Technical University of Munich | Arcisstraße 21 | 80333 Munich
Professorship for Legal Tech | Boltzmannstraße 3 | 85748 Garching

An den
Präsidenten des Landtags Nordrhein-Westfalen
Herrn Andre Kuper MdL
Platz des·Landtags 1
40221 Düsseldorf

An den Rechtsausschuss

LANDTAG
NORDRHEIN-WESTFALEN
18. WAHLPERIODE

STELLUNGNAHME
18/588

A14

Stellungnahme: "Einsatz von ChatGPT im Justizbereich"

Zur Anhörung von Sachverständigen durch den Rechtsausschuss am 13.6.2023

Autor:

Prof. Matthias Grabmair, Ph.D., LL.M.
Assistant Professor für Legal Tech
School of Computation, Information and Technology
Technische Universität München

Fassung vom 5.6.2023

Der Schwerpunkt meiner Expertise liegt im Bereich der Anwendung technischer Methoden aus dem Bereich von künstlicher Intelligenz, Wissensrepräsentationen, maschinellem Lernen und Natural Language Processing auf juristische Daten und Problemstellungen. Ich kann daher auf Fragen 1, 4, 5, 6, 7, 8 und 16 keine sachverständigen Antworten geben. Die übrigen Fragen wurden thematisch gruppiert.

Vorbemerkungen

Die Antworten in dieser Stellungnahme beziehen sich teils spezifisch auf ChatGPT, teils auf das Folgemodell GPT-4 sowie generative *Large Language Models* (LLMs) im Allgemeinen. Diese werden als "generativ" bezeichnet, da sie darauf ausgelegt sind, die Verteilung von Trainingsdaten zu lernen und neue Daten dieser Verteilung zu generieren. Dies ist von sog. "diskriminativen" Modellen zu unterscheiden, mit denen eine Dateneingabe bestimmten Klassen oder numerischen Werten zugeordnet werden (zB die Klassifikation von Dokumenten in ein bestimmtes Typenschema). Typischerweise werden generative Modelle auf großen Mengen Texten anhand der Aufgabe trainiert, das auf einen bestimmten Eingabetext folgende Wort zu bestimmen. Sie erzeugen daher entsprechend ihrem Training in Reaktion auf eine Eingabe (der sog. "Prompt") anhand bestimmter Parameter einen Text. Dieser kann unter Umständen von einem durch einen Menschen geschriebenen Text nicht zu unterscheiden sein, jedoch auch faktisch unrichtige Behauptungen oder unsinnige Inhalte enthalten. ChatGPT ergänzt diese Formel durch das sog. Reinforcement-Learning-from-Human-Feedback (RLHF), in dem das Modell eine weitere Trainingsstufe durchläuft, in der die Eignung der möglichen Generierungen

Technical University of Munich
TUM School of Computation,
Information and Technology
Professorship for Legal Tech

Matthias Grabmair
Ph.D., LL.M.
Boltzmannstraße 3
85748 Garching

Tel. +49 89 289 17 146

matthias.grabmair@tum.de
www.cit.tum.de
www.tum.de

Bayerische Landesbank
IBAN-Nr.:
DE10700500000000024866
BIC: BYLADEMM
Steuer-Nr.: 143/241/80037
USt-IdNr.: DE811193231

als Antwort auf die Frage/Aussage/Aufforderung eines Benutzers mit in die Berechnung einfließt.[1] Es ist als Durchbruchstufe generativer Sprachmodelle zu sehen, die in Zukunft weiterentwickelt werden.

Antworten auf Fragen:

2. Besteht die Gefahr, dass Urteile von Richtern und Beschlüsse von Rechtspflegern in Zukunft vollständig durch ChatGPT gefertigt werden und nähern wir uns damit der Gefahr eines „Robo-Jugdes“?

15. Wie beurteilen Sie die Nutzung von ChatGPT durch Richterinnen und Richter zum Verfassen von Urteilen?

Der prominente KI-Forscher und Unternehmer Richard Socher charakterisiert geeignete Nutzungsszenarien für generative KI als solche, in denen "es eine lange Zeit bräuchte um ein Artefakt zu erschaffen, jedoch sehr wenig Zeit um seine Korrektheit zu verifizieren".[2] Diese Aussage umreißt aus meiner Sicht sehr zutreffend die zentrale Verschiebung in der textzentrierten Arbeit der Rechtspraxis durch generative KI-Modelle.

Betrachten wir zunächst den angesprochenen Fall der Bescheidverfassung: Der/die Richter*in/Rechtspfleger*in/etc. startet nicht mit einem leeren Dokument, sondern kann mit Hilfe einer Prompt und ggf. etwas kontextuellem Inhalt aus der Akte einen Erstaufschlag eines Beschlusses oder anderen Schriftstücks erhalten.

Mit dieser technischen Möglichkeit ändert sich die Natur der Tätigkeit von eigener Textproduktion hin zu einem idealerweise mehrstufigen Vorgehen. Der vom Modell erstellte Text muss auf Richtigkeit und sachliche Vollständigkeit überprüft werden. Der/die Richter*in muss fehlende oder unrichtige Inhalte als solche erkennen und korrigieren bzw. ergänzen. Im Umkehrschluss kann der Text auch Aspekte beinhalten, die bei einer manuellen Erstellung des Dokuments übersehen worden wären. So wird ein finales Dokument erstellt, dessen Inhalt sich der/die Richter*in bei der Verwendung im Prozess zu eigen macht und für das er/sie die volle Verantwortung trägt. In dieser Idealbeschreibung ergeben sich insbesondere folgende Risikodimensionen relativ zum von der Frage aufgeworfenen provokanten "Robo-Judge" Szenario:

- **Welche Rolle spielt die rechtliche Bewertung bei der Textgenerierung?** Die anfängliche Prompt kann bereits die wesentlichen rechtlichen Entscheidungen vorwegnehmen (zB "Generiere einen Beschluss zur Ablehnung der Leistungsklage im voranstehenden Sachverhalt wegen der erfolgreichen Einrede der Verjährung mit entsprechender Kostenverteilung") oder offen lassen (zB "Generiere einen Urteilsentwurf im voranstehenden Sachverhalt für die Entscheidung über die Leistungsklage und behandle dabei die mögliche Verjährungseinrede"). Im ersten Fall geht es um die

---

[1] Diese Technik kam ua auch schon bei dem Vorgängermodell *InstructGPT* zum Einsatz. Siehe: Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, et al. “Training Language Models to Follow Instructions with Human Feedback.” arXiv, March 4, 2022. https://doi.org/10.48550/arXiv.2203.02155.

