

VERSTÄNDLICHE RECHTSTEXTE MIT HILFE GROSSER SPRACHMODELLE?

Bettina Mielke¹ / Christian Wolff²

¹Präsidentin des Landgerichts Ingolstadt, Lehrbeauftragte an der Universität Regensburg, Landgericht Ingolstadt, Auf der Schanz 37, 85049 Ingolstadt, DE, bettina.mielke@lg-in.bayern.de

²Professor, Lehrstuhl für Medieninformatik, Fakultät für Informatik und Data Science, Universität Regensburg, 93040 Regensburg, DE, christian.wolff@ur.de, <http://go.ur.de/mi>

Schlagworte: *Künstliche Intelligenz, Große Sprachmodelle, Large Language Models, ChatGPT, GPT, Verständlichkeit, Textsorte, Juristische Fachtexte*

Abstract: *In diesem Beitrag wird die Fähigkeit großer Sprachmodelle, juristische Fachtexte in einfache Sprache zu übersetzen, untersucht. Dazu erfolgt die Analyse verschiedener Textsorten – Gesetzestext, Gerichtsurteil, Fachaufsatz – hinsichtlich der vorgenommenen Änderungen nach der Bearbeitung durch Sprachmodelle und ihrer jeweiligen Verständlichkeit. Dabei kommen einfache quantitative Metriken sowie qualitative Methoden zum Einsatz. Es zeigt sich Potenzial für die Erstellung von Texten in einfacher Sprache, auch als Beitrag zur Barrierefreiheit im juristischen Bereich, wobei das Vereinfachungspotential je nach Textsorte unterschiedlich ist. Zudem ist intensive redaktionelle Nacharbeit nötig, um die Qualität der Texte zu sichern.*

1. Einleitung

Die Veröffentlichung von ChatGPT Ende 2022 hat die intensiven Diskussionen um die Nutzung von Techniken und Methoden der Künstlichen Intelligenz (KI) in der Rechtswissenschaft noch einmal befeuert. Auch wenn das Sprachmodell GPT 3 mit 175 Milliarden Parametern bereits seit dem Frühjahr 2020 bekannt war und ebenso lange Beispiele für seine Anwendung im Bereich des Rechtswesens bzw. Legal Tech diskutiert wurden, hat ChatGPT dieser Auseinandersetzung einen enormen Schub verliehen (Regalia 2023; Pettinato Oltz 2023; Schwarz/Choi 2023). Aktuelle Einschätzungen, wie sie im Rahmen einer Anhörung im Landtag des Landes Nordrhein-Westfalen zu den Perspektiven seiner Nutzung erfolgten, zeigen ein differenziertes Bild künftiger Szenarien (Mielke/Wolff 2023).

Ein Aspekt, der bereits frühzeitig diskutiert und exemplarisch vorgeführt wurde, ist die Möglichkeit, mithilfe großer Sprachmodelle juristische Fachtexte in einfache(re) Sprache zu übersetzen. Bereits im Juli 2020 postete Twitter-Nutzer @michaeltefula ein Beispiel, wie er englischsprachige Rechtstexte nach Training mit nur zwei Beispielen von GPT-3 in einfache Sprache übersetzen ließ (Abbildung 1).

AI Translations (AI Output in Blue)

Legal clause: Upon liquidation of the Company, the Series A Shareholders will receive in preference to all other shareholders an amount in respect of each Series A Share equal to one times the Original Issue Price (the "Liquidation Preference"), plus all accrued but unpaid dividends. To the extent that the Company has assets remaining after the distribution of that amount, the Series A Shareholders will participate with the holders of Ordinary Shares pro rata to the number of shares held on an as converted basis.

Plain English: If the startup is wound up, the Series A investors will be paid back at least what they invested and they will also share any leftover assets with ordinary shareholders.

Abbildung 1: Frühes Beispiel der Nutzung von GPT-3 zur Vereinfachung von Rechtstexten¹

¹ Quelle: Michael Tefula, Twitter Post, <https://twitter.com/michaeltefula/status/1285505897108832257>, 21. 7. 2020, Zugriff 12 / 23.

Hier setzt der vorliegende Beitrag an: Wir untersuchen, welche Möglichkeiten große Sprachmodelle für derartige Übersetzungsprozesse bieten. Damit wird eine ganze Reihe von Fragen aufgeworfen:

1. Welche Texte bzw. juristische Textsorten² lassen sich auf diese Weise verarbeiten?
2. Welche Anforderungen sind an geeignete Prompts zu stellen, um eine gute Ergebnisqualität zu erlangen (vgl. White et al. 2023; Ekin 2023; Lo 2023)?
3. Mit welchen Kriterien und Methoden lassen sich die Ergebnisse evaluieren? Dabei ist zu klären, ob bisherige Ansätze zur Evaluierung großer Sprachmodelle auch für diese Problemstellung geeignet sind (Liang et al. 2022)?
4. Welche der mittlerweile bekannten und viel diskutierten Schwächen der Systeme (Halluzinieren, mangelnde Wahrheitstreue, unterschiedliche Antworten auf gleiche Fragen) treten auch für diesen spezifischen Anwendungsfall auf?

Der Beitrag widmet sich dem Themenfeld Sprache und Recht und setzt eine Reihe von Studien fort, die sich im Rahmen des IRIS mit der Frage der Verständlichkeit juristischer Fachtexte auseinandergesetzt haben (Berteloot/Mielke/Wolff 2018; Mielke/Wolff 2016). Wir kombinieren qualitative Methoden des Vergleichs von Texten mit einfachen quantitativen Metriken zur Bewertung der Textkomplexität und Lesbarkeit. Die intellektuelle Bewertung der Ausgabequalität ist dabei nur von juristischen Fachexperten zu leisten.

