Abkürzungen:

WP-Artikel = Wikipedia-Artikel

KurzInfo = Kurzinformation erzeugt vom KI-Tool Wiki-KurzInfo

1. Einführung

Viele Artikel in der Wikipedia sind komplex und schwer zugänglich. Das hat [Salino01](#) auf dem [96. Digitalen Themenabend \(DTA\) der Wikipedia](#) und auf der Wikimedia-Konferenz [WikiCon 2024](#) [WikiProjekt KI und Wikipedia](#) dargelegt. Die internationale Studie zur Lesbarkeit von WP-Artikeln ([Gordejeva et al. 2022](#)) hat ergeben, dass insbesondere Artikel zu Krankheiten schwer verständlich sind. Die Studie empfiehlt, dass die Herausgeber der Artikel eine kurze und verständliche Zusammenfassung anbieten.

Im [Beitrag zum 104. DTA der Wikipedia](#) hat [BineMan](#) angeregt, eine solche Zusammenfassung mithilfe von KI zu erstellen und in Tests zu erproben. Hier setzt das Modellprojekt an.

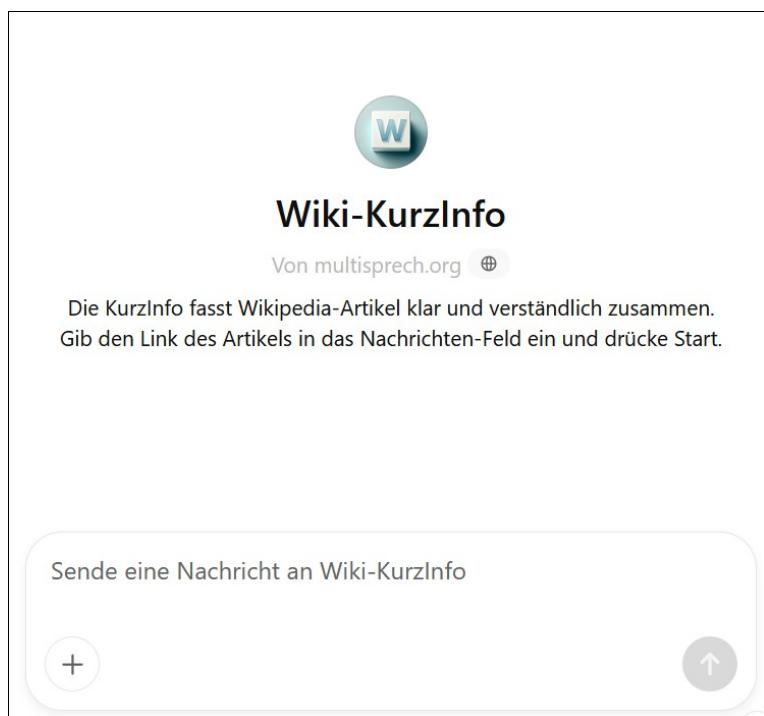
Der thematische Rahmen für das Modellprojekt ist das umfassende , konzipiert und moderiert von [Wortulo](#). Auf der Diskussionsseite dieses WikiProjekts stellte [BineMan](#) am 17.1.2025 das Modellprojekt "Verständliche Zusammenfassung von WP-Artikeln mithilfe von KI" vor ([Vorschlag für Modellprojekt](#)).

1.1 Wie kann das KI-Tool Wiki-KurzInfo den Lesern helfen?

Für das Modellprojekt hat ein Team von [multisprech.org](#) das KI-Tool **Wiki-KurzInfo** ([wiki-kurzinfo.org](#)) entwickelt. Die erzeugte Information ist für alle gedacht, die mühelos und schnell erfahren wollen, worum es in einem WP-Artikel geht.

Das Tool läuft auf ChatGPT ([Original-Link](#)). Hier ist die Startseite:

Abb. 1
Das Tool Wiki-KurzInfo



Eine ausführliche Beschreibung des Tools ist auf multisprech.org veröffentlicht ([Aufruf zum Testen](#)).

Das folgende Beispiel einer KurzInfo zeigt, welche Leistung von dem KI-Tool zu erwarten ist:

Abb. 2
Beispiel einer KurzInfo

<https://de.wikipedia.org/wiki/Osteoporose>

Osteoporose

Was ist Osteoporose?

Osteoporose, auch Knochenschwund genannt, ist eine Krankheit, bei der die Knochen dünner und poröser werden. Dadurch brechen sie leichter, besonders an der Wirbelsäule, am Oberschenkelhals und am Handgelenk.

Wer ist betroffen?

Vor allem ältere Menschen bekommen Osteoporose. In Deutschland haben etwa 6 % der Bevölkerung diese Krankheit. Besonders häufig sind Frauen nach den Wechseljahren betroffen.

Wie entsteht Osteoporose?

Unsere Knochen bauen sich ständig um. Bei Osteoporose wird mehr Knochen abgebaut als aufgebaut. Das führt zu weniger Knochendichte und schwächeren Knochen.