[2] Tweet von Richard Socher [https://twitter.com/RichardSocher/status/1648735187088601088]; Originaltext: "GenAI is amazing when it would take a long time to create an artifact but very little time to verify its correctness"; vom Autor übersetzt.

Erstellung eines Dokuments auf der Basis einer menschlich getroffenen Entscheidung.[3] Im zweiten Fall wird diese Entscheidung an das Modell delegiert und der/die Richter*in muss sich im Anschluss kognitiv gewissermaßen "über das Modell hinwegsetzen" falls der Vorschlag als nicht sachgerecht empfunden wird.

- **Wie viel Prüfung und Überarbeitung des generierten Textes und der darin enthaltenen rechtlichen Wertungen ist regelmäßig zu erwarten?** Obgleich der erste Textaufschlag potenziell große Zeitersparnis mit sich bringt, muss der/die Richter*in nach wie vor mit intensiver mentaler Arbeit die Entscheidung in der Sache verifizieren, fehlende Aspekte ergänzen, und neu aufgeworfene prüfen. Durch Zeitdruck und wiederholte Nutzung des Modells können sich neue "Abkürzungen" im gerichtlichen Arbeitsalltag bilden. Diese können zu einem punktuellen Aufmerksamkeitsdefizit sowie einer Tendenz führen, die vorgeschlagenen Inhalte mit zusehends weniger Prüfung zu übernehmen. Dies ist freilich bei detailliert ausformulierten Prompts weniger problematisch als bei solchen, die rechtliche Wertungen offen lassen (siehe Beispiel oben).
- **Wird generative KI zur Erstellung des Prozessdokuments oder als Werkzeug zur Falldurchdringung und Recherche genutzt?** Neben der Unterstützung bei der Erstellung von Prozessdokumenten können auf generativer KI basierende Dialogsysteme (wie zB ChatGPT) auch zur effizienten inhaltlichen Erschließung von Dokumenten genutzt werden. Hierbei ergibt sich der Mehrwert abermals aus der gesparten manuellen Arbeit bei der Sichtung von großen Textmengen. Beispielsweise kann GPT-4 aus einer Sachverhaltsschilderung eine tabellarische Darstellung relevanter Ereignisse mit Datum erstellen. Diese schafft eine kompakte Übersicht und kann ggf. schnell manuell verifiziert werden, wird aber möglicherweise nicht Teil eines Beschlusses. Es ist außerdem damit zu rechnen, dass künftig Dokumentenanalysesysteme generative Sprachmodelle intern als Kernkomponenten für die inhaltliche Erschließung von Texten und Generierung von Zusammenfassungen, etc., einsetzen, wobei hier die Komplexität der Prompt-Formulierung dem Benutzer wohl weitestgehend angenommen werden wird.

Diese Risikodimensionen werden in den Antworten auf die folgenden Fragen aufgegriffen und teilweise vertieft. In jedem Fall sollte aus den Erläuterungen hervorgehen, dass generative Sprachmodelle zwar verschiedene Tätigkeiten in der Justiz effektiv unterstützen können, dies jedoch eine sachkundige und achtsame Benutzung voraussetzt.

3. In dem Artikel: WISSEN, Freitag, 17. März 2023, Artikel 1⁄4, „KI - Bald intelligenter als ein Mensch?“ heißt es: „Jedenfalls macht GPT-4 nochmals Fortschritte bei Aufgaben, die auch der Vorgänger GPT-3.5 schon konnte. So schnitt GPT-4 bei akademischen Tests teilweise deutlich besser ab. Bei einigen, etwa einem juristischen Test, lag seine Leistung im besten Zehntel der menschlichen Testteilnehmer.“ Was bedeutet dies für die Arbeit von Gerichten, Verwaltungen, Rechtspflegern, etc. in Zukunft?

9. Eignet sich ChatGPT für den Einsatz in der Justiz?

---

[3] Diese Nutzung ist teilweise analog zur Arbeitsweise mit dem *Frauke* Werkzeug zur Baukasten-artigen Erstellung Urteilen von Fluggastrechteverfahren. Siehe Werkstattvortrag von Christian Metz zu FRAUKE bei der Legal Tech Tagung 2022 der MLTech Student Association vom 15.10.2022; https://www.youtube.com/watch?v=tzRT45w9Dy8&t=2506s&ab_channel=MunichLegalTech%28MLTech%29

10. Wenn ja, in welchen Bereichen und in welchem Umfang?

14. Welche Potenziale sehen Sie im Einsatz solcher oder ähnlicher Systeme in der Justiz unter welchen Voraussetzungen?

11. Wenn nein, weshalb nicht?

12. Wo sehen Sie mögliche Gefahren und Risiken beim Einsatz solcher und ähnlicher Programme in der Justiz?

Frage 3 zielt wohl auf die Ergebnisse von GPT-4 auf dem **amerikanischen Uniform Bar Exam** ab (UBE).[4] Hier erzielte es eine Gesamtpunktzahl von 297 Punkten, was oberhalb typischer Bestanden-Schwellen verschiedener Bundesstaaten liegt. Ungeachtet der an der Berechnung der Ergebnis-Perzentilen angebrachten Zweifel[5] ist hier anzumerken, dass das UBE zu substanziellen Teilen aus Multiple-Choice-Fragen besteht, die zwar zur Wissensprüfung von Studierenden geeignet sein mögen, jedoch nicht der Arbeit in der Rechtspraxis entsprechen. Die Qualität der Texte, die GPT-4 zu den Freitext-Fragen des UBE generierte, ist teilweise durchaus bemerkenswert und legt in der Tat nahe, dass generative Sprachmodelle juristische Textarbeit künftig effektiv unterstützen können. Anbieter von Software zur Unterstützung der Anwaltschaft haben bereits begonnen, Modelle der GPT-Familie in ihre Produkte zu integrieren oder zur Grundlage neuer Produkte zu machen.[6] Es ist daher zu empfehlen, die Nutzbarkeit von generativen Sprachmodellen in der Justiz prinzipiell auszuloten. Wie zu Frage 2 bereits angesprochen, ist neben den funktionalen Limitierungen der Modelle hierbei vor allem zu unterscheiden, zu welchen Zwecken generative Sprachmodelle verwendet werden sollen und wie sie in bestehende Arbeitsabläufe achtsam integriert werden können.

