

**Richter am Landgericht
Dr. Christian Schlicht**

**Richter am Landgericht
Prof. Dr. Simon J. Heetkamp, LL.M.**

An den
Präsidenten des Landtags
Nordrhein-Westfalen



Köln, 23.05.2023

Einsatz von ChatGPT im Justizbereich

Vorlage 18/1022

Anhörung des Rechtsausschusses am 13.06.2023

Sehr geehrte Damen und Herren,

für die Einladung zu der vorgenannten Anhörung bedanken wir uns und nehmen vorab wie folgt Stellung. Zur Vorbereitung formulieren wir fünf Thesen zu möglichen Veränderungen im Justizbereich durch einen Einsatz von Large-Language-Modellen (fortan: LLMs) (**Teil A.**), bevor wir uns mit den Fragen des Rechtsausschusses befassen (**Teil B.**).

Teil A.

Unsere Überlegungen zu einem möglichen Einsatz von LLMs im Justizbereich lassen sich in den folgenden Thesen, zu denen sodann ausgeführt wird, zusammenfassen:

1. Status quo: Keine weitreichenden Einsatzmöglichkeiten von ChatGPT in der Rechtsprechung
2. ChatGPT ist nur ein Beispiel für weitere, zu beachtende LLMs und KI-Anwendungen; perspektivisch: Einsatz eines Justiz-LLM
3. Justiz wird in großem Umfang Empfänger von KI-erstellten Texten
4. Überlegungen zum Einsatz von LLMs nicht auf Richterschaft beschränken
5. Justiz-Task Force sollte Anwendungsfelder von LLMs praktisch erproben und Entwicklungen auf dem Rechtsmarkt beobachten

These 1: Status quo: Keine weitreichenden Einsatzmöglichkeiten von ChatGPT in der Rechtsprechung

Die Verfasser haben ausgiebig erprobt, ob/wie ChatGPT¹ bei dem Entwurf von Entscheidungen unterstützen könnte.

a) Im Ausgangspunkt stehen der Eingabe konkreter Sachverhaltsinformationen aus realen Gerichtsverfahren in eine browserbasierte Version von ChatGPT bereits datenschutzrechtliche Bedenken entgegen, so dass sich derzeit die Frage nach einer Unterstützung durch ChatGPT jedenfalls für alle konkret verfahrensbezogenen Texte² in der Praxis nicht stellt. Lediglich für Textbestandteile, in denen es um die abstrakte Darstellung etwa von technischen oder rechtlichen Fragen geht, könnten per KI erstellte Textpassagen in die richterliche Arbeit einfließen:

Beispiel:

Aufforderung an ChatGPT:

Beschreibe, was eine polymermodifizierte Bitumendickbeschichtung ist!

¹ Soweit nicht besonders gekennzeichnet, sind mit „ChatGPT“ alle zum Zeitpunkt der Stellungnahme verfügbaren Varianten von ChatGPT gemeint.

² D.h. insbesondere den Tatbestand im Urteil, Subsumtionen in den (Entscheidungs-)Gründen und Beweiswürdigungen.

Antwort von ChatGPT (von den Verfassern leicht gekürzt):

Eine polymermodifizierte Bitumendickbeschichtung (PMBD) ist eine Art von Flüssigabdichtungssystem, das hauptsächlich für die Abdichtung von Gebäudedächern und anderen Bauwerken verwendet wird. Es besteht aus einer Mischung aus Bitumen und synthetischen Polymeren, die chemisch modifiziert wurden, um ihre physikalischen Eigenschaften zu verbessern. Diese Polymermodifikation verbessert die Eigenschaften des Bitumens wie seine Elastizität, Flexibilität, Dichtigkeit, Haftung und Dauerhaftigkeit.

Es gibt verschiedene Arten von Polymeren, die in der PMBD verwendet werden können, darunter Styrol-Butadien-Styrol (SBS), Styrol-Ethylen-Butylen-Styrol (SEBS), Ethylen-Propylen-Dien-Monomer (EPDM) und Polyvinylchlorid (PVC). Jeder dieser Polymere hat unterschiedliche Eigenschaften, die sich auf die Leistung und Anwendung der PMBD auswirken können.

Die PMBD besteht aus mehreren Schichten, die auf das Dach aufgetragen werden. Jede Schicht kann aus einer anderen Art von Bitumen und Polymeren bestehen, um eine optimale Kombination von Eigenschaften zu erzielen. Die PMBD wird normalerweise auf eine Trägermatte aufgebracht, um ihre Stabilität und Dicke zu erhöhen und um sicherzustellen, dass die PMBD in allen Bereichen des Dachs gleichmäßig verteilt wird. (...)

Insgesamt ist die polymermodifizierte Bitumendickbeschichtung ein hochwertiges Abdichtungssystem, das für seine hervorragenden Leistungseigenschaften und seine Langlebigkeit bekannt ist. Es wird in vielen verschiedenen Anwendungen eingesetzt, darunter Gebäudedächer, Terrassen, Balkone, Parkdecks und Brückenabdichtungen.

ChatGPT antwortet auf eine kurze Aufforderung mit umfangreichen Ausführungen, die sich noch mit einer weiteren Aufforderung ergänzen ließen. Dies könnte dem Gericht zeitintensive Schreibarbeit ersparen und Gedankenanstöße bieten; auch liefere das Gericht nicht Gefahr, (unbewusst) die schriftlichen Ausführungen einer Partei zu übernehmen. Bei einem solchen Vorgehen müsste das Gericht darauf achten, dass die von ChatGPT generierten Ausführungen inhaltlich zutreffend sind und gegebenenfalls an-

passen. Für den Zivilprozess wäre den Parteien regelmäßig eine Möglichkeit zur Stellungnahme einzuräumen.³ Soweit es sich um streitige Tatsachen handelt, würde das Gericht ohnehin ein Sachverständigengutachten einholen, so dass sich das vorgenannte Vorgehen erübrigte.

b) Bezüglich der rechtlichen Recherche bietet ChatGPT keine Vorteile gegenüber dem herkömmlichen Einsatz von Suchmaschinen und insbesondere juristischen Datenbanken. Nach unseren Tests ist ChatGPT beim Erarbeiten und Abfassen deutscher Rechtstexte (noch) so schlecht, dass die Antworten in aller Regel unbrauchbar sind. Einfache Sachverhalte mögen LLMs noch mit einer im Ergebnis richtigen rechtlichen Einschätzung beantworten, doch scheitern sie schon an leicht gehobenen Rechtsproblemen. Auch sind belastbare Begründungen nicht zu erzielen. Häufig kann ChatGPT noch nicht einmal die für einen Sachverhalt entscheidungserheblichen Rechtsnormen vollständig identifizieren.