Welche Folgen hat das?

Die Knochen verlieren an Stabilität und können leichter brechen. Oft passiert das ohne große Verletzung. Typische Anzeichen sind Rückenschmerzen, eine gekrümmte Haltung und ein kleinerer Körper.

Wie kann man Osteoporose erkennen?

Mit einer Knochendichtemessung kann man feststellen, wie fest die Knochen sind. So lässt sich Osteoporose früh erkennen

Quelle: <https://de.wikipedia.org/wiki/Osteoporose>

1.2 Welche Aufgabe hat das Modellprojekt?

Das Tool Wiki-KurzInfo ist ein **Test-Angebot**, um Erfahrungen für ein internes KI-Tool der Wikipedia zu sammeln. Die erzeugte KurzInfo könnte als **Pop-up** zum jeweiligen Artikel erscheinen.

Die Teilnehmer des Modellprojekts sollen WP-Artikel mithilfe des KI-Tools zusammenfassen und die erzeugten KurzInfos testen sowie bewerten.

Die Tests sollen helfen, folgende Fragen zu beantworten:

- Gibt die KurzInfo die Hauptinhalte des Artikels zuverlässig wieder?
- Ist die KurzInfo ansprechend und verständlich?
- Kann man die KurzInfo als Original anbieten?

In der **Pilotphase** sollten die Teilnehmer WP-Artikel möglichst auf ihrem eigenen Fachgebiet auswählen und entlang der Fragen testen.

1.3 Wer nahm an der Pilotphase teil?

In der **Pilotphase** beteiligten sich zunächst 6 Wikipedianer spontan auf der [Diskussionsseite des WikiProjekts](#). Darüber hinaus wurden 42 Wikipedianer vom [WikiProjekt](#) und vom [104. DTA der Wikipedia](#) persönlich angefragt. So konnte der Austausch parallel per E-Mail erweitert werden. Insgesamt kamen Auskünfte von 14 Teilnehmern zusammen. Darunter führten 9 Teilnehmer eigene Tests aus und gaben Feedback zu KurzInfos von 16 WP-Artikeln.

1.4 Welche Meinung hatten die Teilnehmer zum Modellprojekt?

Die Haltung zum Modellprojekt war zumeist aufgeschlossen. Einige Teilnehmer äußerten jedoch grundsätzliche Bedenken

- gegen jeglichen Einsatz von KI bei WP-Artikeln: Es bestehe die Sorge, dass KI ohne tiefere Themenkenntnis Texte produziert, die oberflächlich oder sogar sachlich falsch sind. Besonders in fachlich sensiblen Bereichen sei Genauigkeit oberstes Gebot – eine Aufgabe, die nur Menschen mit Expertise leisten können.
- gegen Versuche, WP-Artikel mithilfe von KI verständlicher zu machen: Die enzyklopädische Qualität beruhe auf präziser, oft auch komplexer Sprache. Deren Vereinfachung könne zu inhaltlichen Verzerrungen oder Ungenauigkeiten führen und damit die Vermittlung von Wissen unterlaufen.

Diese Bedenken sind für die Tests und den Einsatz des KI-Tools zu beachten: Verständlichkeit ist wichtig – aber nicht um den Preis von Präzision. Die erzeugten KurzInfos müssen daher sorgfältig kontrolliert und von Verantwortlichen freigegeben werden.

1.5 Wie kamen die Teilnehmer mit dem Testen zurecht?

Zur Bedienung des KI-Tools gab es keinerlei Einwände oder Rückfragen. In den ersten Tests haben jedoch einige Teilnehmer den *Titel* des WP-Artikels verwendet, statt den *Link* einzugeben.

Infolgedessen entstanden KurzInfos, die sich auf mehrere WP-Artikel oder andere Quellen bezogen und zusätzliche Informationen aus diesen integrierten. Diese Fehlleistung fiel den Teilnehmern sofort auf, doch der Grund dafür war ihnen nicht klar. Wir haben daher die Teilnehmer wiederholt aufgefordert, den Link statt des Titels einzugeben.

Mehrere Teilnehmer kommentierten ihre ersten Versuche spontan, andere entwickelten eine strukturierte Bewertung. Das Feedback wurde sofort diskutiert und ausgewertet. So war es möglich, die internen Anweisungen für das KI-Tool laufend zu präzisieren.

Damit sollte während der Pilotphase erreicht werden, dass die KurzInfos:

- leichter verständlich sind,
- den Inhalt des WP-Artikels korrekt und ohne Wertung zusammenfassen,
- die hauptsächlichen Aspekte des WP-Artikels hervorheben,
- nur den betreffenden WP-Artikel als Quelle nutzen und ausweisen,
- das Interesse für den WP-Artikel wecken und zum Lesen des Artikels anregen.