Neben dem Generieren von (einfachen) Schreiben, der Erschließung von Akten in vielfältigen Ausgestaltungen von der Visualisierung, der Klassifizierung von Dokumenttypen bis hin zum Erstellen von Textvergleichen (Mielke/Wolff 2023) wird als möglicher Anwendungsfall großer Sprachmodelle die Übertragung von Texten in einfache Sprache genannt (Schlicht/Heetkamp 2023, S. 8; Biallaß 2023, S. 17), die einen Beitrag zur Barrierefreiheit nicht nur bei der Justiz leisten kann.

Zunächst werden wir exemplarisch für einschlägige juristische Textsorten eine derartige Übersetzung vornehmen, wobei zwei Varianten der großen Sprachmodelle zum Einsatz kommen (ChatGPT 3.5 im Vergleich mit ChatGPT 4).

2. Textauswahl

Sinnvoll erscheint, unterschiedliche juristische Textsorten zu wählen. Wir haben uns dafür entschieden, Texte aus dem Themenfeld Legal Tech und konkret aus dem Umfeld der juristischen Bewertung der Geschäftsaktivitäten der Fa. Smartlaw zu wählen. Smartlaw (vgl. <https://www.smartlaw.de/>) ist eine vom Informationskonzern Wolters Kluwer seit geraumer Zeit betriebene Plattform, auf der sich für unterschiedliche Anwendungsfälle juristische Fachtexte generieren lassen, z. B. Mietverträge, Testamente oder Arbeitsverträge. Dabei werden anhand von Fragen und vom Nutzer auszuwählenden Antworten standardisierte Vertragsklauseln abgerufen. Die Nutzung erfolgt im Browser, was auch in der mobilen Variante auf dem Smartphone möglich ist. Die Hanseatische Rechtsanwaltskammer sah in dieser Dienstleistung einen Verstoß gegen das deutsche Rechtsdienstleistungsgesetz (RDG). Der Rechtsstreit wurde 2021 vom Bundesgerichtshof zugunsten von Smartlaw entschieden und war eine wichtige höchstrichterliche Grundsatzentscheidung zu den Möglichkeiten von Legal Tech-Firmen in Deutschland. Das Dienstleistungsangebot von Smartlaw ist mit Blick auf technische Realisierung und Gebrauchstauglichkeit (*user experience*) sowie die rechtliche Bedeutung regelmäßig auch Unterrichtsgegenstand im LL.M. Legal Tech an der Universität Regensburg. Nur am Rande sei erwähnt, dass in der aktuellen Diskussion, ob im Fall der Lösung von Rechtsproblemen mittels ChatGPT (bzw. anderen großen Sprachmodellen mit Dialogfunktion) Rechtsdienstleistungen erbracht würden, häufig

² Wir verwenden den Begriff Textsorte als einschlägigen *terminus technicus* der Textlinguistik: „Der Terminus *Textsorte* wird als Sammelbegriff verwendet für eine finite Menge von Textexemplaren mit spezifischen Gemeinsamkeiten. Textsorten werden durch Zuordnungsoperationen einer Menge invarianter stereotyper Parameter auf niedriger Abstraktionsebene zu konkreten Texten konstituiert“ (HEINEMANN 2000, S. 519).

auf die Smartlaw-Entscheidung Bezug genommen wird bzw. versucht wird, die Funktionalität von Smartlaw von den Möglichkeiten, die durch ChatGPT et al. gegeben sind, abzugrenzen, vgl. Römermann/Lolou 2023, S. 67 oder Remmert 2023, S. 82f.

Rund um diese Rechtsstreitigkeit haben wir folgende Textbeispiele zur „Vereinfachung“ durch ChatGPT ausgewählt:

1. Normtext: Das Rechtsdienstleistungsgesetz (RDG, online vgl. <https://www.gesetze-im-internet.de/rdg/>).
2. Gerichtsurteil: Das Urteil des Bundesgerichtshofs im Smartlaw-Fall (BGH I ZR 113/20, ECLI:DE:BGH:2021:090921UIZR113.20.0, online vgl. https://t1p.de/BGH_I_ZR_113_20).
3. Ein Fachaufsatz zu diesem Urteil: Uwer 2022, der sich in seiner Urteilsbesprechung kritisch mit der Diskussion zur Smartlaw-Entscheidung auseinandersetzt.

3. Prompt Engineering

Mit den neuen Möglichkeiten von ChatBots bzw. Dialogsystemen auf der Basis großer Sprachmodelle ist schnell deutlich geworden, dass ein entscheidender Faktor für die Ausgabe- und Ergebnisqualität die Art und Weise ist, wie Fragen an das System gestellt werden. Das *Prompt Engineering* hat sich rasch zu einem neuen Qualifikationsfeld zwischen Informationskompetenz und fachwissenschaftlicher Anwendung entwickelt. Sowohl aus informationswissenschaftlicher (Lo 2023) wie auch aus fachwissenschaftlicher Perspektive (Pettinato Oltz 2023; Regalia 2023; Schwarcz/Choi 2023; Noonan 2023) liegen erste Modelle und Handreichungen vor, die Hinweise auf gute Prompts entwickeln.