Geeignete **Einsatzbereiche** sind meines Erachtens nach technischen Funktionen abzuschichten, für die generative Modelle besonders geeignet sind und im Anwendungskontext Mehrwert erzeugen. Diese beinhalten zB Dokumentgenerierung, gezielte Informationsextraktion, Dokumentzusammenfassungen, Informations-Umformatierung und andere Aufgaben, deren Ergebnis vom Benutzer effizient überprüft werden kann. Die **Potenziale** dieser Unterstützung liegen auf der Hand: Automatische Zusammenfassungen ersparen den Zeitaufwand zur manuellen Sichtung von umfangreichem Schriftgut. Eine vergleichende Informationsextraktion aus mehreren Dokumenten kann in Form einer Liste von Fragen geschehen, für die das Modell binnen Minuten eine tabellarische Übersicht erstellt, die einen Menschen Stunden beschäftigen würde. Gut ausformulierte Prompts verkürzen den Weg von der getroffenen Entscheidung in der Sache zu einer Textvorlage, die idealerweise nur noch geprüft und ggf. geschliffen werden muss. Konsequent weitergedacht birgt ein auf Interaktion spezialisiertes generatives Sprachmodell (wie zB ChatGPT) eine Zukunftsvision von Justizarbeit, in der ein Assistenzsystem zur Verfügung steht, das den Akteninhalt kennt und so jederzeit Fragen darüber beantworten kann sowie in

---

[4] Katz, Daniel Martin and Bommarito, Michael James and Gao, Shang and Arredondo, Pablo, GPT-4 Passes the Bar Exam (Version vom 15.3.2023). Abrufbar unter SSRN: https://ssrn.com/abstract=4389233 or http://dx.doi.org/10.2139/ssrn.4389233

[5] vgl. Martínez, Eric, Re-Evaluating GPT-4's Bar Exam Performance (May 8, 2023). Verfügbar unter: https://ssrn.com/abstract=4441311; Anmerkung: Nach bestem Wissen des Autors unterlief dieser Artikel (noch) keinem Peer-Review.

[6] Beispiele in USA: CaseText CoCounsel [https://casetext.com/blog/casetext-announces-cocounsel-ai-legal-assistant ; Harvey AI [https://www.harvey.ai] in Partnerschaft mit Allen & Overy; in Deutschland: JUNE [https://www.june.de]

begrenztem aber nützlichem Maße Recherche- und Informationsaufbereitungsaufgaben nahezu augenblicklich erledigen kann. Diese Benutzungsformen pauschal zu untersagen erscheint mir nicht sachgerecht. Verschiedene Funktionen von generativen Sprachmodellen bergen verschiedene Risiken und machen eine nuancierte Betrachtung notwendig.

Die Frage nach dem **Umfang** des Einsatzes ist kontextunabhängig nur schwer zu beantworten. Wenn in einem komplexen Verfahren die Funktionalität eines generativen Sprachmodells eine schnelle Navigation in der Akte ermöglicht, dann ist dies mit Sicherheit vergleichsweise unkritisch zu sehen. Sollten hingegen Richter*innen bei der Generierung von Texten substanzielle rechtliche Entscheidungen offen lassen (heißt: nicht bereits in der Prompt ausreichend spezifizieren) und/oder Dokumente in so großem Umfang automatisch generieren lassen, dass eine angemessene fachliche Prüfung nicht mehr stattfinden kann, sind aus meiner Sicht Grenzen überschritten.

Im Folgenden werden die aus meiner Sicht bedeutsamsten Gefahren/Risiken und allgemeinen Beschränkungen generativer Sprachmodelle für den Justizkontext behandelt. Zunächst zu den durch die technische Funktionalität der Modelle bedingten Aspekten:

- **"Halluzinationen"**: Der Begriff "Halluzinationen" bezeichnet im Bereich generativer Sprachmodelle die Erzeugung von Text, der faktisch inkorrekt ist, Informationen aus Quellen verfälscht oder unsinnige Aussagen enthält.[7] Diese Herausforderung ist den Entwicklern solcher Systeme bekannt. So enthält der technische Bericht von OpenAI zu GPT-4 einen Abschnitt mit quantitativen Ergebnissen einer internen Korrektheitsanalyse.[8] Im Justizkontext ist dieses Verhalten offensichtlich hochproblematisch, da Entscheidungen auf der Basis aller korrekten Sachverhaltselemente unter Anwendung der korrekten juristischen Definitionen und Regeln erfolgen muss. Es ist davon auszugehen, dass im Laufe der Weiterentwicklung dieser Modelle diese Phänomene weniger prominent auftreten (wie zB auch die Verbesserungen von GPT-4 im Vergleich zu ChatGPT) und spezielle Modellarchitekturen näher mit "korrektem" Text arbeiten (zB unter Einbeziehung einer Dokumentensuche). Dennoch sollte generierter Text vor der Verwendung stets fachkundig geprüft werden. Meines Erachtens ist die Gefahr inhaltlicher Verfälschung allerdings kein ausreichender Grund, die Nutzung dieser Modelle in der Justiz prinzipiell zu verbieten. Stattdessen sind spezifische mehrwertschaffende Funktionen (zB Zusammenfassungen, tabellarische Aufarbeitung von Dokumenteninhalt, Frage-Antwort-Navigation in der Akte) dahingehend zu prüfen, (1) ob der generierte Inhalt von dem/der Richter*in effizient auf Fehler überprüft werden kann und (2) ob punktuelle inhaltliche Verfälschungen tatsächlich in der Sache gefährlich oder als Ungenauigkeiten akzeptabel sind. Analog zu Halluzinationen können generative Sprachmodelle auch unrichtige Quellenangaben und ganze fiktive Zitate auf externe Dokumente, Urteile, etc. erstellen. Ein jüngst prominentes Beispiel ist der Fall, in dem ein Anwalt vor einem amerikanische Bundesgericht in einem Schriftsatz durch ungeprüfte Übernahme von

[7] siehe für eine Typologie dieser Phänomene: Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of Hallucination in Natural Language Generation. ACM Comput. Surv. 55, 12; https://doi.org/10.1145/3571730
[8] siehe OpenAI, GPT-4 Technical Report, https://arxiv.org/abs/2303.08774, Seite 10 f.

durch ChatGPT generiertem Text mehrere Präzedenzfälle zitierte, die nicht existierten.[9] Ironischerweise gab er an, ChatGPT nach der Echtheit der Zitate gefragt und sich auf die entsprechende Bestätigung des Systems verlassen zu haben. Ungeachtet dieses außergewöhnlichen Falls anwaltlicher Fahrlässigkeit ist dennoch technisch zum einen zu erwarten, dass künftige Modellgenerationen und spezielle Architekturen besser geeignet sind, Referenzen auf externe Dokumente verlässlich zu produzieren. Das von Deepmind entwickelte, LLM-basierte Question Answering System *GopherCite* beispielsweise kann seine Antworten auf Allgemeinwissensfragen durch Zitate aus externen Dokumenten belegen.[10] Im rechtlichen Kontext kombiniert die *CoCounsel* Software der amerikanischen Firma CaseText bereits jetzt generative Sprachmodelle mit einer Urteilsdatenbank.[11] In jedem Fall bleibt hinsichtlich der Authentizität generierter Inhalte jedoch der Benutzer in der Prüfpflicht.