Soweit in der (Fach-)Öffentlichkeit und auch in Frage Nr. 8 des Rechtsausschusses auf ein Urteil eines kolumbianischen Richters abgestellt wird, das auch von ChatGPT erstellte Passagen enthält, werden die – als von Chat-GPT erstellt gekennzeichneten – Passagen von der Fachöffentlichkeit als oberflächlich, missverständlich oder unzutreffend kritisiert.⁴

c) Im Ergebnis gehen die Verfasser davon aus, dass jenseits der vorgenannten, begrenzten Anwendungsbeispiele gegenwärtig keine weitreichenden Einsatzmöglichkeiten im Kernbereich richterlicher Tätigkeit bestehen. Dabei ist auch darauf hinzuweisen, dass die gängigen Textbausteine für verfahrensbegleitende Verfügungen – etwa zur Einleitung des schriftlichen Vorverfahrens, zur Terminbestimmung oder zur Fristverlängerung – und für Beschlüsse – etwa hinsichtlich der (vorläufigen) Festsetzung des Streitwerts, Hinweis auf Unzuständigkeit des angerufenen Gerichts, die (Nicht-)Gewährung von Prozesskostenhilfe – in Nordrhein-Westfalen im *TextSystem Justiz* (TSJ) bereits enthalten sind. Diese Vorstücke werden zentral durch die zuständige Verfahrenspflegestelle bereitgestellt. Innerhalb der TSJ-Vorlagen besteht die Möglichkeit, diese individuell anzupassen, einzelne Textbausteine aus-/abzuwählen und dy-

³ Vgl. BGH, Beschluss vom 27.01.2022, Az. III ZR 195/20 im Hinblick auf eine gerichtliche Internetrecherche.

⁴ Siehe dazu auch die Beantwortung der Frage Nr. 8 in Teil B. dieser Stellungnahme.

namische Angaben (automatisiert) zu berücksichtigen (z.B. die durch die Serviceeinheit vorher erfassten Zeugen in einem Beweisbeschluss anzuwählen). Zudem können die Vorstücke von der Richterschaft individualisiert, entsprechend gespeichert und geteilt werden. Nahezu alle Mitglieder der Richterschaft arbeiten zugleich mit eigenen Textbausteinen, die bei Bedarf in die Entscheidung einkopiert werden können. Hinsichtlich einer etwaigen Verringerung von Schreiarbeit ist auch zu berücksichtigen, dass in der NRW-Justiz mit der Diktiersoftware „Dragon“ gearbeitet werden kann, die ein schnelles und effizientes Diktat von Schriftstücken ermöglicht.

Im Ergebnis sehen wir in dem Bereich der verfahrensbezogenen Schriftstückerstellung für die Richterschaft derzeit keine wesentlichen Anwendungsfälle für Unterstützungsmöglichkeiten durch ChatGPT.

These 2: ChatGPT ist nur ein Beispiel für weitere, zu beachtende LLMs und KI-Anwendungen; perspektivisch: Einsatz eines Justiz-LLM

Die weitere Debatte sollte sich nicht auf ChatGPT oder entsprechende Nachfolgeprogramme der Firma OpenAI beschränken. Zwar kann der Einsatz von LLMs in der Justiz anhand von ChatGPT beispielhaft erläutert werden. ChatGPT ist aber weder der Ausgangspunkt noch das für Justizzwecke am besten geeignete Programm.⁵ Von Anfang an sollten auch andere LLMs und KI-Anwendungen bedacht werden.

a) Schon heute werden KI-Sprachtools (unter Berücksichtigung bestimmter Vorgaben) in der Justiz gewinnbringend eingesetzt. Aus der Praxis kann berichtet werden, dass Richterinnen und Richter das KI-Sprachtool DeepL Write⁶ als „Lektoratsersatz“ einsetzen, um die Qualität der schriftlichen Ausführungen zu steigern.⁷ Denn ein „Vier-Augen-Prinzip“ (allein mit Blick auf sprachliche und grammatikalische Aspekte) ist bei Einzelrichterentscheidungen nicht vorgesehen. Aber auch bei Kammerentscheidungen gibt die KI-Assistenz oft wertvolle Verbesserungsvorschläge.

⁵ Ein wesentliches Problem besteht darin, dass die Daten außerhalb der EU gespeichert werden. Jedenfalls dieses Problem ließe sich durch den Einsatz eines LLM aus Deutschland lösen.

⁶ Nähere Informationen unter <https://www.deepl.com/de/write>.

⁷ Die Verfasser haben als Initiatoren der „digitalen richterschaft“, einer Austauschplattform zu Digitalisierungs- und Innovationsprozessen in der Justiz, entsprechende Erfahrungsberichte vorliegen.

Die Funktionsweise wird anhand der folgenden Beispiele deutlich:

The screenshot shows the DeepL Write interface. At the top left is the DeepL Write logo and a dropdown menu. At the top right is the 'Anmelden' button and a hamburger menu icon. The main heading is 'Bessere Texte im Handumdrehen' with the subtitle 'Schreiben Sie klar, präzise und fehlerfrei'. Below this is a language selector set to 'Deutsch' and a bidirectional arrow icon. The interface is split into two columns. The left column contains the original text: 'Denn im Anwendungsbereich des Art. 5 Abs. 1 Satz 1 GG liegt in jeder Verwaltungsmaßnahme, die die Meinungsäußerung und -verbreitung verbietet, behindert oder gebietet, ein relevanter Eingriffsakt (Schemmer, in: Epping/Hillgruber, BeckOK GG, Stand November 2022, Art. 5 Rn. 18)'. The right column shows the corrected text: 'Denn im Anwendungsbereich des Art. 5 Abs. 1 Satz 1 GG ist jede behördliche Maßnahme, die die Äußerung und Verbreitung von Meinungen verbietet, behindert oder anordnet, ein relevanter Eingriffsakt (Schemmer, in: Epping/Hillgruber, BeckOK GG, Stand November 2022, Art. 5 Rn. 18)'. The corrected text has several words underlined in green.

The screenshot shows the DeepL Write interface. At the top left is the DeepL Write logo and a dropdown menu. At the top right is the 'Anmelden' button and a hamburger menu icon. The main heading is 'Bessere Texte im Handumdrehen' with the subtitle 'Schreiben Sie klar, präzise und fehlerfrei'. Below this is a language selector set to 'Deutsch' and a bidirectional arrow icon. The interface is split into two columns. The left column contains the original text: 'Auch sogenannten hinkenden Ehen, die - etwa bei unterschiedlichem Eheschließungsstatut der Eheschließenden - nach der einen beteiligten Rechtsordnung wirksam, nach der anderen aber unwirksam sind, kommt der Schutz des Ehegrundrechts zu (vgl. BVerfGE 62, 323 <330 f.>).'. The right column shows the corrected text: 'Auch sogenannte hinkende Ehen, die - etwa bei unterschiedlichem Eheschließungsstatut der Ehegatten - nach einer der beteiligten Rechtsordnungen wirksam, nach der anderen aber unwirksam sind, genießen den Schutz des Ehegrundrechts (vgl. BVerfGE 62, 323 <330 f.>).'. The corrected text has several words underlined in green.