Darüber hinaus entdeckten die Teilnehmer weitere Anwendungen des KI-Tools:

- Einige stellten Zusatzfragen zur KurzInfo, um besondere Inhalte des WP-Artikels zu erfahren. Die Antworten der KI waren zwar nützlich. Doch im Modellprojekt sollte die KurzInfo keinen externen Dialog auslösen, sondern zum Lesen des WP-Artikels anregen. Das wäre auch die Aufgabe eines internen KI-Tools der Wikipedia.
- Andere nutzten die Möglichkeit, alternative KurzInfos zum gleichen WP-Artikel zu erzeugen. Damit konnten sie die beste Variante auswählen. Dieses Vorgehen ist sinnvoll, wenn die KurzInfo in bearbeiteter Form verwendet werden soll. Für das Modellprojekt sollte jedoch eine einmalige Version erzeugt werden, um zu testen, ob die KurzInfo als Original eingesetzt werden könnte.

Nachfolgend stellen wir die hauptsächlichen Testergebnisse aus der Pilotphase vor. Insbesondere geht es um zwei Aspekte: die inhaltliche Relevanz (Abschnitt 1) und die Verständlichkeit (Abschnitt 2) der KurzInfos.

2. Bewertung der inhaltlichen Relevanz von KurzInfos

Aufschlüsse über die Leistung des KI-Tools haben wir aus einer Analyse der Testergebnisse gewonnen. Sie beziehen sich auf die KurzInfos zu 16 WP-Artikeln, die von den Teilnehmern erzeugt und kommentiert wurden.

2.1 Wie bewerten die Teilnehmer die erzeugten KurzInfos?

Hauptsächlich wollten wir wissen, wie zuverlässig die Aussagen in den KurzInfos sind und wie passend die Formulierung ist. Wichtig war uns auch, wie übersichtlich die KurzInfos gestaltet sind.

Die Testergebnisse haben wir unter drei Aspekten gebündelt und bewertet:

- **Inhalt:** Wie treffend sind die Aussagen: Ist der Inhalt korrekt und die Auswahl relevant?
- **Stil:** Wie passend ist die Formulierung: Ist der Stil angemessen und ansprechend?
- **Form:** Wie übersichtlich ist die Gestaltung: Hat der Text Absätze, Zwischenüberschriften und eventuell Anstriche für Aufzählungen?

Abb. 3
Bewertung*) von 13 getesteten KurzInfos
 basierend auf Einschätzungen der Testteilnehmer

KurzInfos	Aussagen	Formulierung	Gestaltung
3	Gut	Gut	Gut
3	Gut	Mittel	Gut
3	Mittel	Mittel	Gut
2	Mittel	(<i>)</i>	(<i>)</i>
1	Schwach	Mittel	Mittel
1	Schwach	(<i>)</i>	Mittel

*) Kriterien der Bewertung:

Gut = ohne Fehler bzw. ohne Mängel

Mittel = ein Fehler bzw. geringe Mängel

Schwach = zwei und mehr Fehler bzw. deutliche Mängel

(*)* = keine Bewertung

Der Überblick ergibt, dass nur 3 KurzInfos einwandfrei (ohne Fehler oder Mängel) waren. Weitere 3 KurzInfos waren inhaltlich ebenfalls in Ordnung, hatten jedoch stilistische Mängel. Die übrigen 7 KurzInfos enthielten sowohl inhaltliche Fehler als auch stilistische Mängel.

In den Worten von Testteilnehmern reichte die Qualität von einer "gut formulierten und korrekten Zusammenfassung" bis hin zu einer Zusammenfassung mit "vollkommen inhaltlich falscher" Aussage.

2.2 Welche Beispiele zeigen unkorrekte Aussagen und Mängel?

Hier sind einige Beispiele für unkorrekte Aussagen und Mängel in der Formulierung (entnommen aus den Tests):

Abb. 4
Beispiele für unkorrekte Aussagen

Zitate aus KurzInfos (Mängel kursiv)	Anmerkungen
"Behörden [...] bestehen <i>zum Beispiel</i> auf supranationaler, nationaler und subnationaler Ebene."	Das sind keine Beispiele, sondern politikwissenschaftliche Fachbegriffe, die ein Mehrebenensystem beschreiben.
"Ihre Werke zeigen <i>das harte Leben</i> dieser	Im WP-Artikel steht ausdrücklich, dass die

Arbeiterinnen und <i>machten auf ihre Situation aufmerksam.</i> "	genannten Künstler nicht die sozialen Missstände anklagten.
"Besonders störten sie sich an Gemälden von Vincent van Gogh, Honoré Daumier und Paul Signac. Einige Bürger nannten diese Werke 'perverse Kunst' ..."	Protest richtete sich nicht explizit gegen die aufgeführten Künstler, sondern die Verwendung der Gelder für moderne statt traditionelle Kunst der Region.
"Sie [...] befürchteten, dass die <i>Museumskuppel mit "Linoleumschnitten" ausgemalt</i> werden könnte."	Ausmalung der Kuppel mit Linolschnitten ist blander Unsinn.
"bunte Glasfenster, die <i>biblische Geschichten</i> darstellen"	Diese Aussage ist inhaltlich vollkommen falsch. Da kombiniert die KI die im WP-Artikel erwähnten Wörter "Bibel" und "Geschichte". Faktisch hat keines der Glasfenster eine biblische Geschichte als Thema.