- **Können generative Sprachmodelle rechtlich Argumentieren?** Ein Sprachmodell kann mittels einer entsprechenden Prompt dazu gebracht werden, den Sachverhalt eines Falles schrittweise ähnlich dem Gutachtenstil unter rechtliche Regeln zu subsumieren. Dies bedient sich dem gleichen Mechanismus wie das sog. "Chain of Thought Prompting", bei dem die Prompt durch den Hinweis ergänzt wird, dass das Modell seine Antwort Schritt für Schritt herleiten soll.[12] Dies wurde auf Fallangaben im deutschen Recht in einzelnen Versuchen erprobt und in den sozialen Medien anekdotisch besprochen.[13] Es wurde so demonstriert, dass aktuelle GOT-Modelle gutachtenstilartigen Text produzieren können, die darin enthaltenem juristische Schlussfolgerungen und Argumente allerdings noch teils recht flach und von gemischter Güte sind. Meines Erachtens können solche vereinzelten Ergebnisse noch nicht so weit verallgemeinert werden, dass generativen Sprachmodellen die Fähigkeit zur robusten juristischen Argumentation zugesprochen werden kann. Dies gilt nicht zuletzt aufgrund der stark begrenzten Kenntnisse über die Trainingsdaten und Architekturen der im Fokus stehenden OpenAI Modelle. Belastbare, systematische empirische Ergebnisse aus akademischer Forschung auf deutschen juristischen Texten existieren meines Wissens nach noch nicht[14], was auch an der hohen Geschwindigkeit der Entwicklung im privaten Sektor liegt.
- **Begrenzte Eingabelänge**: Generative Sprachmodelle können aus Rechenkapazitätsgründen im Regelfall nur eine begrenzte Anzahl von Worten als

---

[9] Weiser, Benjamin. “Here’s What Happens When Your Lawyer Uses ChatGPT.” The New York Times, May 27, 2023, sec. New York. https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html.

[10] Menick, Jacob, Maja Trebacz, Vladimir Mikulik, John Aslanides, Francis Song, Martin Chadwick, Mia Glaese, et al. “Teaching Language Models to Support Answers with Verified Quotes.” arXiv, March 21, 2022. http://arxiv.org/abs/2203.11147.

[11] https://casetext.com/blog/casetext-announces-cocounsel-ai-legal-assistant

[12] Wei, Jason, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. "Chain of thought prompting elicits reasoning in large language models." arXiv preprint arXiv:2201.11903 (2022) [https://arxiv.org/abs/2201.11903].

[13] zB https://www.linkedin.com/posts/braegel_so-l%C3%B6st-bings-chatbot-den-ersten-fall-rutschige-activity-7033021141573070848-24vc?utm_source=share&utm_medium=member_desktop

[14] Siehe aber für GPT-3 auf amerikanischen Daten: Andrew Blair-Stanek, Nils Holzenberger, Benjamin Van Durme, Can GPT-3 Perform Statutory Reasoning?, https://arxiv.org/abs/2302.06100 [wird nach Peer Review im Juni 2023 als Teil des ICAIL 2023 Konferenzbands erscheinen].

Eingabe verarbeiten.[15] Die GPT-3.5-Generation (zu der auch ChatGPT gehört) beispielsweise arbeitet mit einer maximalen Eingabelänge von 4.096 Tokens[16], wogegen GPT-4 in der Standard-Variante bis zu 8.192 Tokens verarbeitet und in einer "extended" Version sogar 32.768 Tokens erreicht. In Justizkontexten und Analyseaufgaben auf Akten würden diese Limits wohl regelmäßig überschritten. Meines Erachtens ist dies jedoch kein langfristig relevanter Grund, die Anwendbarkeit der Modelle prinzipiell zu beschränken. Die Erweiterung dieser Grenzen wird intensiv beforscht und es existieren verschiedene technische Ansätze, um konkrete Analysefunktionen auch auf langen Dokumenten effektiv zu implementieren.

- **Erklärbarkeit des generierten Textes**: Siehe Antwort zu Fragen 13 und 17 unten.
- **Bias und toxische Inhalte**: Von generativen Sprachmodellen erstellter Text kann sozial-stereotypische Muster aus den Trainingsdaten enthalten.[17] Ferner können sie auch erlernte schädliche (zB beleidigende) Inhalte in Reaktion auf bestimmte Prompts wiedergeben.[18] Die Entwickler von GPT-4 begegnen dieser Herausforderung mit Sicherheitstests durch Experten sowie der Nutzung von speziellen Datensätzen[19], doch ein vollständiger Schutz vor solchen Inhalten und gefährlichen Modellausgaben ist bislang nicht erreicht. Beim Einsatz im Justizkontext muss meines Erachtens noch erörtert werden, welche spezifischen Risiken diesbezüglich bei der Verwendung in Gerichten existieren und wieviel Sicherheitsvorkehrungen und Tests für den Einsatz genügen. Zum einen wirkt die bereits erklärte notwendige manuelle Prüfung von generierten Dokumenteninhalten schützend. Zum anderen können sich aber auch subtile Risiken ergeben, falls beispielsweise Modelle durch Akteninhalte in ausländischen Sprachen weniger genau arbeiten als auf rein deutschen Daten.
- **LLMs als "Foundation Models" und als Backend-Service**: Generative Sprachmodelle haben den Bereich des Natural Language Processing (sowie KI im Allgemeinen) derart beeinflusst, dass sie inzwischen auch als "Foundation Models" bezeichnet werden.[20] Dies bedeutet, dass sie als Kerntechnologie nach ihrem "neutralen" Vortraining die technische Grundlage für spezialisierte Anwendungskontexte bilden können. Es wäre daher zu kurz gegriffen, sich lediglich mit der Nutzung von einzelner Systeme wie ChatGPT im Justizkontext zu beschäftigen. Es ist mit Sicherheit zu erwarten, dass Dokumentenanalysesysteme in Zukunft regelmäßig intern generative Sprachmodelle auf verschiedene Weisen nutzen, wobei die erwähnten Risiken (Halluzinationen, etc.) nicht

---

[15] "Worte" im Sinne von Sprachmodellen sind typischerweise sog. "Tokens". Dabei handelt es sich um ein flexibles, in der Größe begrenztes Vokabular aus natürlichen Worten und Wortkomponenten (dh häufigen Buchstabenkombinationen), mit denen Texte für das Modell kodiert werden.