Typischerweise kommt DeepL Write nach dem ersten Abfassen der Entscheidungsgründe in der Korrekturphase zum Einsatz. Noch ist die Handhabung relativ umständlich, da die Kolleginnen und Kollegen nur das auf der Website frei verfügbare Tool nutzen, das in der Bearbeitung auf 2.000 Zeichen beschränkt ist. Entscheidungen können daher nur abschnittsweise überprüft werden. Zudem können nur anonymisierte Passagen korrigiert werden. Die Einarbeitung der Vorschläge ist mit einem zusätzlichen Zeitaufwand verbunden, erhöht aber spürbar die Qualität der Entscheidungen. Der Einsatz ist daher derzeit vor allem bei häufig verwendeten Bausteinen sinnvoll.

Die Fehlerquote bei Grammatik- und Rechtschreibfehlern tendiert bei DeepL Write gegen Null. Die sprachlichen Vorschläge dürften nach Schätzungen der Kolleginnen und Kollegen in mehr als 75% der Fälle hilfreich sein. Der Mehrwert ergibt sich vor allem

aus der Summe der Vorschläge. Die Texte werden dadurch für die anderen Verfahrensbeteiligten verständlicher und präziser. Insgesamt sind die Kolleginnen und Kollegen von der digitalen Unterstützung sehr angetan und würden eine Implementierung dieser (oder einer vergleichbaren) KI-Unterstützung in Word bzw. die eAkte begrüßen. In vielen (Groß-)Kanzleien ist DeepL Write bereits in das Textverarbeitungsprogramm Word integriert und kommt dadurch standardmäßig zum Einsatz.

b) Zu erwähnen sind zudem KI-Sprachassistenten, die bei angefangenen Sätzen die nächsten Worte vorschlagen bzw. den restlichen (Ab-)Satz vervollständigen. Dies geschieht entweder durch Rückgriff auf nutzerseits geschriebene Texte oder durch stochastische Berechnungen. So könnten entsprechende Programme etwa den folgenden Einleitungsabsatz, der von vielen Gerichten zu Beginn einer Auslegung von Allgemeinen Versicherungsbedingungen bemüht wird, nach der Eingabe der ersten Worte erkennen und automatisch vervollständigen:

„Allgemeine Versicherungsbedingungen sind so auszulegen, wie ein durchschnittlicher, um Verständnis bemühter Versicherungsnehmer sie bei verständiger Würdigung, aufmerksamer Durchsicht und unter Berücksichtigung des erkennbaren Sinnzusammenhangs versteht. Dabei kommt es auf die Verständnismöglichkeiten eines Versicherungsnehmers ohne versicherungsrechtliche Spezialkenntnisse und damit auch auf seine Interessen an. In erster Linie ist vom Bedingungswortlaut auszugehen. Der mit dem Bedingungswerk verfolgte Zweck und der Sinnzusammenhang der Klauseln sind zusätzlich zu berücksichtigen, soweit sie für den Versicherungsnehmer erkennbar sind (Senat BGHZ 232, 344 = NJW 2022, 872 Rn. 10; stRspr).“

Entsprechende KI-Sprachassistenten könnten so konfiguriert werden, dass nur zuvor validierte oder durch die Person bereits selbst eingegebene Textpassagen automatisch vervollständigt werden. Die Richterschaft könnte durch entsprechende Autovervollständigungen erhebliche Zeitersparnisse bei der Erstellung von Entscheidungen erzielen. Auch hier wäre eine Integration in Word bzw. die eAkte sinnvoll.

c) Jenseits der gegenwärtigen Diskussionen um ChatGPT sollte die Justiz den grundsätzlichen und perspektivischen Einsatz von LLMs durchdenken. LLMs können schon heute allein oder ergänzend auf der Grundlage eigener Textdateien trainiert werden

(sog. *customized large language models*). Die Kanzlei Allen & Overy arbeitet etwa mit dem KI-Programm „Harvey“, das auf ChatGPT beruht, mit Daten der Kanzlei trainiert wurde und fortlaufend trainiert wird.⁸ Entsprechend könnte die Justiz ein LLM auf der Grundlage von Entscheidungen eines oder mehrerer Gerichte trainieren und dann Entscheidungen in einem Fachgebiet im Duktus des Gerichts/der Gerichte entwerfen. Sollten diese *customized large language models* auf justizeigenen Servern betrieben werden, dürfte dies datenschutzrechtliche Bedenken minimieren. In diesem perspektivischen Szenario könnte ein Justiz-LLM ein wahrer Gamechanger für die Arbeit der Justiz sein. Beispielhaft können folgende Anwendungsmöglichkeiten genannt werden:

- Extrahieren von Metadaten in Bezug auf Eingänge;
- Übersetzen von Texten in andere Sprachen oder leichte Sprache;
- Schreibassistenz für den richterlichen und nichtrichterlichen Bereich;
- Vergleichsfunktion zwischen Schriftsätzen verschiedener Verfahren – sehr hilfreich in Massenverfahren wie sog. „Diesel“-, Fluggastrechte-, Beitragsanpassungsverfahren;
- Anonymisierungsfunktion;
- Umfassende Co-Pilot- oder Assistenzfunktion
 - LLM weist auf Widersprüche oder Logikfehler in der Arbeit hin,
 - LLM erstellt Zusammenfassungen von Aktenteilen oder recherchierten Urteilen,
 - LLM erstellt Relationstabelle mit Nachweis der Fundstellen;
 - LLM sucht Akteninhalte heraus und bereitet diese auf – etwa Extrahieren aller benannten Zeugen mit ladungsfähiger Adresse,
 - LLM unterstützt bei weiteren Verfahrensschritten – etwa Vorschlag eines zu bestellenden Sachverständigen anhand des Beweisthemas,
 - LLM unterstützt bei der Erstellung von Leitsätzen.

Die vorstehenden Beispiele zeigen, dass ein Justiz-LLM nicht isoliert gedacht werden sollte; vielmehr müsste es mit den bestehenden Justizprogrammen verknüpft werden, um einen maximalen Wirkungsgrad zu erzielen. Die Entwicklung und der Betrieb eines solchen Justiz-Sprachmodells setzen entsprechende finanzielle, personelle und organisatorische Ressourcen (etwa Rechenleistung) voraus.

⁸ <https://www.juve.de/markt-und-management/allen-overly-kooperiert-mit-ki-plattform/>.