Abb. 5
Beispiele für Mängel in der Auswahl von Inhalten

Manche Infos fallen unter den Tisch, wie z.B. bei einer Politikerin das 2. Staatsexamen.
Das Kapitel "Bedeutung und Vermächtnis" (im Text über eine Persönlichkeit) ist ziemlich sinnentleert.
Viele wichtige Details und Listen gingen verloren. Warum bestimmte Personen als bedeutende Künstlerinnen genannt werden, ist unklar. Eine dieser Personen wird im WP-Artikel überhaupt nicht erwähnt.
Dass eine Orgel Gottesdienste musikalisch begleitet, ist der Sinn einer Orgel in einer Kirche und somit unsinnig. Interessanter wäre z.B. der Orgelbauer gewesen.

Abb. 6
Beispiele für Mängel in der Formulierung

Das Tool verwendet gelegentlich eigene Formulierungen anstelle von eindeutigen Aussagen, z.B. "war die letzte Auseinandersetzung" (Original) wird zu "gilt als die letzte Auseinandersetzung". Das ist eine unnötige Relativierung.
Bewertungen wie „eindrucksvolles Porträt“ bringen eine subjektive Perspektive hinein, die sich vom Grundsatz des neutralen Standpunktes entfernt und in enzyklopädischen Texten fehl am Platz ist.
"Dies führte zu einer raschen Verbreitung und einem neuen Wirtschaftszweig."
In mehreren KurzInfos ist von „maßgeblich geprägt“ die Rede. Das sollte in einer KurzInfo eigentlich nicht vorkommen.

Diese Beispiele machen deutlich, welch detaillierte Sachkenntnis nötig ist, um inhaltliche Fehler oder Schwächen in der KurzInfo aufzudecken. Zugleich bedarf es sprachlicher Kompetenz, um die Mängel in den Formulierungen zu erkennen. Damit ist klar, dass die KurzInfos ein kundiges Post-Editing erfordern.

2.3 Wie schätzen Teilnehmer die KurzInfos allgemein ein?

- Die Zuverlässigkeit des KI-Tools wird kritisch, aber differenziert bewertet. Positiv wird angemerkt, dass bei einigen Tests inhaltlich korrekte und gut verständliche Zusammenfassungen erstellt wurden. Auch die Struktur der Zusammenfassungen wird teilweise gelobt, da das Tool dazu tendiert, komplexe Informationen kompakter und zugänglicher darzustellen. In diesen Fällen ähnelt der Text einem knappen Lexikoneintrag.
- Die Qualität der Zusammenfassung hängt stark vom Umfang und der Komplexität des Themas ab. Während bei schmalen, klar abgegrenzten Themen relativ gute Ergebnisse erzielt werden, bleiben die Zusammenfassungen bei breiten oder vielschichtigen Themen oft oberflächlich und lückenhaft.
- Bei der Zusammenfassung berücksichtigt die KI nicht immer die Besonderheiten eines Landes. So wählt sie für die Wirtschaft Deutschlands die drei Sektoren Industrie, Dienstleistungen und Landwirtschaft aus. Diese Aufteilung ist jedoch nicht typisch für Deutschland, sondern für Länder allgemein üblich.
- Ein weiterer Aspekt betrifft die potenzielle Zielgruppe: Während für fachkundige Leser die KurzInfos oft zu oberflächlich bleiben, könnten sie für Laien oder Schüler möglicherweise hilfreich sein – vorausgesetzt, sie werden als externes, klar gekennzeichnetes Angebot eingebunden.
- Insgesamt wird das Tool als gelegentlich hilfreich, aber nicht vollständig zuverlässig eingeschätzt. Entscheidend sei, dass der neutrale Standpunkt gewahrt bleibt und keine automatisierte Veränderung an bestehenden Artikeln vorgenommen wird.

3. Analyse zur Verständlichkeit von WP-Artikeln und KurzInfos

Wie die Diskussion zum Modellprojekt erkennen lässt, sind sich die Teilnehmer nicht einig, ob WP-Artikel verständlicher sein sollten. Wir wollen diese Frage daher voranstellen, bevor wir die Verständlichkeit der KurzInfos bewerten.

Zunächst betrachten wir die **Lesbarkeit**, die mithilfe des [Flesch-Lesbarkeits-Index](#) gemessen wird. Danach vergleichen wir das **Sprachniveau** der betreffenden Texte. Abschließend prüfen wir, ob die KurzInfos ihre vorgegebene **Textlänge** gegenüber den unterschiedlich langen WP-Artikeln eingehalten haben.