[16] siehe https://platform.openai.com/docs/guides/chat/managing-tokens

[17] siehe zB für geschlechterbezogene Stereotypen in generierten Erzählungen: Lucy, Li, and David Bamman. “Gender and Representation Bias in GPT-3 Generated Stories.” In Proceedings of the Third Workshop on Narrative Understanding, 48–55. Virtual: Association for Computational Linguistics, 2021. https://doi.org/10.18653/v1/2021.nuse-1.5.

[18] Gehman, Samuel, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A. Smith. “RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models.” In Findings of the Association for Computational Linguistics: EMNLP 2020, 3356–69. Online: Association for Computational Linguistics, 2020. https://doi.org/10.18653/v1/2020.findings-emnlp.301.

[19] siehe OpenAI, GPT-4 Technical Report, https://arxiv.org/abs/2303.08774, Seite 11 ff.

[20] Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, et al. “On the Opportunities and Risks of Foundation Models.” CoRR abs/2108.07258 (2021). https://arxiv.org/abs/2108.07258.

zwangsläufig für alle Nutzungsarten problematisch sind. Eine unqualifizierte Beschränkung des Einsatzes solcher Modelle als ganze Kategorie würde die Natur von Dokumentanalyse-Fachanwendungen verfehlen.

Ferner ergeben sich Fragestellungen hinsichtlich der Benutzung der Modelle durch Anwender im Justizkontext:

- **Abhängigkeit von Prompts:** Die Qualität und Nutzbarkeit eines generierten Textes für eine zu lösende Aufgabe sind stark von der Prompt abhängig. Die Fähigkeit zur Erstellung guter Systemeingaben (sog. "Prompt Engineering") erfordert zeitintensives systematisches Vorgehen und Erfahrung.[21] Meines Erachtens wäre es aktuell problematisch, Richter*innen und Rechtspfleger*innen regelmäßig zuzumuten, Experten in der Prompt-basierten Bedienung von generativen Sprachmodellen zu werden, um die Vorteile dieser Technologie nutzen zu können. Es ist daher sachgerechter, dass auf die Rechtspraxis zugeschnittene Software für gängige Aufgaben die unmittelbare Interaktion mit dem Modell übernimmt, wie das Produkt CoCounsel[22] der amerikanischen Firma CaseText beispielhaft aufzeigt. Dort kann der Nutzer zwar direkt Prompts in das System eingeben, doch im Wesentlichen stellt die Software spezielle Analysefunktionen (zB Zusammenfassungen, Essay-Erstellung, tabellarische Analysen) mittels eines traditionellen User Interfaces zur Verfügung. Obgleich davon auszugehen ist, dass CoCounsel intern mit einer von Experten entwickelten Prompt-basierten Interaktion mit GPT-4 arbeitet, wird diese vor dem Benutzer versteckt, womit eine unsachgemäße Nutzung erschwert wird.
- **"Automation Bias" & neue Arbeitsautomatismen**: "Automation Bias" bezeichnet die menschliche Tendenz, ein automatisiertes Entscheidungsunterstützungssystem als heuristischen Ersatz für gründliches und wachsames Suchen und Bearbeiten von Informationen zu verwenden.[23] Dies kann zum einen dazu führen, dass der Benutzer die Ausgaben des Systems nicht prüft und möglicherweise falsch handelt. Zum anderen kann er fahrlässig darauf vertrauen, dass in Abwesenheit eines Hinweises durch das System nichts zu tun ist, und so eine möglicherweise notwendige Handlung unterlassen.[24] Im Rahmen einer Nutzung von generiertem Texten in der Justiz muss eine Prüfung durch den Verwender sowohl die im Text explizit angesprochenen als auch die nicht vorkommenden Aspekte umfassen. Ob und wie effizient dies unter realistischen Bedingungen möglich ist, muss kontextabhängig betrachtet werden. Ferner ist zu untersuchen, wie sich Tätigkeiten bei Gericht längerfristig verändern, wenn sich generative Modelle für bestimmte Aufgaben über einen substanziellen Zeitraum als ausreichend verlässlich erwiesen haben.

---

[21] Liu, Pengfei, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. “Pre-Train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing.” ACM Computing Surveys 55, no. 9 (2023): 1–35.
[22] https://casetext.com/blog/casetext-announces-cocounsel-ai-legal-assistant
[23] Skitka, Linda J., Kathleen L. Mosier, and Mark Burdick. “Does Automation Bias Decision-Making?” International Journal of Human-Computer Studies 51, no. 5 (November 1999): 991–1006. https://doi.org/10.1006/ijhc.1999.0252.https://doi.org/10.1177/0018720810376055.
[24] Parasuraman, Raja, and Dietrich Manzey. “Complacency and Bias in Human Use of Automation: An Attentional Integration.” Human Factors 52 (June 1, 2010): 381–410. https://doi.org/10.1177/0018720810376055.