Lo 2023 schlägt beispielsweise das CLEAR-Modell vor, das folgende Aspekte und Schritte beinhaltet:

Concise: Kürze und Klarheit der Prompts

Logical: Strukturierte und kohärente Prompts

Explicit: Klare Beschreibung der Spezifikationen für die Ausgabe

Adaptive: Flexibilität und Anpassbarkeit der Prompts

Reflective: Kontinuierliche Evaluierung und Verbesserung der Prompts

Auf dieser Basis haben wir folgende Formulierung als Prompt entwickelt und jeweils an die Textsorte angepasst:

Norm: „Nachfolgend übergebe ich die ersten fünf Paragraphen des Rechtsdienstleistungsgesetzes (RDG). Diese sollen bitte in leicht verständliche Sprache umgewandelt werden: Einfache Sätze, einfache Begriffe, am besten in leicht verständlicher Sprache. Hier ist der Text: [...]“

Aufsatz: „Nachfolgend übergebe ich einen juristischen Fachaufsatz. Dieser soll bitte in leicht verständliche Sprache umgewandelt werden: Einfache Sätze, einfache Begriffe, am besten in leicht verständlicher Sprache. Hier ist der Text: [...]“

Urteil: „Nachfolgend übergebe ich einen Teil eines Urteils des Bundesgerichtshofs. Dieses soll bitte in leicht verständliche Sprache umgewandelt werden: Einfache Sätze, einfache Begriffe, am besten in leicht verständlicher Sprache. Hier ist der Text: [...]“

Die „Übersetzung“ wurde zunächst mit der frei verfügbaren Fassung ChatGPT 3.5 über die webbasierte Schnittstelle durchgeführt. Da dort einschließlich des Prompts nur etwa vier Kilobyte Text übergeben werden können, haben wir uns auf ausgewählte Passagen des Aufsatzes und des Urteils sowie beim Normtext auf die ersten fünf Paragraphen des RDG beschränkt. In einem zweiten Lauf wurden die Textauschnitte mit demselben Prompt an ChatGPT 4, die gebührenpflichtige Variante, übergeben. Insgesamt sind so sechs tabellarische

Gegenüberstellungen entstanden, die hier aus Platzgründen nicht wiedergegeben werden können, die aber Online verfügbar sind.³

Im Anschluss wurde die ChatGPT-Ausgabe kopiert und dem Prompt-Text in einer Tabelle gegenübergestellt, die Zuordnung der einzelnen Paragraphen erfolgt intellektuell nach erster Durchsicht und Vergleich von Input und Output. Wir sind uns dabei der Tatsache bewusst, dass wiederholte Anfragen mit identischem Prompt unterschiedliche Ausgaben erzeugen können. Eine erste – qualitative und stichprobenartige – Annäherung an die Frage, ob aktuelle große Sprachmodelle für die Übersetzung von Fachtexten geeignet sind und welche Phänomene dabei zu beobachten sind, dürfte dies nicht beeinträchtigen, auch wenn klar sein sollte, dass mit mehr Texten, mehr Variation in den Prompts und deren Wiederholung weitere interessante Effekte auftreten dürften.

4. Evaluierung großer Sprachmodelle

Zur Evaluierung großer Sprachmodelle liegt mittlerweile umfangreiche Literatur vor. Neben Überblicksaufsätzen, die sich nicht nur auf die GPT-Familie von OpenAI beschränken, gibt es spezifische Studien, die

- ein bestimmtes Anwendungsfeld in den Blick nehmen (z. B. Einsatz für die Codeproduktion (Frankston 2023) oder Einsatz in der Medizin (Elmarakeby et al. 2023)) und
- Studien, die bestimmte Merkmale (etwa Sprache oder Faktentreue) untersuchen (Koo et al. 2023).

Spezifische Evaluationsstudien für den rechtlichen Bereich sind bisher noch kaum verfügbar, eine entsprechende Suche im Web of Science erbrachte keine Ergebnisse.⁴ Die bekannteste und umfangreichste Evaluationsstudie dürfte die *Holistic Evaluation of Language Models* (HELM) an der Universität Stanford sein (Liang et al. 2022). Hier werden zahlreiche große Sprachmodelle umfassend und weitgehend automatisiert bewertet, wobei sowohl nach typischen *Funktionen* (u. a. Klassifikation, Frage-Antwort, Zusammenfassung) als auch nach konkreten *Bewertungsmerkmalen* (Akkuratheit, Robustheit, Fairness) unterschieden wird. Der Schwerpunkt liegt dabei – wie so häufig im Bereich *Natural Language Processing* – auf der englischen Sprache. Der Aspekt der Transformation von (komplexerer) Fachsprache in einfachere Sprache wird dort bisher nicht betrachtet.

Für unsere Analyse ziehen wir einfache quantitative Metriken (Anzahl Zeichen, Wörter) sowie ein Maß zur Verständlichkeit heran. Im Zentrum steht aber die intellektuelle Analyse der Textbeispiele („close reading“, Byron 2021). Hier sollen sprachliche und inhaltliche Unterschiede gleichermaßen Berücksichtigung finden.

5. Analyse

Nachfolgend berichten wir für die drei herangezogenen Textsorten Norm, Urteil und Aufsatz jeweils ausgewählte Beobachtungen und ergänzen diese durch statistische Basisdaten.

5.1. Auffälligkeiten bei der RDG-Vereinfachung (§§ 1 – 5 RDG)

Auf den ersten Blick sind die Unterschiede nicht sehr groß, allerdings ergibt sich – ungefragt – ein nicht unerheblicher Kürzungseffekt. ChatGPT 4 verdichtet den Ausgangstext fast auf die Hälfte des ursprünglichen Umfangs. Während der Ausgangstext des RDG einen Umfang von 409 Wörtern hat, ergibt sich durch ChatGPT 3.5 eine leichte Verkürzung auf 365 Wörter und durch ChatGPT 4 auf 211 Wörter.

³ Online abrufbar: <https://t1p.de/MielkeWolffMaterialdokumentationVerstaendlichkeitIRIS2024>, Zugriff 12 / 23.

⁴ Suche nach: TS=(„large language models“ OR GPT OR ChatGPT) AND evaluation) AND TS=(legal OR law) im Web of Science (<https://www.webofscience.com/>) sowie nach “evaluation „large language models“ legal law” in Google Scholar.