These 3: Justiz wird in großem Umfang Empfänger von KI-erstellten Texten

Die Justiz wird sich darauf einstellen müssen, in großem Umfang Empfänger von KI-erstellten Texten zu sein. Dabei kommen als Absender insbesondere rechtssuchende Bürgerinnen und Bürger (hierzu a)) und die Anwaltschaft in Betracht (hierzu b)). Die Justiz wird reagieren müssen (hierzu c)).

a) Einsatz von Large Language Models durch die rechtssuchende Bevölkerung

Die Verbreitung von LLMs könnte dazu führen, dass sich eine größere Zahl von Bürgerinnen und Bürger in der Lage sieht, sich in den nicht vom Anwaltszwang erfassten Gerichtsverfahren selbst zu vertreten. Denn auf eine kurze Sachverhaltsschilderung generiert ChatGPT ein Schreiben an das Gericht, das manche Bürgerin bzw. mancher Bürger als ordnungsgemäße und ausreichende Klage ansehen könnte.

Aufforderung an ChatGPT:

„Ich habe von Herrn Meier am 5.5.2022 einen PKW für 4.000 EUR gekauft. Jetzt ist der Motor kaputt und ich bin vom Vertrag durch Schreiben vom 20.03.2023 zurückgetreten und habe eine Frist bis zum 27.03.2023 zur Rückzahlung des Kaufpreises gesetzt. Herr Meier hat nicht gezahlt. Verfasse eine Klage auf Rückzahlung des Kaufpreises an das Amtsgericht Köln!“

Antwort von ChatGPT:

„Sehr geehrte Damen und Herren,

hiermit erhebe ich Klage auf Rückzahlung des Kaufpreises in Höhe von 4.000 EUR gegen Herrn Meier, wohnhaft in Köln, wegen eines fehlerhaften PKW-Kaufvertrags, den ich am 5.5.2022 abgeschlossen habe.

Ich habe den PKW für 4.000 EUR gekauft, aber der Motor ist defekt. Ich bin am 20.03.2023 vom Vertrag zurückgetreten und habe Herrn Meier eine Frist bis

zum 27.03.2023 zur Rückzahlung des Kaufpreises gesetzt. Trotz meiner Anforderung hat Herr Meier bis heute nicht gezahlt.

Ich fordere hiermit die Rückzahlung des vollen Kaufpreises in Höhe von 4.000 EUR sowie die Erstattung der Kosten, die mir durch den Kauf des defekten PKWs entstanden sind.

Ich bitte Sie, das Gerichtsverfahren einzuleiten und Herrn Meier zur Zahlung zu verurteilen.

Mit freundlichen Grüßen,

[Name des Klägers]“

b) Einsatz von Large Language Models in der Anwaltschaft

Bei der Auseinandersetzung mit den Auswirkungen von LLMs auf die Justiz sollten auch die entsprechenden Einsatzmöglichkeiten in der Anwaltschaft bedacht werden, da sich diese mittelbar auf die Justiz auswirken werden.⁹ So ist etwa damit zu rechnen, dass die Anwaltschaft vermehrt LLMs einsetzen wird, um Schriftsätze zu erstellen. Dies könnte dazu führen, dass anwaltliche Schriftsätze deutlich umfangreicher werden. Mussten Rechtsanwältinnen und Rechtsanwälte bislang (jedenfalls beim erstmaligen Erstellen eines entsprechenden Schriftsatzes) etwa technische Sachverhaltsbeschreibungen selbstständig ausarbeiten, könnte dies nun weitgehend durch LLMs übernommen werden.¹⁰ Auch könnte die Anwaltschaft aus von der Mandantschaft mitgeteilten Informationen einen Schriftsatz(-entwurf) durch ein LLM erstellen lassen. Des Weiteren ist denkbar, dass die Anwaltschaft LLMs dazu einsetzt, die Arbeit der Gericht zu überprüfen. So kann ChatGPT einen Text auf Logikfehler und Widersprüche überprüfen.

⁹ Bereits der basale Einsatz von Legal Tech-Tools insbesondere in Massenverfahren führt zu Diskussionen um einen befürchteten „Kollaps der Ziviljustiz“, vgl. hierzu <https://www.bundestag.de/dokumente/textarchiv/2023/kw19-pa-recht-massenverfahren-945422>.

¹⁰ Siehe ein entsprechendes Beispiel aus Perspektive der Richterschaft unter These 2 a).

Aufforderung an ChatGPT:

Sind in der folgenden Beweiswürdigung Widersprüche oder Logikfehler erkennbar? [an dieser Stelle wurde von den Verfassern eine fiktive Beweiswürdigung nebst einer schriftlichen Darstellung und Bewertung von Parteianhörungen eingekopiert]

Antwort von ChatGPT:

Nein, es sind keine Widersprüche oder Logikfehler in der obigen Beweiswürdigung erkennbar. Der Text gibt eine umfassende Darstellung der Aussagen der Beklagten zu 1) und deren Glaubwürdigkeit sowie der Frage, ob und wann der Kläger über den Deckungsausschluss informiert wurde. Dabei werden auch mögliche Ungereimtheiten in den Aussagen der Beklagten aufgezeigt und diskutiert. Die Argumentation ist schlüssig und nachvollziehbar und es werden keine Widersprüche oder Logikfehler erkennbar.

Entsprechende Textpassagen könnten – mithilfe von ChatGPT beliebig erweiterbar – bei Vorliegen von Widersprüchen oder Logikfehlern z.B. in eine Berufungserwiderung kopiert werden, wenn der Berufungsführer die Beweiswürdigung des erstinstanzlichen Gerichts rügen will.

c) Mögliche Reaktionen der Justiz

Sollte sich durch die Nutzung von KI, insbesondere LLMs, die Anzahl und der Umfang von Eingängen bei Gericht erhöhen, könnte dies eine Anpassung des Personalbedarfs erfordern. Sollten Parteien vermehrt ohne anwaltlichen Beistand auftreten, wäre auch dies zu beachten.

In diesem Zusammenhang könnte erwogen werden, ob Bürgerinnen und Bürgern eine erweiterte, staatliche Hilfestellung angeboten werden sollte, um eine Anspruchsdurchsetzung zu erleichtern (etwa digitale Rechtsantragsstellen oder digitale Klagewege).