3.1 Wie lesbar sind WP-Artikel auf verschiedenen Fachgebieten?

Am ehesten wird das Problem unverständlicher WP-Artikel bei medizinischen Themen anerkannt. Die erwähnte **Studie zur Lesbarkeit von Artikeln über Krankheiten** ([Gordejeva et al. 2022](#)) hat für die deutsche Wikipedia Folgendes erbracht:

- Nahezu alle Artikel (99,7 %) waren schwer oder sehr schwer lesbar (unter dem Flesch-Wert 50).
- Die Artikel zu kodifizierten Krankheiten (77 % aller Artikel) waren im Durchschnitt sehr schwer lesbar (Flesch-Wert 20).

Wie lesbar sind demgegenüber WP-Artikel zu anderen Themen? Hierzu haben wir für das Modellprojekt eine **Stichprobe zu 6 Fachgebieten** gemacht. Für die Stichprobe haben wir ChatGPT aufgefordert,

- in der Wikipedia Artikel auf den Fachgebieten Kunst, Literatur, Medizin, Rechtswesen, Technik und Wirtschaft zu finden und
- für jedes Fachgebiet 2 Artikel zu allgemeinen Themen und 3 Artikel zu speziellen Themen auszuwählen.

Das sind insgesamt **30 WP-Artikel**. Von diesen haben wir jeweils die Einleitungen auf Lesbarkeit getestet, denn sie sind der gewöhnliche Einstieg in die Lektüre.

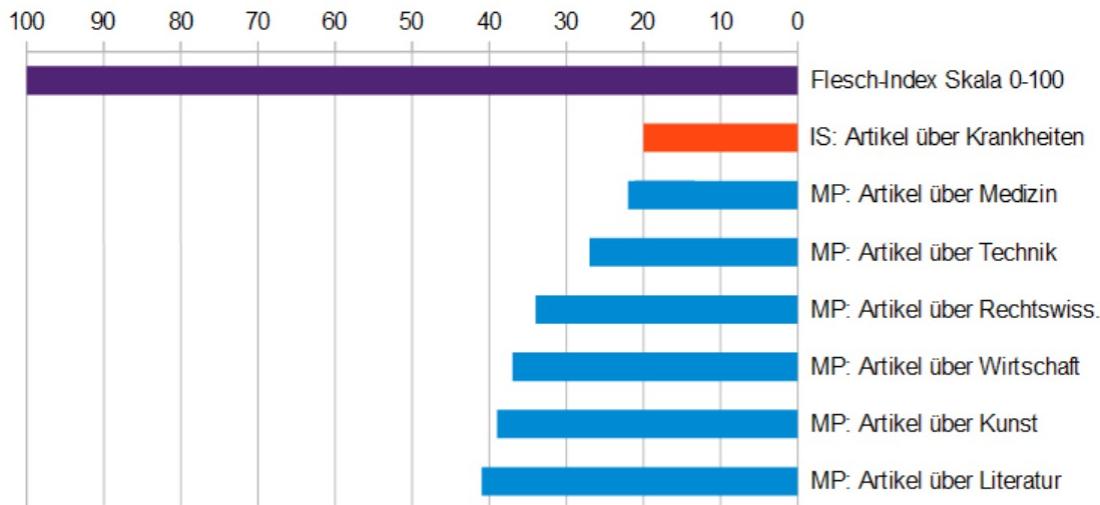
Wie bei der internationalen Studie nutzen wir den Flesch-Index ([Berechnung](#) / [Erläuterung](#)). Hier sind die Hauptstufen der Lesbarkeit dargestellt:

Abb. 7
Flesch-Lesbarkeits-Index
Skala mit Stufen der Lesbarkeit

Leicht	Mittel	Schwer	Sehr schwer
100 – 80	80 – 50	50 – 30	30 – 0

Die folgende Übersicht zeigt die Index-Ergebnisse der Stichprobe im Vergleich zu dem durchschnittlichen Indexwert aus der internationalen Studie:

Abb. 8
Lesbarkeit von Artikeln der deutschen Wikipedia
 Mittelwert gemessen mit Flesch-Index
 Internationale Studie 2022 (IS)* und Modellprojekt 2025 (MP)**