5.1.1. Textstruktur

Bei ChatGPT 3.5 bleibt die Paragraphenstruktur erhalten, bei ChatGPT 4 ist eine Struktur mit Paragraphen und Absätzen nicht mehr erkennbar. Zwei Beispiele sollen die Unterschiede illustrieren (lexikalische Unterschiede zwischen Original und der Ausgabe ChatGPT 3.5 hervorgehoben):

Originaltext RDG	Ausgabe ChatGPT 3.5	Ausgabe ChatGPT 4
<p>§ 2 Begriff der Rechtsdienstleistung</p> <p>(1) Rechtsdienstleistung ist jede Tätigkeit in konkreten fremden Angelegenheiten, sobald sie eine rechtliche Prüfung des Einzelfalls erfordert.</p>	<p>§ 2 Begriff der Rechtsdienstleistung:</p> <p>(1) Eine Rechtsdienstleistung ist jede Tätigkeit in konkreten rechtlichen Angelegenheiten, die eine Prüfung des Einzelfalls erfordert.</p>	<p>2. Was ist eine Rechtsdienstleistung?</p> <p>– Jede Arbeit, die eine rechtliche Prüfung in einer bestimmten Angelegenheit erfordert.</p>
<p>§ 5 Abs. 1 Satz 1 RDG</p> <p>Erlaubt sind Rechtsdienstleistungen im Zusammenhang mit einer anderen Tätigkeit, wenn sie als Nebenleistung zum Berufs- oder Tätigkeitsbild gehören.</p>	<p>Erlaubt sind Rechtsdienstleistungen im Zusammenhang mit einer anderen Tätigkeit, wenn sie eine Nebenleistung zum Berufs- oder Tätigkeitsbild darstellen.</p>	<p>– Rechtsdienstleistungen sind erlaubt, wenn sie eine Nebenleistung zu einer anderen Haupttätigkeit sind.</p>

Die Unterschiede in der Definition der Rechtsdienstleistung (§ 2 Abs. 1 RDG) sind bei ChatGPT 3.5 noch überschaubar, bei ChatGPT 4 aber bereits recht deutlich. Auch bei § 5 Abs. 1 Satz 1 RDG ist die Verkürzung bei ChatGPT 4 weitaus stärker als bei ChatGPT 3.5.

5.1.2. Lexikalische Unterschiede

Gerade bei der Ausgabe von ChatGPT 3.5, die die Ausgangsstruktur des Normtextes weitgehend beibehält, fallen lexikalische Unterschiede auf, etwa beim Austausch von Fremdwörtern. So heißt es in § 1 Abs. 1 RDG „vor unqualifizierten Rechtsdienstleistungen“, bei ChatGPT 3.5 „vor schlechter Qualität bei Rechtsdienstleistungen“, bei ChatGPT 4 ist die Umformulierung wieder deutlich stärker, der Begriff „qualifiziert“ bleibt aber hier (die Stelle lautet insgesamt „nur qualifizierte Personen Rechtsberatung geben“). § 2 Abs. 3 RDG nennt „die Erstattung wissenschaftlicher Gutachten“, ChatGPT 3.5 „das Erstellen von wissenschaftlichen Gutachten“, ChatGPT 4.0 verkürzt auf „wissenschaftliche Gutachten“ (Erstellen/Erstatten fehlt). In § 2 Abs. 3 RDG ist von „Erörterung“ die Rede, in ChatGPT 3.5 von „Bespprechung“, in ChatGPT 4 von „Diskussionen“. In § 3 RDG heißt es „die selbständige Erbringung“, bei ChatGPT 3.5 „die eigenständige Erbringung“, bei ChatGPT 4 fehlt „selbständig/eigenständig“. Ähnliches kann man bei § 4 Satz 1 RDG beobachten, der Normtext spricht von „Rechtsdienstleistungen, die unmittelbaren Einfluss auf die Erfüllung“, ChatGPT 3.5 von „Rechtsdienstleistungen, die direkten Einfluss auf die Erfüllung“ haben, bei ChatGPT 4 fehlt „unmittelbar/direkt“. In § 2 Abs. 3 RDG ist von „Schiedsrichterinnen und Schiedsrichtern“ die Rede, ChatGPT 3.5 verkürzt dies auf „Schiedsrichter“, ChatGPT 4 lässt den Begriff weg.

5.1.3. Syntaktische Änderungen

Die syntaktischen Änderungen sind heterogen. So gibt es die Umformulierung eines Relativsatzes in eine (komplexere) Nominalphrase. Bei § 5 Abs. 1 Satz 2 RDG heißt es „unter Berücksichtigung der Rechtskenntnisse zu beurteilen, die für die Haupttätigkeit erforderlich sind“. ChatGPT 3.5 formuliert dies um in „unter Berücksichtigung der für die Haupttätigkeit erforderlichen Rechtskenntnisse“. Bei der Ausgabe durch