- **Beziehung zu Textbaustein-Systemen**: Im Rahmen der Bewältigung von Massenverfahren stellt sich die Frage, welche Vorteile generative Sprachmodelle gegenüber regelbasierten Modellen mit Textbausteinen bringen. Ein geeignetes Referenzbeispiel ist hier das FRAUKE-Pilotprojekt am Amtsgericht Frankfurt a.M. zur Bearbeitung von Fluggastrechteverfahren.[25] Ausgangspunkt ist, dass der Fallausgang richterlich in der Sache bereits entschieden wurde und effizient ein Urteil verfasst werden soll. Das System erlaubt die Beschreibung der Entscheidung mittels einer Formularmaske und schlägt dann geeignete Textbausteine für das Urteil vor. Diese enthalten Platzhalter-Elemente, die automatisch mit fallspezifischen Angaben aus der Akte befüllt werden, wofür ein NLP-Modul zum Einsatz kommt.[26] Für ein solches System müssen die Logik zur Entscheidungsspezifikation, die zugehörigen Textbausteine und die gezielte Extraktionsfunktionalität entsprechend entwickelt werden. Einmal erstellt verhält es sich (mit Ausnahme möglicher statistischer NLP-Komponenten) deterministisch und birgt nur minimale Unsicherheiten hinsichtlich des erstellten Urteilsentwurfs für die bereits getroffene richterliche Entscheidung. Dies stellt das konzeptionelle Gegenbeispiel zum generativen Sprachmodell dar. Letzteres ist flexibel und kann möglicherweise mit weniger Modellierungsaufwand eingesetzt werden, bringt jedoch die hier dargelegten Risiken mit sich, insbesondere wenn Teile der Entscheidung in der Sache in der Modelleingabe nicht spezifiziert werden. Möglicherweise können der generative und regelbasierte Ansatz in Mischformen synergetisch zusammenarbeiten. Meines Erachtens benötigt es hier noch experimentelle Forschung/Entwicklung und qualifizierte Diskussion in konkreten rechtlichen Anwendungskontexten.

13. Wie schätzen Sie die Problematik der Intransparenz und fehlenden Nachvollziehbarkeit solcher Programme ein?

17. Inwieweit ist die Nutzung von ChatGPT, insbesondere die durch ChatGPT generierten Texte, Textkörper, Entscheidungen und sonstigen Resultate, für die Richterschaft transparent und nachvollziehbar?

Die Erklärbarkeit von tiefen neuronalen Netzen ist ein in weiten Teilen ungelöstes Problem und Gegenstand aktiver Forschung. Generative Sprachmodelle bestehen aus vielschichtigen Berechnungen mit teils Milliarden von Parametern, deren Rolle für das Verhalten des Modells nicht unmittelbar erkennbar ist. Es existieren zahlreiche mathematische Methoden[27] zur Ableitung von "Erklärungen" aus dem Verhalten eines bestimmten Modells anhand einer Dateneingabe (sog. *lokale* Erklärbarkeit). In den meisten Fällen geschieht dies in Form einer sog. *Saliency Map*,

[25] https://www.hessenschau.de/panorama/amtsgericht-frankfurt-kuenstliche-intelligenz-hilft-bei-massen-urteilen-v1,amtsgericht-roboter-100.html

[26] Die Erklärung zur Funktionsweise des Systems entspricht dem Verständnis des Autors auf der Basis des Werkstattbericht-Vortrags von Christian Metz zu FRAUKE bei der Legal Tech Tagung 2022 der MLTech Student Association vom 15.10.2022; https://www.youtube.com/watch?v=tzRT45w9Dy8&t=2506s&ab_channel=MunichLegalTech%28MLTech%29

[27] Siehe Danilevsky, Marina, Kun Qian, Ranit Aharonov, Yannis Katsis, Ban Kawas, and Prithviraj Sen. “A Survey of the State of Explainable AI for Natural Language Processing.” In Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, 447–59. Suzhou, China: Association for Computational Linguistics, 2020. https://aclanthology.org/2020.aacl-main.46.

die Teile der Eingabe farblich hervorhebt um zu markieren, ob der jeweilige Teil des Textes für die Modellausgabe entscheidend war oder nicht bzw. die Wahrscheinlichkeit der Ausgabe beeinflusst hat.

Das deutsche Unternehmen Aleph Alpha hat in einer neuen Version seines *Luminous*-Modells eine solche Funktion integriert, mit der Teile der generierten Inhalte durch Einfärbungen mit Elementen des Eingabekontextes verknüpft werden, sofern diese in der Berechnung des Modells miteinander in Beziehung stehen.[28] Die OpenAI-Modelle ChatGPT und GPT-4 stellen meines Wissens nach aktuell eine solche Funktion nicht zur Verfügung. Die generierten Texte des früheren (und offen verfügbaren) GPT-2 Modells können allerdings mittels der Open Source Bibliothek Ecco[29] ebenfalls mittels solcher Einfärbungen untersucht werden. Es ist somit damit zu rechnen, dass künftig Einfärbungen als eine mögliche "Erklärungstechnik" in Anwendungen auf der Basis generativer Sprachmodelle weitere Verbreitung finden.

Saliency Maps sind als Erklärungstechnik begrenzt und können nur einen Teil der Transparenzbedürfnisse von Nutzern bedienen. Beispielsweise kann es wichtiger sein zu wissen, warum das Modell eine bestimmte Wortfolge *nicht* generiert hat.[30] Alternativ könnte das Modell aufzeigen, wie die Eingabe verändert werden müsste, um eine bestimmte Ausgabe zu erreichen. Ungeachtet der Einfachheit und Intuitivität von Erklärung-durch-Texteinfärbungen ist meines Erachtens im Justizkontext in jedem Fall zu erörtern, welche Form von Nachvollziehbarkeit eines generierten Textes den Anforderungen der Richter*innen und ihren rechtsgebietsspezifischen Notwendigkeiten gerecht wird. Insbesondere die Notwendigkeit einer effizienten Verifizierung der Modellausgabe vor ihrer Verwendung gibt dieser Frage besonderes Gewicht. Nach meinem Kenntnisstand ist dies noch weitestgehend ungeklärt. In einer kleinen Nutzerstudie im Bereich der Entscheidungsklassifikation kam eine amerikanische Forschungsgruppe der MITRE Corporation zu dem Ergebnis, dass anhand eines neuronalen Modells erstellte Texteinfärbungen für menschliche Entscheidungsvoraussage nicht als nützlich empfunden wurde.[31] Diese und ähnliche Ergebnisse stammen allerdings meist von diskriminativen (dh nicht oder nur eingeschränkt generativen) Klassifikations-Systemen und sind meines Erachtens nach in der aktuellen Generation von generativen Sprachmodellen neu zu bewerten.

---

[28] Siehe https://www.aleph-alpha.com/aleph-alpha-reaches-first-milestone-on-the-way-to-content-correct-explainable-and-trustworthy-ai

[29] siehe Alammar, J. “Ecco: An Open Source Library for the Explainability of Transformer Language Models.” In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: System Demonstrations, 249–57. Online: Association for Computational Linguistics, 2021. https://doi.org/10.18653/v1/2021.acl-demo.30.

[30] vgl. Yin, Kayo, and Graham Neubig. “Interpreting Language Models with Contrastive Explanations.” In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 184–98. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022. https://aclanthology.org/2022.emnlp-main.14.