*) 4625 analysierte Artikel über Krankheiten der ICD-10 Klassifikation

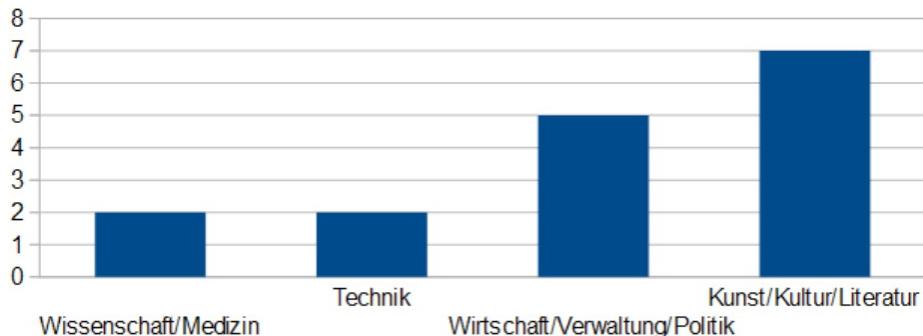
**) 30 analysierte Artikel-Einleitungen auf 6 Fachgebieten (je 5 Artikel)

Wie die Stichprobe zeigt, sind die Einleitungen zu den 30 WP-Artikeln sehr schwer bis schwer lesbar (Indexwert im Durchschnitt 33). Bei fast allen Einleitungen liegt der Indexwert unter 50 (nur 2 erreichen die Werte 52 bzw. 54). Unter den Fachgebieten sind die Artikel in der Medizin am schwersten. Der durchschnittliche Indexwert (22) liegt nur geringfügig über dem Indexwert aus der internationalen Studie (20).

Die Stichprobe kann nur ein kleiner Anhaltspunkt sein. Gleichwohl bestätigt sie, dass WP-Artikel nicht nur in der Medizin, sondern auch auf anderen Fachgebieten schwer zu lesen sind. Insofern ist es sinnvoll, das Modellprojekt auf WP-Artikel insgesamt auszurichten.

Zugleich sollten wir die fachliche Struktur der WP-Artikel berücksichtigen. Hier ist die Aufteilung der Artikel, die die Teilnehmer in der Pilotphase frei ausgewählt haben:

Abb. 9
Fachliche Aufteilung der ausgewählten WP-Artikel
 Anzahl der Artikel pro Fachgebiet

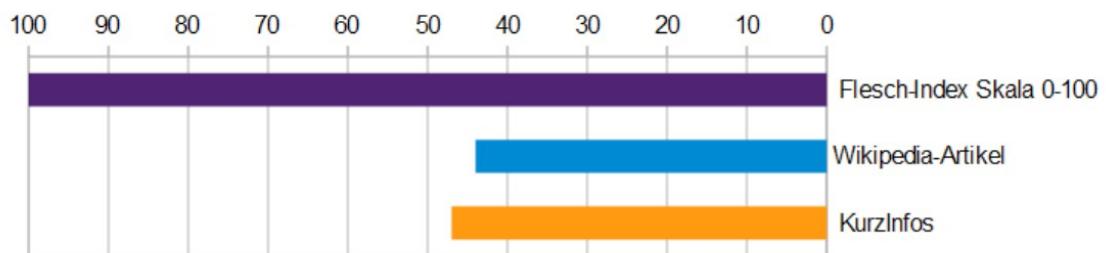


In der Auswahl sind vier große Fachgebiete vertreten. Die Texte sind also im Inhalt breit gefächert. Zugleich können wir annehmen, dass sie sich auch in der Lesbarkeit unterscheiden.

3.2 Wie lesbar sind die KurzInfos verglichen mit den WP-Artikeln?

In der folgenden Grafik vergleichen wir die durchschnittliche Lesbarkeit von KurzInfos und dazugehörigen WP-Artikeln.

Abb. 10
Lesbarkeit von WP-Artikeln und KurzInfos
 Mittelwert gemessen mit Flesch-Index
 (Modellprojekt 2025*)



*) 16 Artikel auf verschiedenen Fachgebieten

Die WP-Artikel sind im Durchschnitt schwer lesbar. Die KurzInfos erreichen zwar insgesamt einen höheren Indexwert, sind jedoch im Durchschnitt auch schwer lesbar. Nur 5 von 16 KurzInfos haben eine mittlere Lesbarkeit (bis zu Flesch-Wert 67).

Dieses Ergebnis erfüllt nicht den Anspruch, dass die KurzInfos gut lesbar sein sollten. In dieser Hinsicht muss das KI-Tool nachgebessert werden.

Allerdings misst der Lesbarkeits-Index nur die Anzahl und Länge von Wörtern, Silben und Sätzen. Er beachtet weder die Satzstruktur und Grammatik noch die Gebräuchlichkeit von Wörtern. Diese Faktoren sind aber nicht minder wichtig, um einen Text lesen und verstehen zu können. Um diese Faktoren einzubeziehen, messen wir zusätzlich das Sprachniveau der Texte.