ChatGPT 4 fehlt diese Passage. Umgewandelt werden zudem (und im Sinne einer Vereinfachung eher nachvollziehbar) Infinitiv- in Relativsätze, wenn ChatGPT 3.5 den Satz „Regelungen in anderen Gesetzen über die Befugnis, Rechtsdienstleistungen zu erbringen, bleiben unberührt“ (§ 1 Abs. 3 RDG) in „Andere Gesetze, die die Befugnis zur Erbringung von Rechtsdienstleistungen regeln, bleiben hiervon unberührt“ umformuliert. Bei ChatGPT 4 heißt es: „Andere Gesetze, die Rechtsdienstleistungen regeln, sind davon nicht betroffen.“ § 1 Abs. 1 Satz 1 RDG lautet: „Dieses Gesetz regelt die Befugnis, in der Bundesrepublik Deutschland außergerichtliche Rechtsdienstleistungen zu erbringen.“ Bei ChatGPT 3.5 heißt es: „Das Gesetz regelt, wer in Deutschland außerhalb von Gerichtsverfahren Rechtsdienstleistungen erbringen darf.“ Bei ChatGPT 4 lautet der entsprechende Satz: „Das Gesetz regelt, wer in Deutschland außergerichtliche Rechtshilfe anbieten darf.“ Teilweise wird auch der immer wieder als typisches Merkmal der Rechtssprache genannte Nominalstil abgeschwächt. Statt „wenn ihr Gegenstand deutsches Recht ist“ (§ 1 Abs. 2 RDG), findet sich bei ChatGPT 3.5 die Formulierung „wenn es um deutsches Recht geht“ und bei ChatGPT 4 „wenn sie deutsches Recht betreffen“.

5.1.4. Fazit

Festzuhalten ist, dass der Regelungsgehalt bei der Umformulierung durch ChatGPT 3.5 insgesamt erkennbar bleibt, bei nicht unerheblichen Abweichungen im Detail. Der Umfang reduziert sich bei ChatGPT 3.5 nur unwesentlich, auch ist die Normstruktur noch gut sichtbar, während der Umfang bei ChatGPT 4 deutlich kürzer ist und manches fehlt.

5.2. Auffälligkeiten bei der Urteilsvereinfachung (BGH-Urteil zu Smartlaw)

Bei dieser Textsorte fällt auf, dass der Text stark verkürzt wird und zwar sowohl bei ChatGPT 3.5 als auch bei ChatGPT 4. Zum Teil fehlen mehrere Sätze völlig, formale Angaben wie Urteilszitationen und Literaturangaben entfallen durchgängig. Der Ausgangstext hat einen Umfang von 488 Wörtern, die Ausgabe von ChatGPT 3.5 enthält 178 Wörter, die von ChatGPT 4 beinhaltet 161 Wörter. Aufgrund der so großen Verkürzung ist eine lexikalische und syntaktische Feinanalyse kaum möglich. Festzustellen ist auf jeden Fall, dass es sich um deutlich kürzere Sätze handelt. Auch lexikalisch fällt auf, dass bestimmte Begriffe nicht mehr vorkommen. Dies ist im Hinblick auf den Begriff *konkret* nicht unproblematisch, da es bei diesem BGH-Urteil wesentlich darauf ankommt, dass das System (Smartlaw) nicht in einer *konkreten* Angelegenheit im Sinne des § 2 Abs. 1 RDG tätig wird. Bei ChatGPT 3.5 kommt der Begriff noch in der Wendung „auf einen konkreten Fall ausgerichtet“ vor, bei ChatGPT 4 fehlt *konkret* gänzlich. Damit ist fraglich, ob der Kern dieser Urteilspassage noch getroffen wird.

5.3. Auffälligkeiten bei der Vereinfachung des Aufsatzes zur Diskussion zum Smartlaw-Urteil

Auch hier zeigt sich der Verdichtungseffekt in einer vergleichbaren Größenordnung wie bei dem Urteilstext: Von 512 Wörtern im Ausgangstext auf 282 Wörter bei ChatGPT 3.5 und 220 Wörter bei ChatGPT 4. In der Ausgabe von ChatGPT 3.5 und 4 finden sich deutlich kürzere Sätze. Zudem fehlen der in dem Aufsatz zitierte Leitsatz der Entscheidung sowie weitere Zitate aus anderen Quellen.

Interessant ist der Umgang mit dem Titel des ausgewählten Abschnitts „Smartlaw: Much Ado about Nothing – or Little?“. Bei ChatGPT 3.5 wird der Titel umgesetzt in „Smartlaw: Was bedeutet das BGH-Urteil für die Rechtsberatung“, ChatGPT 4 *übersetzt* hingegen den Titel in „Smartlaw: Viel Lärm um nichts – oder doch etwas?“

Inhaltlich wirkt die Ausgabe von ChatGPT 3.5 wesentlich neutraler als der Ausgangstext. So fehlen etwa die Ausdrücke „mitunter überzeichnet wirkende Diskussion“, „seit langem schwelender Konflikt“ oder „berufspolitische Beschwörungsformeln“ des Originaltextes.

Originaltext Aufsatz	Ausgabe ChatGPT 3.5	Ausgabe ChatGPT4
Deren Adaptionsgeschwindigkeit lässt viele verbittert zurück, die sich, häufig der traditionellen anwaltlichen Berufsbildpflege verhaftet, mit einem unbehaglichen Kontrollverlust konfrontiert sehen. Der Notwendigkeit einer zukunftsorientierten regulatorischen Neuausrichtung des immer volatileren Rechtsdienstleistungsmarktes wird dann gerne mit berufspolitischen Beschwörungsformeln begegnet:	Einige, die an das traditionelle Bild der Anwaltschaft gebunden sind, fühlen sich durch die schnelle Entwicklung von Legal Tech unsicher und haben Angst vor Kontrollverlust.	Viele traditionelle Anwälte fühlen sich durch die schnelle Entwicklung und den Kontrollverlust bedroht.

Zusammenfassend kann man sagen, dass der Sinn grundsätzlich verständlich geblieben ist, allerdings wird nicht mehr deutlich, dass der Autor die Kritik an dem Urteil als übertrieben ansieht. ChatGPT 4 ist hier z. T. etwas näher am Ausgangstext als ChatGPT 3.5. Bei ChatGPT 4 werden beispielsweise ein Artikel im Anwaltsblatt sowie ein Papier der Bundesrechtsanwaltskammer, von denen der Autor berichtet, genannt. Beide kommen in ChatGPT 3.5 nicht vor. Allerdings wird im Fall des Artikels im Anwaltsblatt die Stoßrichtung des Autors falsch wiedergegeben, wenn es bei ChatGPT 4 heißt: „In einem Artikel im Anwaltsblatt wurde draufhingewiesen, wie hitzig die Debatte über Legal Tech und Anwälte ist“, während im Originalaufsatz der Autor die Diskussion im Anwaltsblatt als hitzig bezeichnet.