[31] Branting, L. Karl, Craig Pfeifer, Bradford Brown, Lisa Ferro, John Aberdeen, Brandy Weiss, Mark Pfaff, and Bill Liao. “Scalable and Explainable Legal Prediction.” Artificial Intelligence and Law 29, no. 2 (June 1, 2021): 213–38. https://doi.org/10.1007/s10506-020-09273-1.

18. In Anlehnung an Frage Nummer 17: Auf welche Daten, Datensätze und sonstigen Inhalte greift ChatGPT zurück, um solche Texte und sonstigen Resultate wie unter Frage Nummer 3 zu generieren?

Nach meinem bestem Wissen ist unbekannt, auf welchen genauen Datensätzen die OpenAI Modelle ChatGPT und GPT-4 trainiert wurden.[32] Als Referenzpunkt kann hier die umfangreiche Untersuchung von Googles großem *C4* Datensatz der Washington Post dienen, die in Kollaboration mit dem non-profit *Allen Institute for AI* erstellt wurde.[33] Prominente Komponenten des Korpus sind beispielsweise *Google Patents* (0,46% des Gesamtkorpus), Wikipedia (0,19% des Gesamtkorpus), eine extensive Sammlung von Nachrichtenseiten, offene Bücherarchive und große Mengen persönlicher Webseiten. Obgleich Google die Inhalte proaktiv filtert, finden sich laut dem Bericht in dem Korpus unter anderem auch Seiten mit extremistischen Inhalten. Ferner dominieren christliche Inhalte die religiösen Materialien innerhalb des Datensatzes. Die Website des Untersuchungsberichts enthält ein Suchelement zur Prüfung der Inklusion von bestimmten Webseiten im Korpus. Relevante Quellen in C4 für das in Frage 3 angesprochene amerikanische Universal Bar Exam sind beispielsweise neben Seiten mit allgemeinen juristischen Inhalten (zB Urteile des von hohen Gerichten[34], dem US Code[35], dem Code of Federal Regulations[36], etc.) und frei verfügbaren Studienliteratur[37] beispielsweise auch Seiten von Anwaltskammer der amerikanischen Bundesstaaten sein, die Fragen und Antworten vergangener Bar Exams enthalten.[38] Es ist davon auszugehen, dass die aktuellen Modelle der GPT-Familie auf mindestens genauso umfangreichen Daten trainiert wurden.

19. Inwieweit werden die unter Frage Nummer 4 genannten Daten, Datensätze und sonstigen Inhalte aktualisiert und durch wen? Wie werden diese Daten und von wem durch externe schädliche Beeinflussung geschützt?

Bzgl. Datensatz-Aktualisierung: Wie erwähnt sind die genauen in ChatGPT und GPT-4 verwendeten Trainingsdaten unbekannt. Es ist davon auszugehen, dass OpenAI intern in regelmäßigen Abschnitten Internetinhalte in großem Ausmaß selbst automatisiert herunterlädt (mittels sog. *Web Scraping*) und/oder von anderen Organisationen *en bloc* bezieht[39]. Diese

---

[32] OpenAI's technischer Bericht zu GPT-4 spricht ohne detailliertere Angaben lediglich davon, dass das Modell sowohl auf öffentlich verfügbaren als auch auf von dritten lizensierten Daten trainiert wurde.
[33] Schaul, Kevin, Szu Yu Chen, und Nitasha Tiku. “Inside the Secret List of Websites That Make AI like ChatGPT Sound Smart.” Washington Post. Letzter Zugriff 27. Mai 2023. https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning.
[34] zB https://www.supremecourt.gov, Umfang laut Bericht ca. 170,000 Tausend Tokens, wobei davon auszugehen ist, dass die Einzeldokumente der Urteile nicht in diese Berechnung eingeflossen aber gleichwohl im Korpus vorhanden sind. Beispielsweise ist die Urteilssammlung https://caselaw.findlaw.com laut Bericht mit ca. 41 Millionen Tokens in C4 vertreten.
[35] https://uscode.house.gov, Umfang laut Bericht ca. 3,6 Millionen Tokens.
[36] zB via https://www.govinfo.gov, Umfang laut Bericht ca. 2,1 Millionen Tokens.
[37] zB via der kommerziellen Online-Büchersammlung https://www.scribd.com (Umfang laut Bericht ca. 100 Millionen Tokens), die auch zahlreiche Werke zur juristischen Ausbildung enthält; siehe https://www.scribd.com/books/Law/Legal-Education
[38] zB https://www.nybarexam.org/examquestions/examquestions.htm
[39] zB von https://commoncrawl.org, das unbearbeitete Internetinhalte in großen Mengen zum Download anbietet.

müssen dann typischerweise nach bestimmten Kriterien gefiltert und ausgewählt werden, um Daten hoher Qualität (sprachlich sauber, inhaltlich richtig, etc.) zu erhalten, die keinen schädigenden Charakter haben (zB rassistische oder gewaltverherrlichende Inhalte). Der so erstellte Datensatz kann periodisch aktualisiert werden. Im Fall von OpenAI liegt nahe, dass Datensatzbildung und -Wartung typischerweise innerhalb des Unternehmens geschieht.

Bzgl. Schutz vor schädlichen Beeinflussungen: Neben den erwähnten Filtern von Trainingsdaten gibt OpenAI an, seine Modelle mittels Benutzerfeedback zu "sicheren" Antworten zu trainieren. In der Entwicklung von ChatGPT wurden beispielsweise in großem Ausmaß Daten als schädlich annotiert, mit denen dem Modell schließlich das Generieren solcher Inhalte "abtrainiert" werden sollte.[40] Dennoch konnten diese internen Mechanismen umgangen werden. Wenn der Benutzer beispielsweise das Modell aufforderte, Text im Still einer bestimmten, zu toxischer Sprache tendierenden Person zu verfassen, produzierte das Modell mit höherer Wahrscheinlichkeit ebenfalls solche Inhalte.[41] GPT-4 verbesserte seine Sicherheit in Bezug auf solche und andere Phänomene. Laut dem technischen Bericht kommen hierbei u.a. menschliche Experten verschiedener Disziplinen zum Einsatz, die das Modell mittels "Adversarial Testing" zu schadhaftem Verhalten verleiten sollen, um so neue Testfälle zu sammeln, die im Training zur Steigerung der Sicherheit beitragen können.[42]

In Anlehnung an Frage Nummer 3: Bietet ChatGPT mehrere Texte und sonstigen Resultate an mit divergierenden Inhalten zu einer konkreten Anfrage (mithin einer konkreten Nutzung), die transparent und nachvollziehbar sind, mithin der dem Programm anwendenden Richterschaft eine Auswahl zwischen mehreren Texten und sonstigen Resultaten ermöglichen?