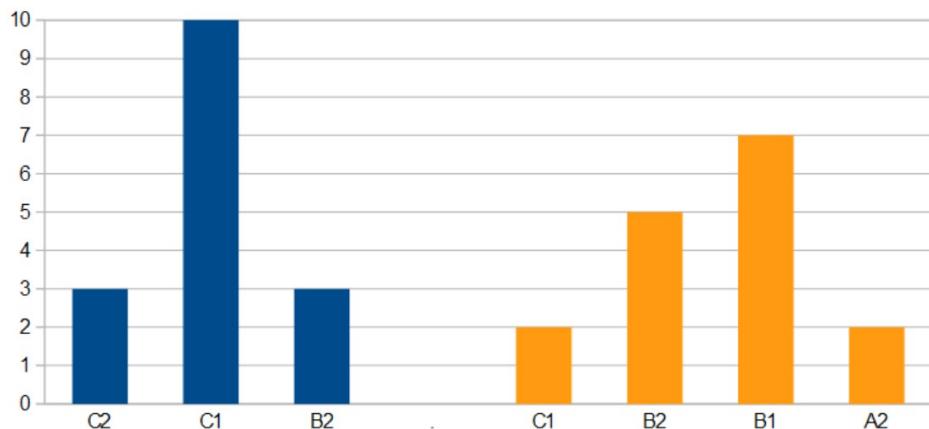
3.3 Welches Sprachniveau erreichen die KurzInfos gegenüber den WP-Artikeln?

Wir nutzen dafür das Sprachanalyse-Programm [Wortliga](#). Wortliga geht von den Sprachstufen des Gemeinsamen Europäischen Referenzrahmens ([GER](#)) aus. Im Menü von Wortliga sind die Stufen wie folgt gekennzeichnet:

Abb. 11

GER-Niveaustufen auf Wortliga
A1 Einfache Sätze
A2 Kurze Alltagstexte
B1 Sprache aus Alltag und Beruf
B2 Sachtexte und Standpunkte
C1 Komplexe und literarische Texte
C2 Anspruchsvolle und abstrakte Texte

Abb. 12
Sprachniveau von WP-Artikeln (blau) und KurzInfos (orange)
GER-Niveaustufen auf Wortliga
Modellprojekt 2025*)

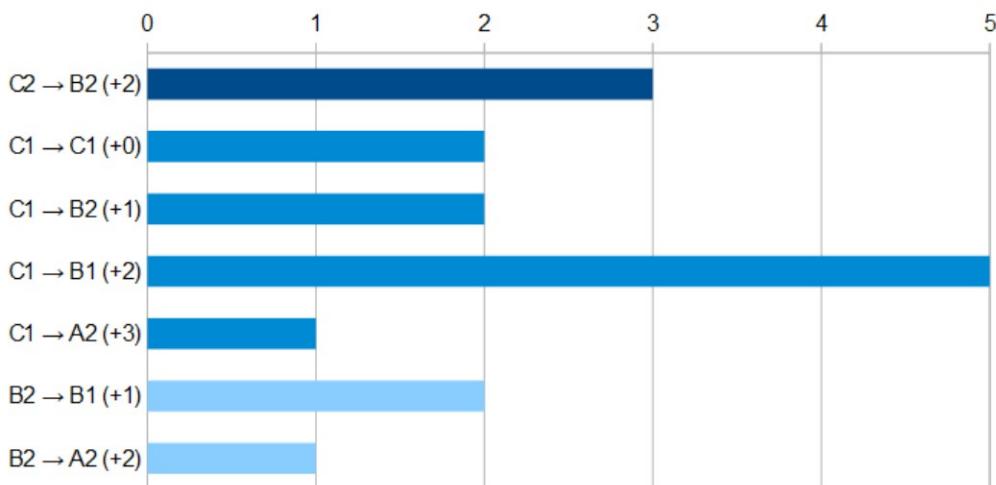


*) 16 WP-Artikel und die dazugehörigen KurzInfos

Der Vergleich der Niveaustufen ergibt, dass die WP-Artikel vorwiegend komplex formuliert sind, während die KurzInfos zumeist ein allgemeinverständliches Niveau erreichen. Damit sind die KurzInfos eindeutig verständlicher als die entsprechenden WP-Artikel.

Beide Gruppen sind jedoch in sich differenziert. Daher haben wir noch untersucht, ob es einen Zusammenhang zwischen dem Ausgangsniveau der WP-Artikel und dem erreichten Niveau der KurzInfos gibt. Steigt sich das sprachliche Niveau vom WP-Artikel zur KurzInfo?

Abb. 13
Vergleich des Sprachniveaus von WP-Artikeln und KurzInfos
 GER-Niveaustufen auf Wortliga
 Steigerung des Niveaus von WP-Artikel zu KurzInfo (Anzahl der Stufen in Klammern)
 Modellprojekt 2025*)



*) 16 Artikel und KurzInfos auf verschiedenen Fachgebieten

Hier ist das Ergebnis: Die KurzInfos haben das Sprachniveau zumeist um 2 Stufen (9-mal) oder um 1 Stufe (4-mal) verbessert. Selten hingegen blieb das Niveau gleich (2-mal) oder stieg um 3 Stufen (1-mal). Insgesamt lässt sich ableiten, dass die Wiki-KurzInfo das Sprachniveau der WP-Artikel um 1 bis 2 Stufen verbessert.