5.4. Bewertung der Verständlichkeit der Texte

Die insgesamt neun Textausschnitte wurden mit Hilfe des Online-Tools <http://www.leichtlesbar.ch/> bewertet und es wurde jeweils die Verständlichkeitsmetrik des Flesch-Index berechnet:

	RDG §§ 1-5	BGH-Urteil	Aufsatz
Originaltext	Unsere Analyse hat ergeben: Ihr Text besteht aus 17 Sätzen mit 408 Wörtern, wovon 217 verschiedene. Sie haben total 904 Silben benutzt (Endsilben auf -e zählen nicht). Das ergibt einen Flesch-Wert von -5.	Unsere Analyse hat ergeben: Ihr Text besteht aus 39 Sätzen mit 486 Wörtern, wovon 286 verschiedene. Sie haben total 979 Silben benutzt (Endsilben auf -e zählen nicht). Das ergibt einen Flesch-Wert von 24.	Unsere Analyse hat ergeben: Ihr Text besteht aus 18 Sätzen mit 516 Wörtern, wovon 317 verschiedene. Sie haben total 1078 Silben benutzt (Endsilben auf -e zählen nicht). Das ergibt einen Flesch-Wert von 1.

	RDG §§ 1-5	BGH-Urteil	Aufsatz
Ausgabe ChatGPT 3.5	Unsere Analyse hat ergeben: Ihr Text besteht aus 21 Sätzen mit 365 Wörtern, wovon 202 verschiedene. Sie haben total 778 Silben benutzt (Endsilben auf -e zählen nicht). Das ergibt einen Flesch-Wert von 9.	Unsere Analyse hat ergeben: Ihr Text besteht aus 9 Sätzen mit 179 Wörtern, wovon 110 verschiedene. Sie haben total 338 Silben benutzt (Endsilben auf -e zählen nicht). Das ergibt einen Flesch-Wert von 27.	Unsere Analyse hat ergeben: Ihr Text besteht aus 17 Sätzen mit 293 Wörtern, wovon 180 verschiedene. Sie haben total 595 Silben benutzt (Endsilben auf -e zählen nicht). Das ergibt einen Flesch-Wert von 18.
Ausgabe ChatGPT 4	Unsere Analyse hat ergeben: Ihr Text besteht aus 18 Sätzen mit 203 Wörtern, wovon 131 verschiedene. Sie haben total 439 Silben benutzt (Endsilben auf -e zählen nicht). Das ergibt einen Flesch-Wert von 12.	Unsere Analyse hat ergeben: Ihr Text besteht aus 12 Sätzen mit 161 Wörtern, wovon 104 verschiedene. Sie haben total 321 Silben benutzt (Endsilben auf -e zählen nicht). Das ergibt einen Flesch-Wert von 25.	Unsere Analyse hat ergeben: Ihr Text besteht aus 15 Sätzen mit 229 Wörtern, wovon 151 verschiedene. Sie haben total 452 Silben benutzt (Endsilben auf -e zählen nicht). Das ergibt einen Flesch-Wert von 24.

Die Flesch-Formel, nach dem österreichisch-amerikanischen Lesbarkeitsforscher Rudolf Flesch benannt, ergibt sich aus einer gewichteten Berechnung der mittleren Wortlänge (WL) und der mittleren Satzlänge eines Textes (konkret: Flesch-Index $FI = 206,835 - 84,6 \times WL - 1,015 \times SL$).

Auf der hier benutzten Website <http://www.leichtlesbar.ch/> finden sich Übersichten und Beispiele zur Einordnung der Werte des Flesch-Index:

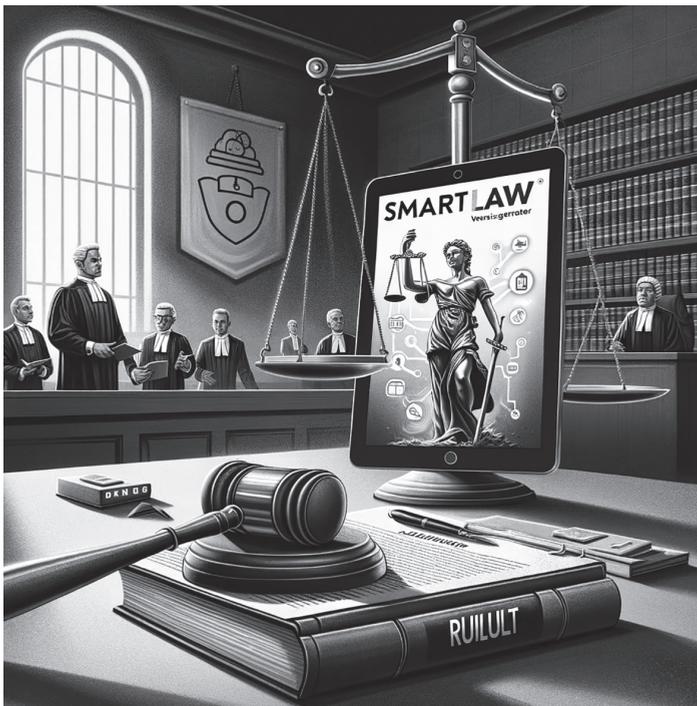
Flesch-Index			Bewertung
81	bis	100	extrem leicht (5. Klasse)
71	bis	80	sehr leicht (6. bis 8. Klasse)
61	bis	70	leicht (Abschlussklasse)
41	bis	60	durchschnittlich (Sek, FOS, Berufsschule)
31	bis	40	etwas schwierig (Mittelschule)
21	bis	30	schwierig (Matura, Abitur)
	bis	20	sehr schwierig (Hochschulabschluss)