Im Einzelfall könnten sich zudem prozessuale Fragen in Bezug auf KI-generierte Texte stellen, etwa wenn kein ausreichender Bezug zum Einzelfall gegeben wäre.¹¹

Klar ist: Die Justiz kann gegenwärtig nicht ChatGPT mit ChatGPT begegnen, also nicht auf ChatGPT-generierte Text mit einer eigenen Nutzung von ChatGPT reagieren. Umso dringlicher erscheint es den Verfassern, dass die übrigen Digitalisierungsdefizite in der Justiz angegangen beseitigt werden.¹²

These 4: Überlegungen zum Einsatz von LLMs nicht auf Richterschaft beschränken

Soweit die (Fach-)Öffentlichkeit den Einsatz von ChatGPT in der Justiz diskutiert, wird die Debatte allzu häufig auf die Rechtsprechung, zuweilen sogar die Kerntätigkeit der Urteilsabfassung, verengt. Wie bei dem Großteil der KI-Anwendungen sind die Einsatzmöglichkeiten und Effizienzgewinne im nichtrichterlichen Bereich um ein Vielfaches höher.¹³

Uns ist bekannt, dass in anderen Bundesländern ChatGPT im justiznahen Verwaltungsbereich bereits erprobt wird, um mögliche Anwendungsbereiche auf den Geschäftsstellen, für die Kostenbeamtinnen und -beamten sowie den Rechtspflegerbereich zu identifizieren und zu erproben. Aus eigener Erfahrung kann berichtet werden, dass es Serviceeinheiten zuweilen schwerfällt, durch den Richter oder die Richterin nur umschriebene Mitteilungen in ein gerichtliches Informationsschreiben zu „übersetzen“. Hierbei könnte ein LLM hilfreich sein.

Beispiel:

Eine Richterin verfügt für die Serviceeinheit: *„Schreiben an Zeugen Schneider, dass sein Erscheinen zum Termin entbehrlich ist.“*

¹¹ Das OLG Köln hat eine Berufung als unzulässig verworfen, weil sie aufgrund des Einsatzes von Textbausteinen den gesetzlichen Anforderungen an die Berufungsbegründung nicht genüge, vgl. OLG Köln Beschl. v. 18.08.2020 – Az. 15 U 171/19.

¹² Siehe dazu auch die Stellungnahme der Verfasser zu Einsatzmöglichkeiten und Grenzen für KI in der Justiz: Stellungnahme 18/163 vom 03.01.2023 und die Stellungnahme von Sina Dörr zur öffentlichen Anhörung des Rechtsausschusses des Deutschen Bundestages vom 10.05.2023 zum Thema „Kollaps der Ziviljustiz verhindern – Wirksame Regelungen zur Bewältigung von Massenverfahren schaffen“, abrufbar unter: <https://www.bundestag.de/resource/blob/947116/a9b1e232ac77b2d18650b58ae3577c9e/Stellungnahme-Doerr-data.pdf>.

¹³ Siehe dazu auch die Stellungnahme der Verfasser zu Einsatzmöglichkeiten und Grenzen für KI in der Justiz: Stellungnahme 18/163 vom 03.01.2023.

Aufforderung an ChatGPT:

Entwirf ein Schreiben an Zeugen Schneider, dass sein Erscheinen zum Termin entbehrlich ist.

Antwort von ChatGPT (von den Verfassern leicht gekürzt):

Betreff: Entbehrlichkeit Ihres Erscheinens zum Termin

Sehr geehrter Zeuge Schneider,

ich hoffe, diese Mitteilung erreicht Sie rechtzeitig und in bester Verfassung. Im Zusammenhang mit dem bevorstehenden Termin als Zeuge in der Angelegenheit [Beschreiben Sie die Angelegenheit oder den Fall] möchte ich Ihnen mitteilen, dass Ihr persönliches Erscheinen zum Termin entbehrlich ist. (...)

Wir schätzen Ihre Bereitschaft, als Zeuge auszusagen, und möchten Ihnen dafür danken. Ihre Zusammenarbeit war äußerst wertvoll und hat zur Klärung dieses Falls beigetragen.

Sollten sich weitere Entwicklungen ergeben, die Ihre Anwesenheit erfordern, werden wir Sie umgehend informieren. In der Zwischenzeit müssen Sie sich nicht weiter auf den Termin vorbereiten oder Sorgen um Ihre Teilnahme machen.

Für den Fall, dass Sie Fragen oder Bedenken haben, stehe ich Ihnen gerne zur Verfügung. Sie können mich per Telefon [Ihre Telefonnummer] oder per E-Mail [Ihre E-Mail-Adresse] erreichen.

Vielen Dank für Ihr Verständnis und Ihre Kooperation.

Mit freundlichen Grüßen,

[Ihr Name]

Ein weiterer Anwendungsbereich wäre etwa die Gerichtsverwaltung. Hier können z.B. Hausmitteilungen, Einladungen und Informationen an Referendarinnen und Referendare oder Proberichterinnen und -richter sowie die Zusammenfassung der Berichte der Amtsgerichte im Bezirk eines Landgerichts im Rahmen der Berichtspflicht gegenüber dem Oberlandesgericht genannt werden.

Dass die vorstehenden Beispiele zeitnah umsetzbar sind, zeigt ein Bericht des Gov-Tech Campus Deutschlands über die Zusammenarbeit des Innovations-Labs des Staatsministeriums Baden-Württemberg mit Aleph Alpha hinsichtlich der Anwendung „F13“, die auf generativer KI basiert. Die Anwendung kann etwa eine automatische Zusammenfassung von umfangreichen Texten auf wenige Zeilen generieren, als Textassistent bei der Recherchearbeit unterstützen oder Vermerke erstellen.¹⁴

These 5: Justiz-Task Force sollte Anwendungsfelder von LLMs praktisch erproben und Entwicklungen auf dem Rechtsmarkt beobachten

Wie KI-Programme rechtskonform und effizient eingesetzt werden können, ist kein Allgemeinwissen. Daher sollte die Justiz jetzt beginnen, Justizbedienstete entsprechend zu schulen. Dabei sollten die Beteiligten – aufbauend auf These 4 – nicht nur aus dem richterlichen Dienst, sondern ebenso aus dem nichtrichterlichen Dienst stammen. Durch entsprechend geschulte Justizbedienstete könnten rechtskonforme Einsatzmöglichkeiten von LLMs zur Effizienzsteigerung in der Justiz erarbeitet und erprobt werden. Zudem sollte die weitere Entwicklung des Einsatzes von LLMs von anderen Akteuren im Rechtsmarkt engmaschig beobachtet werden.

Teil B.

Gerne nehmen wir zu den Fragen des Rechtsausschusses wie folgt Stellung:

1. In welchem rechtlichen Verhältnis stehen die „richterliche Unabhängigkeit“ des Art 97 GG und das Recht jeden Einzelnen auf den „gesetzlichen Richter“ Art 101 I 1 GG bzgl. der Verwendung von ChatGPT oder einer vergleichbaren KI-Software durch die Justiz?

Derzeit ist die wohl einhellige Meinung in Praxis und Rechtswissenschaft, dass KI-Systeme menschliche Richterinnen und Richter nicht ersetzen und deren richterliche

¹⁴ Siehe entsprechende Pressemitteilung des Staatsministeriums Baden-Württemberg: <https://stm.baden-wuerttemberg.de/de/service/presse/meldung/pid/kuenstliche-intelligenz-in-der-verwaltung>.