3.4 Wie lang sind die KurzInfos im Vergleich zu den WP-Artikeln?

Ein kurzer Text ist leichter lesbar als ein langer. Das ist auch ein Kriterium für die KurzInfo: Sie soll etwa 150 Wörter umfassen. Wichtig ist dabei, dass dieser Umfang unabhängig von der Länge der betreffenden WP-Artikel eingehalten wird.

Wir haben die Anzahl der Wörter bei den Artikeln und KurzInfos verglichen. In der folgenden Tabelle sind nur die niedrigsten und höchsten Werte gegenübergestellt.

Abb. 14
Länge der WP-Artikel und KurzInfos
 Anzahl der Wörter
 Modellprojekt 2025*)

Texte	Anzahl der Wörter		
WP-Artikel 1 bis 16		191 <<<<<<>>>>>	8.842
KurzInfo 1 bis 16	112 <<<>>>	226	
WP-Artikel 1 und 16		191	8842
KurzInfo 1 und 16		192	142

*) 16 Artikel und KurzInfos

Wie die Daten zeigen, ist die Spannbreite zwischen niedrigster und höchster Anzahl der Wörter bei den WP-Artikeln gewaltig. Hingegen ist die Spannbreite bei den KurzInfos relativ gering. Zudem liegt die durchschnittliche Länge der KurzInfos (166) dicht bei dem Zielwert 150.

Darüber hinaus ist ersichtlich, dass die Länge der KurzInfos nicht von der Länge der WP-Artikel abhängt: Die KurzInfo vom längsten Artikel (über 8.000 Wörter) liegt sogar unter dem Zielwert von 150 Wörtern!

Damit bestätigt der Test, dass die KurzInfos die vorgegebene Textlänge weitgehend zuverlässig einhalten.

4. Fazit

Unsere Analysen ergeben, dass WP-Artikel schwer lesbar sind. Das betrifft besonders den Bereich Medizin, aber auch andere Fachgebiete. Es ist daher sinnvoll, das Modellprojekt auf WP-Artikel insgesamt auszurichten.

Aufschlüsse über die Leistung des KI-Tools haben wir aus einer Analyse der Testergebnisse gewonnen. Hauptsächlich wollten wir wissen, wie zuverlässig die Aussagen in den KurzInfos sind und wie passend die Formulierung ist. Zusätzlich haben wir die Verständlichkeit der KurzInfos mit statistischen Mitteln analysiert.

Hier sind die wichtigsten Ergebnisse:

- Von den Teilnehmern wird das Tool als gelegentlich hilfreich, aber nicht vollständig zuverlässig eingeschätzt.

- Der Test ergibt, dass die inhaltliche Relevanz der KurzInfos unterschiedlich ist. 3 KurzInfos sind einwandfrei (ohne Fehler oder Mängel), und weitere 3 haben nur stilistische Mängel. Die übrigen KurzInfos enthalten inhaltliche Fehler sowie stilistische Mängel.
- Die KurzInfos sind besser lesbar als die WP-Artikel. Das zeigen die Werte des Flesch-Index. Doch die Unterschiede sind gering: Nur ein Drittel der KurzInfos schafft den Übergang von "schwerer" zu "mittlerer" Lesbarkeit. In dieser Hinsicht muss das KI-Tool nachgebessert werden.
- Die KurzInfos erreichen ein allgemeinverständliches Sprachniveau (B1 bis B2). Damit verbessern sie das Sprachniveau der WP-Artikel (meist C1) um 1 bis 2 Stufen.
- Die KurzInfos halten die vorgegebene Textlänge (150 Wörter) mit geringen Schwankungen ein.

Aus diesen Ergebnissen entnehmen wir, dass die KurzInfos einen nützlichen Einstieg in WP-Artikel bieten können. Sie erfordern jedoch ein sorgfältiges **Post-Editing**: Sie müssen inhaltlich und sprachlich kontrolliert und von den Verantwortlichen freigegeben werden. Wichtig ist zudem, dass die KurzInfos von den betreffenden WP-Artikeln klar abgegrenzt werden, am besten als **Pop-up**.

5. Nachwort

Die Ergebnisse der Pilotphase legen nahe, die Wiki-KurzInfo in einem breiter angelegten Test zu erproben. Dazu wären Überlegungen und Initiativen im Rahmen des WikiProjekts KI und Wikipedia wünschenswert. Wir hoffen, dass unser Bericht dazu beiträgt.

Vielen Dank für die Anregungen und Hinweise von [Stephan Manning](#), für die Unterstützung und Beratung von [Salino01](#) und [Wortulo](#) und für die Mitwirkung von [Artessa](#), [Aschmidt](#), [Daniel Mietchen](#), [Gerhardbeck](#), [Gestumblindi](#), [Goldzahn](#), [Jensbest](#), [Michael](#) und [Weltenspringerin](#).