Die Ausgangstexte sind damit unterschiedlich gut lesbar, wobei der Urteilsausschnitt mit einem Flesch-Wert von 24 nur in die zweitschwierigste Kategorie fällt, d. h. deutlich leichter lesbar ist als der Gesetzestext und der Aufsatztext. Durch die Transformation mit generativer KI verbessert sich die Lesbarkeit jeweils, allerdings mehr oder weniger deutlich. Nur bei dem Aufsatz wird auch eine neue Kategorie der Skala erreicht (vom sehr niedrigen Flesch-Wert 1 auf den Flesch-Wert 24 nach der Bearbeitung mittels ChatGPT 4). In zwei von drei Fällen wird mit ChatGPT 4 eine deutlichere Vereinfachung erreicht als mit ChatGPT 3.5.

6. Fazit und Ausblick

Die Analyse weniger ausgewählter Texte kann nur einen ersten Eindruck von den Möglichkeiten der Transformation von Fachtexten in einfachere Sprache liefern. Immerhin wird dabei das grundsätzliche Potential deutlich: Die Texte werden bei einem vergleichsweise einfachen Prompt nicht nur kürzer, sondern auch einfacher. Komplexe Elemente wie Querverweise, Zitate und Gender-Formulierungen fallen weg. Der wesentliche inhaltliche Kern bleibt zumindest ansatzweise erhalten, wobei die einfachere Version typischerweise kommunikativen Zwecken dient und nicht an die Stelle des Originaltextes treten kann. Durch spezifischere Prompts dürfte sich die Vereinfachung noch besser steuern lassen, weitere Studien, die die Prompts gezielt variieren und mehr und umfangreichere Texte untersuchen, wären sinnvoll. Im Ergebnis dürfte anzunehmen sein, dass für viele Kommunikationssituationen, in denen einfachere Sprache wünschenswert ist, mit der Texttransformation über KI ein nützliches Werkzeug verfügbar geworden ist, das die sonst sehr arbeitsaufwändige intellektuelle Erstellung von Texten in einfacher Sprache leichter möglich macht, selbst wenn im Einzelnen noch erhebliche redaktionelle Nacharbeit erforderlich sein sollte.

ChatGPT bietet neben Möglichkeiten zur Texttransformation mittlerweile auch die Ansteuerung von Bildgeneratoren wie Dall-E als Teil seines Leistungsspektrums an. Ganz im Sinne der Idee der Rechtsvisualisierung und des multisensorischen Rechts (Brunschwig 2018) kann auch die visuelle Darstellung rechtlicher Zusammenhänge bzw. wissenschaftlicher Argumentation einen Beitrag zur verständlichen Rechtssprache leisten. Ein Beispiel, bei dem die ersten Absätze des betrachteten Aufsatzes (Uwer 2022) durch ChatGPT visualisiert wurden (Prompt: „Visualisiere den folgenden Text: [Text]“), soll dies illustrieren:



Here is a visual representation of the text you provided, illustrating the landmark legal decision involving the „Smartlaw Vertragsgenerator“ and its implications in the legal field.

Abbildung 2: Über einen ChatGPT-Dialog mit Dall-E generiertes Bild zu dem hier betrachteten Aufsatz

Das Bild kommt allerdings über eine (erwartbare) Gerichtsszene (mit angelsächsischen Attributen) kaum hinaus. Die Beschriftungen sind Phantasiebegriffe, die entfernt an relevante Konzepte aus dem Text erinnern (auf dem Tablet im Bild: „Veersisgerrater“ → Vertragsgenerator?), es lässt sich aber aus dem Bild keine relevante Aussage des Textes rekonstruieren. Auch wenn man einräumen muss, dass durch gezieltere Prompts konkretere Bildinhalte generiert werden können, bleibt für den produktiven Einsatz der Bildgenerierung hier noch ein recht weiter Weg.

7. Literatur

BTELOOT, PASCALE/MIELKE, BETTINA/WOLFF, CHRISTIAN, Deutsches, österreichisches, europäisches Deutsch? Deutschsprachige Fassungen von Urteilen des europäischen Gerichtshofs im Vergleich. In: Erich Schweighofer/Franz Kummer/Ahti Saarenpää/Burghardt Schafer (Hrsg.), Datenschutz / LegalTech. Tagungsband des 21. Internationalen Rechtsinformatik Symposions (IRIS 2018). Editions Weblaw, Bern, 319-324, 2018.

BIALASS, ISABELLE, Stellungnahme des Deutschen EDV-Gerichtstags e.V. zur Anhörung „Einsatz von ChatGPT im Justizbereich“ im Landtag NRW. Düsseldorf: Landtag Nordrhein-Westfalen, Stellungnahme 18/575, online: <https://www.landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMST18-575.pdf>, 2023.

BRUNSCHWIG, COLETTE R., Perspektiven einer digitalen Rechtswissenschaft: Visualisierung, Audiovisualisierung und Multisensorisierung (Perspectives of Digital Law: Visualization, Audiovisualization, and Multisensorization), Max Planck Institute for European Legal History Research Paper Series, 2018.

BYRON, MARK, Close Reading. In: Oxford Research Encyclopedia of Literature, Oxford: OUP, DOI: <https://doi.org/10.1093/acrefore/9780190201098.013.1014>, 2021.

EKIN, SABIT, Prompt Engineering for ChatGPT: A Quick Guide to Techniques, Tips, and Best Practices, TechRxiv, May 04, 2023, DOI: 10.36227/techrxiv.22683919.v2, 2023.