Unabhängigkeit auch nicht gefährden dürfen. Richterinnen und Richter dürfen kraft ihrer unabhängigen Entscheidungsfindung auch nicht (mittelbar) gehalten sein, KI-Systeme für ihre Tätigkeit nutzen zu müssen. Insofern muss bei der Personalbedarfsberechnung (PeBB§y¹⁵) sichergestellt werden, dass Richterinnen und Richter nicht faktisch auf KI-Lösungen zurückgreifen müssen, weil dies in die Berechnung der Bearbeitungszeit „eingepreist“ ist.¹⁶

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“?

Nach den bisherigen Erfahrungen der Verfasser sind LLMs derzeit noch weit davon entfernt, Entscheidungen in einer überzeugenden Tiefe zu formulieren. Falls sich dies perspektivisch ändern sollte, sehen wir eine Unterstützung beim Erstellen von Entscheidungen im Grundsatz als unbedenklich an, sofern die Letztentscheidung weiter beim Menschen liegt. Im Ausgangspunkt ist es schon heute so, dass Richterinnen und Richtern Entscheidungsentwürfe – etwa durch Referendare bzw. Referendarinnen oder Justizassistenzen¹⁷ – vorgelegt werden, die sie überprüfen und überarbeiten müssen. Auch ist es gelebte Praxis in der Richterschaft, sich gegenseitig Entscheidungen zu (vermeintlich) vergleichbaren Sachverhalten oder Rechtsfragen zur Verfügung zu stellen, die von dem Gericht daraufhin geprüft werden müssen, ob sie für eine Verwendung im vorliegenden Rechtsstreit geeignet und welche Überarbeitungen notwendig sind. Würden entsprechende Vorstücke nunmehr von einem LLM bereitgestellt, müssten die Nutzerinnen und Nutzer in Bezug auf besondere Herausforderungen bei der Inanspruchnahme von KI-Anwendungen – etwa dem sog. Automatisierungsbias (automation bias) – geschult werden. Denn es ist bekannt, dass Menschen aufgrund des Automatisierungsbias dazu neigen können, sich auf automatisierte Entscheidungen oder Prozesse zu verlassen und sie als weniger fehleranfällig einzuschätzen als ihre

¹⁵ PeBB§y ist die Kurzbezeichnung für ein System zur Personalbedarfsberechnung für die deutschen Justizbehörden, vgl. hierzu <https://de.wikipedia.org/wiki/PEBB§Y>.

¹⁶ Siehe dazu auch die Stellungnahme der Verfasser zu Einsatzmöglichkeiten und Grenzen für KI in der Justiz: Stellungnahme 18/163 vom 03.01.2023.

¹⁷ Unter Justizassistenzen werden Personen verstanden, die regelmäßig begleitend zum Referendariat an einem Gericht praktisch als wissenschaftliche Mitarbeiterinnen und Mitarbeiter tätig werden, vgl. https://oberlandesgericht-celle.niedersachsen.de/startseite/karriere_ausbildung/richterassistenz/justizassistenz-in-niedersachsen-185530.html.

eigenen Entscheidungen (oder die eines anderen Menschen). Dies könnte insbesondere dazu führen, dass die von einem KI-Programm vorgeschlagene Entscheidung unkritisch übernommen oder (unterbewusst) als besser eingestuft wird. Das Problem stellt sich allerdings nicht spezifisch für LLMs bzw. konkret für ChatGPT, sondern bei allen KI-Anwendungen. Je näher die KI-Unterstützung am Entscheidungsprozess mitwirkt, desto mehr muss dies berücksichtigt werden.

3. In dem Artikel: WISSEN, Freitag, 17. März 2023, Artikel ¼, „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?

Klar ist, dass die Nutzung von KI-Anwendungen wie ChatGPT schon heute und verstärkt in naher Zukunft die Arbeit im Rechtsmarkt und in der Justiz beeinflusst bzw. beeinflussen wird. Hinsichtlich einer Nutzung von LLMs in Kanzleien hatten wir schon das ChatGPT-basierte Programm „Harvey“ der Kanzlei Allen & Overy genannt. Auch rechtssuchende Bürgerinnen und Bürger werden sich entsprechender Tools bedienen. Daher muss die Justiz sich mit dem Phänomen befassen, auch wenn sie ChatGPT nicht einsetzt.¹⁸

In dem in der Frage genannten Artikel wird erwähnt, dass GPT-4 unter dem besten Zehntel der Teilnehmer eines Tests lag. Dies kann insoweit eingeordnet werden, als dass es sich bei dem „juristischen Test“ wohl um ein amerikanisches *bar exam* handelte. Diese Tests bestehen regelmäßig aus Kurzfällen mit *multiple choice*-Antwortmöglichkeiten. Bei einem Großteil der Fragen handelt es sich um bloße Wissensabfragen. Daher ist nicht erstaunlich, dass GPT-4 den Test gut bestanden hat. Es findet bei solchen Tests aber keine umfassende Fallbearbeitung wie im Richteralltag oder den deutschen Staatsexamina statt. LLMs – wie ChatGPT – „verstehen“ die vom Menschen gestellten Fragen oder Arbeitsaufträge (sog. *prompts*) nicht. Vielmehr analysieren die LLMs lediglich Muster und Zusammenhänge in Texten und berechnen daraufhin, für welche Worte die mathematisch größte Wahrscheinlichkeit besteht, in einem

¹⁸ Siehe hierzu bereits Teil A., These 3 dieser Stellungnahme.

zu erstellenden Text von einem Menschen als die „richtigen“ Worte angesehen zu werden. Daher werden LLMs auch als „stochastische Papageien“ bezeichnet, also mathematisch-basierte Textroboter, die zuvor von Menschen verfasste Textbeiträge neu zusammengesetzt „nachplappern“.¹⁹

4. Die Forderungen nach einer „Charta der Robotik“ geht zurück auf einen Bericht des Rechtsausschusses des Europäischen Parlaments mit Empfehlungen an die Kommission zu zivilrechtlichen Regelungen im Bereich Robotik (2015/2103(INL)) vom 27.1.2017 (A8-0005/2017), dort heißt es: „Der für das Gebiet der Robotik vorgeschlagene ethische Verhaltenskodex schafft die Grundlagen für die Ermittlung, Aufsicht und Einhaltung der ethischen Grundsätze bereits von der Planungs- und Entwicklungsphase an.“ Was bedeutet dies für die Entwicklung von KI-Systemen und die Verwendung durch die Justiz und die Verwaltung?

Die Verfasser sehen von einer Beantwortung dieser Frage ab.