Die Frage wirft mehrere Aspekte einer konkreten Nutzungsweise auf:

Generierung mehrerer Texte: Generative Sprachmodelle produzieren eine Wahrscheinlichkeitsverteilung über Sequenzen von Worten. Es gibt zwar im Regelfall einen wahrscheinlichsten Output, doch es ist ohne weiteres möglich, mehrere Texte auf die gleiche Anfrage zu generieren, indem man kleine Abweichungen von statistischer Optimalität erlaubt (mittels des sog. *Temperature*-Parameters). Lässt die Prompt dabei dem Modell die Freiheit eigene Rechtsfolgen zu schließen, können diese zwischen verschiedenen generierten Texten trotz gleichem Input durchaus zu rechtlich verschiedenen Ergebnissen kommen.

Sonstige Inhalte: Die Kernfunktionalität von generativen Sprachmodellen selbst ist die Textgenerierung. Sie können allerdings modifiziert bzw. in Systeme integriert werden, um dem Benutzer weitere externe Inhalte zu präsentieren. So erlaubt die neue Version von Microsofts Suchmaschine *Bing*, mit einer Version von ChatGPT im Kontext eines Suchergebnisses einen

---

[40] Perrigo, Billy. “Exclusive: The $2 Per Hour Workers Who Made ChatGPT Safer.” Time, January 18, 2023. https://time.com/6247678/openai-chatgpt-kenya-workers.
[41] Deshpande, Ameet, Vishvak Murahari, Tanmay Rajpurohit, Ashwin Kalyan, and Karthik Narasimhan. “Toxicity in Chatgpt: Analyzing Persona-Assigned Language Models.” ArXiv Preprint ArXiv:2304.05335, 2023.
[42] siehe OpenAI, GPT-4 Technical Report, https://arxiv.org/abs/2303.08774, Seite 11 ff.

Dialog zu führen, im Laufe dessen Aussagen des Modells mit fußnotenartigen Referenzen auf externe Inhalte ergänzt werden, die in die generierte Antwort eingeflossen sind.[43]

Transparenz & Nachvollziehbarkeit: Es wird auf die Antworten zu Fragen 13 und 17 oben verwiesen.

20. Inwieweit werden die zur Nutzung von ChatGPT notwendigen Angaben, aus denen heraus das Programm einen Text generiert, gespeichert (vor allem: wo) und wer hat Zugriff auf diese Informationen? Was passiert mit diesen Daten? Inwieweit bestehen insoweit rechtliche Bedenken, vor allem mit Blick auf Grundrechte und datenschutzrechtliche Vorgaben?

Hinweis: Der Autor ist kein Datenschutzexperte und die Antwort basiert primär auf den offiziellen Dokumenten von OpenAI zur Datennutzung.

Zunächst ist zwischen der Nutzung mittels der OpenAI-Benutzeroberfläche und der Nutzung per API-Schnittstelle zu unterscheiden. OpenAI verwendet die von Benutzern eingegebenen Daten seiner "non-API consumer services" (zu denen auch ChatGPT gehört) zur Weiterentwicklung seiner Modelle.[44] Beispielsweise können eingegebene Prompts, der generierte Text des Systems und mögliche positive/negative Nutzerbewertungen des Systems als Trainingsdaten verwendet oder durch Entwickler gesichtet werden.

Eine Drittanwendung (zB eine Software für Anwaltskanzleien) wird im Regelfall den API-Zugang verwenden. Zum 1.3.2023 änderte OpenAI die Nutzungsbedingungen der API-Schnittstellen zu seinen Modellen zu einem Opt-In-Modell. Danach verwendet OpenAI von Nutzern übermittelte Daten zum Training und zur Entwicklung des Modells nur sofern der Nutzer explizit einwilligt.[45] Laut diesem Dokument werden die Daten zunächst von OpenAI selbst sowie von seinen "Sub-Processors"[46] (d.h. Firmen, die auf Vertragsbasis für OpenAI Datenverarbeitung betreiben) gespeichert. Sofern die Nutzungseinwilligung nicht vorliegt, werden die Daten laut den Bedingungen lediglich für 30 Tage zur Missbrauchsüberwachung gespeichert und dann gelöscht. Innerhalb dieses Zeitraums haben Angestellte von OpenAI und spezialisierte Vertragspartner entsprechen zweckgebunden Zugriff auf diese Daten.[47]

Es sei erwähnt, dass aktuell um ChatGPT ein Plugin-Ökosystem entsteht.[48] Hierbei können Webseiten und Software Dritter sich mit ChatGPT verbinden und beidseitig Funktionalitäten

---

[43] https://blogs.microsoft.com/blog/2023/02/07/reinventing-search-with-a-new-ai-powered-microsoft-bing-and-edge-your-copilot-for-the-web
[44] https://help.openai.com/en/articles/7039943-data-usage-for-consumer-services-faq
[45] Open AI APU Data Usage Policies: https://openai.com/policies/api-data-usage-policies ; zuletzt abgerufen am 29.5.2023.
[46] https://platform.openai.com/subprocessors ; zuletzt abgerufen am 29.5.2023.
[47] Originalformulierung: "OpenAI retains API data for 30 days for abuse and misuse monitoring purposes. A limited number of authorized OpenAI employees, as well as specialized third-party contractors that are subject to confidentiality and security obligations, can access this data solely to investigate and verify suspected abuse. OpenAI may still have content classifiers flag when data is suspected to contain platform abuse."
[48] https://openai.com/blog/chatgpt-plugins ; zuletzt abgerufen am 29.5.2023.

anbieten. Zum Beispiel bietet das Plugin *askyourpdf.com* die Möglichkeit, eine PDF-Datei hochzuladen und einer ChatGPT-Instanz über das darin enthaltene Dokument Fragen zu stellen, die sie dann natursprachlich beantwortet. Plugins sind separate Software und haben dementsprechend separate Nutzungsbedingungen, die eine weitergehende Datenspeicherung enthalten können.

Zur Beeinträchtigung von Grundrechten und Datenschutzrechtlichen Aspekten dieser möglichen Datenspeicherung und/oder -Verarbeitung kann ich keine sachverständigen Antworten geben.

Ich hoffe mit dieser Stellungnahme ihre Fragen zufriedenstellend beantwortet zu haben.

Matthias Grabmair, Ph.D., LL.M.