ELMARAKEBY, HAITHAM A./TRUKHANOV, PAVEL S./ARROYO, VIDAL M./RIAZ, IRBAZ BIN/SCHRAG, DEBORAH/VAN ALLEN, ELIEZER M./KEHL, KENNETH L., Empirical Evaluation of Language Modeling to Ascertain Cancer Outcomes from Clinical Text Reports, BMC Bioinformatics, 24: 324, 2023.

FRANKSTON, B., Chatting up a Program, IEEE Consumer Electronics Magazine, 4-6, 2023.

HEINEMANN, WOLFGANG, „Textsorte – Textmuster – Texttyp (*Text Type – Text Pattern*)“. In: Brinker, Klaus/Antos, Gerd/Heinemann, Wolfgang/Sager, Sven F. (Hrsg.), Text- und Gesprächslinguistik. Linguistics of Text and Conversation. Ein internationales Handbuch zeitgenössischer Forschung. An International Handbook of Contemporary Research. 1. Halbband / Volume 1. DeGruyter, Berlin, 507-523, 2000.

KOO, SEONMIN/PARK, CHANJUN/LEE, SEOLHWA/SEO, JAEHYUNG/EO, SUGYEONG/MOON, HYEONSEOK/LIM, HEUISEOK, Uncovering the Risks and Drawbacks Associated with the Use of Synthetic Data for Grammatical Error Correction, IEEE Access, 2023, 11, Vol. 11, 95747-95756, 2023.

LIANG, PERCY/BOMMASANI, RISHI/LEE, TONY/Tsipras, DIMITRIS/SOYLU, DILARA/YASUNAGA, MICHIMIRO/ZHANG, YIAN/NARAYANAN, DEEPAK/WU, YUHUI/KUMAR, ANANYA, Holistic Evaluation of Language Models, arXiv preprint arXiv:2211.09110, 2022.

LO, LEO S., The CLEAR Path: A Framework for Enhancing Information Literacy through Prompt Engineering, The Journal of Academic Librarianship, Vol. 49 No. 4, 102720, 2023, <https://doi.org/10.1016/j.acalib.2023.102720>.

MIELKE, BETTINA/WOLFF, CHRISTIAN, Österreichische und Deutsche Gerichtsentscheidungen im Sprachvergleich. In: Erich Schweighofer/Franz Kummer/Walter Hötzendorfer/Georg Borges (Hrsg.), Netzwerke. Tagungsband des 19. Internationalen Rechtsinformatik Symposium Salzburg (IRIS 2016), Österreichische Computer Gesellschaft & Erich Schweighofer, Wien 129-138, 2016.

MIELKE, BETTINA/WOLFF, CHRISTIAN, Korpuslinguistik in der Rechtswissenschaft. Eine webbasierte Analyseplattform für EuGH-Entscheidungen, Jusletter IT, 27. Mai 2021, DOI: 10.38023/9b2a8ae5-fcc9-4df4-a599-3d2a59868ba5.

MIELKE, BETTINA/WOLFF, CHRISTIAN, Künstliche Intelligenz und Large Language Models in der Rechtsprechung, LRZ – E-Zeitschrift für Wirtschaftsrecht und Digitalisierung, 560-602, 2023, online: www.lrz.legal/2023Rn560.

NOONAN, NICK, Creative Mutation: A Prescriptive Approach to the Use of ChatGPT and Large Language Models in Lawyering, Social Sciences Research Network (SSRN 4406907): <http://dx.doi.org/10.2139/ssrn.4406907>, 2023.

- PETTINATO OLTZ, TAMMY, ChatGPT, Professor of Law, Social Sciences Research Network (SSRN 4347630), online: <http://dx.doi.org/10.2139/ssrn.4347630>, 2023.
- REGALIA, JOSEPH, From Briefs to Bytes: The GAI Renaissance in Legal Writing, Social Sciences Research Network (SSRN 4371460), online: <http://dx.doi.org/10.2139/ssrn.4371460>, 2023.
- REMMERTZ, FRANK, Legal Tech – Update im anwaltlichen Berufsrecht und im RDG, LTZ, 75-84, 2023.
- RÖMERMANN, VOLKER/LOLOU, IRIS-SYNTHIA, ChatGPT und Berufsrecht. Ist Rechtsrat durch ChatGPT eigentlich erlaubt?, *Rethinking:Law*, 2023, S. 64–67.
- SCHLICHT, CHRISTIAN/HEETKAMP, SIMON, Stellungnahme zur Anhörung „Einsatz von ChatGPT im Justizbereich“ im Landtag NRW. Düsseldorf: Landtag Nordrhein-Westfalen, Stellungnahme 18/574, online: <https://www.landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMST18-574.pdf>, 2023.
- SCHWARCZ, DANIEL/CHOI, JONATHAN H., AI Tools for Lawyers: A Practical Guide, Social Sciences Research Network (SSRN 4404017), online: <http://dx.doi.org/10.2139/ssrn.4404017>, 2023.
- UWER, DIRK, Das Vertragsgenerator-Urteil des BGH: A Good Case to Make „Smartlaw“? – Die anwaltsberufsrechtliche Perspektive, *LRZ - E-Zeitschrift für Wirtschaftsrecht und Digitalisierung*, 53-80, 2022, online: www.lrz.legal/2022Rn53.
- WHITE, JULES/FU, QUCHEN/HAYS, SAM/SANDBORN, MICHAEL/OLEA, CARLOS/GILBERT, HENRY/ELNASHAR, ASHRAF/SPENCER-SMITH, JESSE/SCHMIDT, DOUGLAS C., A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT, arXiv preprint arXiv:2302.11382, 2023.