5. "KI darf den Menschen nicht ersetzen", Der deutsche Ethikrat äußert sich in seinem Bericht vom 20.3.2023 zum Verhältnis von Mensch und Maschine – in Schule, Medizin, sozialen Medien und Verwaltung. Auf Seite 249 des Berichts weist der Ethikrat auf die Grundrechtsbindung der öffentlichen Verwaltung hin und der daraus folgenden Anforderungen an die Nutzung der KI durch staatliche Einrichtungen, dort heißt es: „Aufgrund ihrer Grundrechtsbindung sind an staatliche Einrichtungen bei der Entwicklung und Nutzung algorithmischer Systeme hohe Anforderungen in Bezug auf Transparenz und Nachvollziehbarkeit zu stellen, um den Schutz vor Diskriminierung zu gewährleisten sowie Begründungspflichten erfüllen zu können.“ Welche rechtlichen Folgen haben diese Aussagen für die Verwendung von ChatGPT oder eines ähnlichen KI-Systems für das Rechtssystem?

und

¹⁹ Die inhaltliche Qualität des „Nachplapperns“ wird durch eine sinnvolle Auswahl der Trainingsdaten und – wie bei ChatGPT im Einsatz – eine Feedback-Funktion verbessert. Beides sind wesentliche Faktoren, warum *customized large language models* innerhalb des zugeschnittenen Bereichs deutlich bessere Resultate erzielen können.

6. Wo sind die ethischen Grenzen des Einsatzes künstlicher Intelligenz bei juristischer Entscheidungsfindung zu finden und durch den Gesetzgeber zu ziehen?

und

7. Wo sind die verfassungsrechtlichen Grenzen des Einsatzes künstlicher Intelligenz bei juristischer Entscheidungsfindung zu finden oder sind diese durch den Gesetzgeber neu zu ziehen?

Das Ministerium der Justiz Nordrhein-Westfalen weist in seinem schriftlichen Bericht vom 22.03.2023 zutreffend darauf hin, dass eine abstrakte rechtliche Bewertung derzeit nur schwerlich möglich ist. Zutreffend ist auch, dass der Einsatz umso aufmerksamer geprüft werden muss, je näher ein KI-System zur Unterstützung von Richterinnen und Richtern an die richterliche Kerntätigkeit heranrückt. Darüber hinaus wird der Einsatz von LLMs in der Justiz sich an den Vorgaben der noch nicht verabschiedeten KI-Verordnung der Europäischen Union messen lassen.²⁰ Sofern keine wesentlichen Änderungen in Bezug auf die Einordnung der unterschiedlichen KI-Anwendungen mehr erfolgen sollten,²¹ würden LLMs wie ChatGPT, die einen Entscheidungsentwurf erstellen, als sog. Hochrisiko-KI eingestuft werden. Denn nach dem Entwurf sind in der Justiz sind alle KI-Systeme als hochriskant anzusehen, wenn sie Justizbehörden bei der Ermittlung und Auslegung von Sachverhalten und Rechtsvorschriften und bei der Anwendung des Rechts auf konkrete Sachverhalte unterstützen sollen. Damit würde nach gegenwärtigem Stand bereits der europäische Rechtsrahmen detaillierte Regulierungsvorgaben machen.

8. Das Urteil des kolumbianischen Richters Padilla vom 30. Januar 2023 umfasste mit gut 29 % der Entscheidungsgründe Textteile von ChatGPT. Damit waren die Aussagen von ChatGPT nicht die einzigen Entscheidungsgründe, jedoch waren sie ein entscheidender Beitrag zur Entscheidungsfindung. (Vgl. ChatGPT in Colombian Courts, Why

²⁰ Vgl. den Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (COM/2021/206 final).

²¹ Das Ministerium der Justiz Nordrhein-Westfalen weist im Bericht vom 15.05.2023 zum TOP „Justiz in der KI-EU-VO“ der 16. Sitzung des Rechtsausschusses zutreffend darauf hin, dass sich die Vorschriften (insgesamt) im Laufe des weiteren europäischen Gesetzgebungsverfahrens aller Voraussicht nach noch ändern werden.

we need to have a conversation about the digital literacy of the judiciary, in: <https://verfassungsblog.de/colombian-chatgpt/>) Wie wäre es rechtlich zu bewerten, wenn ein Urteil zu 50 % und mehr, eventuell 100%, der Entscheidungsgründe aus Textteilen von ChatGPT besteht?

Bei genauer Betrachtung des Urteils des kolumbianischen Richters Juan Manuel Padilla vom 31.01.2023²² lässt sich zunächst festhalten, dass der Richter ChatGPT bezüglich der aufgeworfenen Rechtsfragen „konsultiert“ und dies transparent dargestellt hat. Insbesondere ist für die Parteien erkennbar, ob der Richter die entscheidungserheblichen Fragen gestellt hat. Zudem hat der Richter in dem Urteil ausdrücklich festgehalten, dass er die von der KI gelieferten Informationen validiert und auf dieser Grundlage eine eigenständige Entscheidung getroffen habe. In einer solchen Konstellation wäre es im Grundsatz zweitrangig, zu welchem Prozentsatz ChatGPT Vorschläge für die Abfassung des Urteils gemacht hat.

Gutiérrez weist in dem in der Frage aufgeführten Artikel²³ darauf hin, dass die Validierung der Informationen durch den Richter letztlich nicht verifiziert werden könne. Zwei Aspekte könnten tatsächlich dagegen sprechen, dass dies geschehen ist: Zum einen beschreibt *Gutiérrez*, dass er bei Eingabe der an ChatGPT gestellten, identischen Fragen leicht abweichende Antworten erhalten habe.²⁴ Auch begründet der Richter den Einsatz von ChatGPT mit der damit verbundenen Zeitersparnis. Bei einer gewissenhaften Kontrolle der Antworten von ChatGPT ist gegenwärtig allerdings nicht davon auszugehen, dass ein Urteil durch den Einsatz von ChatGPT signifikant schneller abgefasst werden kann.²⁵

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

und

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

²² Das Urteil ist abrufbar unter <https://forogpp.files.wordpress.com/2023/01/sentencia-tutela-segunda-instancia-rad.-13001410500420220045901.pdf>.

²³ <https://verfassungsblog.de/colombian-chatgpt/>.

²⁴ Siehe hierzu auch die Beantwortung von Frage 19.

²⁵ Daher sehen wir derzeit auch keine nennenswerten Einsatzbereiche im Kernbereich richterlicher Tätigkeit, vgl. Teil A., These 1. Vgl. zudem die Ausführungen zu Fragen 2 und 12.

und

11. Wenn nein, weshalb nicht?

Die Verfasser sind der Ansicht, dass der Einsatz von LLMs grundsätzlich möglich und mit Blick auf eine Effizienzsteigerung in Teilbereichen auch wünschenswert ist. Die größten Potentiale sehen wir gegenwärtig im nichtrichterlichen Bereich und dem der Justizverwaltung. Für richterliche Kernaufgaben dürfte es erforderlich sein, ein *customized large language model* für die Justiz zu trainieren. Wir verweisen hinsichtlich der Einzelheiten auf unsere eingangs aufgestellten und – anhand von Beispielen – erläuterten Thesen (Teil A.).

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

Selbstverständlich sind mit dem Einsatz von KI-Anwendungen und LLMs wie ChatGPT Gefahren und Risiken verbunden.²⁶ Neben dem bereits erläuterten Automatisierungsbias sehen wir etwa die Gefahr, dass es sich bei ChatGPT um ein selbstreferenzielles System handelt. Dies kann etwa dazu führen, dass einmal im System vorhandene Fehler – in Kombination mit dem *automation bias* – so oft bestätigt werden, bis sie (ggf. verdeckt) zur Grundlage weitergehender Informationen gemacht werden.²⁷ Zugleich könnte durch den gemeinsamen Rückgriff auf dasselbe System und dieselben Informationen die Fortentwicklung des Rechts durch die Rechtsprechung gehindert werden. Dieser Gefahren müssen sich die Akteure bewusst sein und fortlaufend hierfür sensibilisiert werden.²⁸

²⁶ Vgl. auch BSI, Große KI-Sprachmodelle, Chancen und Risiken für Industrie und Behörden, abrufbar unter https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Grosse_KI_Sprachmodelle.pdf?__blob=publicationFile&v=2.

²⁷ Es ist nicht einmal erforderlich, dass die Informationen ursprünglich fehlerhaft waren. Vielmehr können bestimmte Informationen z.B. durch eine neue EuGH-Rechtsprechung zu einer bestimmten Rechtsfrage ihre Gültigkeit verlieren.

²⁸ Ein hiermit im Zusammenhang stehender Aspekt ist die ggf. bestehende Möglichkeit des sog. „data poisoning“ von LLMs, vgl. etwa Wan/u.a., Poisoning Language Models During Instruction Tuning, abrufbar unter: <https://arxiv.org/abs/2305.00944>.

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

Vielfach wird im Zusammenhang mit der Intransparenz und fehlenden Nachvollziehbarkeit von KI-Anwendungen auf das sog. *Blackbox*-Problem hingewiesen. Vereinfacht ist von einer *KI-Blackbox* die Rede, wenn für den Anwender der Prozess zwischen der Eingabe der Information (bei ChatGPT des *prompts*) und des Ergebnisses (bei ChatGPT der „Antwort“) nicht „einsehbar“ ist, wie die KI zu dem Ergebnis gelangt ist. In den allermeisten Konstellationen, die wir gegenwärtig für umsetzbar halten, kommt dieses Problem aber nicht zum Tragen. Denn für den Anwender wird es in aller Regel irrelevant sein, wie ChatGPT zu dem Ergebnis gelangt ist. Der Anwender muss das Ergebnis ohnehin auf die Richtigkeit hin überprüfen. In dem oben aufgeführten Beispiel der Abladung eines Zeugen wird die Geschäftsstelle selbstverständlich nur die Teile übernehmen können, die in der konkreten Situation passen und das Schreiben im Übrigen ändern. Dabei muss sie nicht nachvollziehen können, aus welchen (Trainings-)Daten ChatGPT das Schreiben generiert hat.

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

und

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

Hinsichtlich der Beantwortung dieser Fragen verweisen wir auf unsere Thesen in Teil A. der Stellungnahme.

16. Inwieweit wird durch die Anwendung von ChatGPT durch die Richterschaft Grundrechte verletzt?

Sofern die Vorgaben, wie wir sie im Rahmen der Beantwortung der Fragen 1., 2., 5. bis 7. und 8. skizziert haben, beachtet werden, gehen wir nicht von Grundrechtsverletzungen durch die Richterschaft aus. Allerdings ist die Frage rein theoretischer Natur,

da ein Einsatz von ChatGPT aus unserer Sicht schon wegen datenschutzrechtlicher Bedenken in konkreten Einzelfällen nicht in Betracht kommt.

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

Wie ChatGPT die jeweiligen Texte erstellt, ist für die Richterschaft nicht transparent und nicht nachvollziehbar. Wir sehen darin – wie unter Fragen Nr. 13 dargestellt – jedoch nicht das zentrale Problem hinsichtlich der Nutzung von ChatGPT. Wichtiger wäre es Maßnahmen zu ergreifen, um den unter Nr. 12 dargestellten Gefahren entgegenzuwirken. Dies gilt nicht nur für ChatGPT, sondern für alle vergleichbaren LLMs.

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?

Da es sich bei ChatGPT um ein tiefes neuronales Netz handelt, ist es nach unserem Verständnis nicht einmal der Entwicklerfirma OpenAI möglich, hinsichtlich eines erstellten Textes konkret die entsprechenden Quellen zu benennen.

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?

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?

Ob und durch wen Daten, Datensätze und sonstigen Inhalte aktualisiert werden, lässt sich aus öffentlich zugänglichen Quellen nicht ermitteln und ist uns nicht bekannt.

Das Bundesamt für Sicherheit in der Informationstechnik (BSI) weist zutreffend darauf hin, dass LLMs wie ChatGPT unter anderem mit einer fehlenden Sicherheit von generiertem Code, einer Anfälligkeit für „versteckte“ Eingaben mit manipulativer Absicht und einer Verletzung der Vertraulichkeit der eingegebenen Daten verbunden sein können.²⁹

Es ist zutreffend, dass dieselbe Eingabe einer Frage oder eines Textes unterschiedliche „Antworten“ liefert.

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?

Aus der Datenschutzrichtlinie von OpenAI geht hervor, dass die eingegebenen Daten auf Server in den USA gelangen (siehe Privacy Policy, 9. International Users³⁰). Wer Zugriff auf diese Daten hat und was genau mit diesen Daten passiert, ist den Verfassern nicht bekannt. Mit Blick auf Grundrechte und datenschutzrechtliche Vorgaben bestehen daher weitere erhebliche rechtliche Bedenken hinsichtlich des Einsatzes von ChatGPT in der Justiz, die über das zuvor Gesagte hinausgehen.³¹ Die Beurteilung könnte anders bezüglich solcher LLMs ausfallen, die in der Europäischen Union, speziell in Deutschland, betrieben werden.³² Erst recht gilt dies für ein justizeigenes LLM.

Wir freuen uns auf eine Diskussion der aufgeworfenen Punkte im Rahmen der Anhörung.

Dr. Christian Schlicht

Prof. Dr. Simon J. Heetkamp, LL.M.

²⁹ Vgl. BSI, Große KI-Sprachmodelle, Chancen und Risiken für Industrie und Behörden, abrufbar unter https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Grosse_KI_Sprachmodelle.pdf?__blob=publicationFile&v=2.

³⁰ Privacy Policy, Stand 27.04.2023, abrufbar unter: <https://openai.com/policies/privacy-policy>.

³¹ So stellen sich etwa auch komplexe urheberrechtliche Fragen.

³² Zur Frage, ob ein „deutsches ChatGPT“ möglich ist, vgl. etwa die Stellungnahme 18/569 von Dr. Gaida für die Anhörung des Wissenschaftsausschusses des Landtags NRW am 19.06.2023.