

Aufsatz | Justiz | Künstliche Intelligenz

Künstliche Intelligenz und Large Language Models in der Rechtsprechung

24. Juli 2023



Dr. Bettina Mielke
Vorsitzende Richterin, OLG Nürnberg



Prof. Dr. Christian Wolff
Lehrstuhl für Medieninformatik, Universität Regensburg

Zitiervorschlag: *Mielke/Wolff*, LRZ 2023, Rn. 560, [●], www.lrz.legal/2023Rn560.
Permanente Kurz-URL: LRZ.legal/2023Rn560

In jüngster Zeit wird vermehrt über den Einsatz von Large Language Models wie ChatGPT in der Rechtsprechung nachgedacht. Der vorliegende Beitrag erläutert mögliche und bereits praktizierte Einsatzbeispiele dieser und anderer KI-Anwendungen und hinterfragt deren Potenzial in der Justiz.

Inhaltsverzeichnis

1. Einführung
2. Was ist KI?
 - 2.1. Wissensbasierte Systeme
 - 2.2. Maschinelles Lernen, Deep Learning
3. Was ist ChatGPT?
 - 3.1. Entwicklung künstlicher neuronaler Netze seit fast 80 Jahren
 - 3.2. Vom Deep Learning zu den heutigen großen Sprachmodellen (Large Language Models, LLM)
 - 3.3. Wesentliche Merkmale der aktuellen großen Sprachmodelle
4. Künstliche Intelligenz im Rechtswesen
5. Einsatz von ChatGPT im Rechtswesen
 - 5.1. Mögliche Einsatzszenarien
 - 5.2. Gefahren / Herausforderungen beim Einsatz von Sprachmodellen wie ChatGPT
 - 5.3. Nutzen von ChatGPT in den vorgeschlagenen Einsatzszenarien
6. Exkurs: Nutzung großer KI-Modelle bei Anwält:innen
7. Fazit und Ausblick

1. Einführung

Seit längerem wird über den Einsatz von Künstlicher Intelligenz (KI) in der

Rechtsprechung nachgedacht, über die möglichen Anwendungsfälle, über rechtliche Implikationen, bis hin zur Frage, ob und inwiefern dies rechtspolitisch gewollt und ethisch verantwortbar ist. Erste Beispiele zur Unterstützung der Richter:innen sind bereits verwirklicht und tragen Frauennamen wie FRAUKE und OLGA. Die Entwicklung von ChatGPT hat diese Überlegungen forciert. Nachfolgend gehen wir zunächst allgemein auf das Konzept Künstlicher Intelligenz und seine Varianten ein (unter 2.), erläutern ChatGPT als bekannte Anwendung großer Sprachmodelle (unter 3.), bevor wir die Nutzung von KI im Rechtswesen (unter 4.) und die aktuelle Diskussion um die Potentiale von Large Language Models wie ChatGPT in der Justiz thematisieren (unter 5.). Ergänzend folgt ein kurzer Blick auf die anwaltliche Perspektive zu diesem Thema (unter 6.).¹

560

2. Was ist KI?

Der (englische) Begriff *artificial intelligence* (AI) geht auf John McCarthy zurück, der ihn 1955 in einem Förderantrag für eine 1956 stattfindende Konferenz erstmals verwendete und wie folgt umschrieb: „For the present purpose the artificial intelligence problem is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving.“²

561

Auch wenn es keine einheitliche Definition von KI gibt, erfolgt häufig eine Unterscheidung zwischen den sog. logik- und wissensbasierten Expertensystemen, oftmals als „klassische“ oder „good old-fashioned AI“ (GOFAI) bezeichnet, und den Verfahren des maschinellen Lernens einschließlich *deep learning* mit Hilfe neuronaler Netze.³

562

Abzugrenzen ist KI von dem Begriff Legal Tech. Nicht alles, was zu Legal Tech zu zählen ist, ist auch KI. Umgekehrt ist der Einsatz von KI im Rechtswesen als Teilgebiet von Legal Tech anzusehen. Unter Legal Tech fällt beispielsweise die reine Automatisierung ohne den Einsatz von Methoden der KI oder die Nutzung anderer digitaler Möglichkeiten, z.B. zur Strukturierung des Parteivortrags.⁴ Da weder der Begriff Legal Tech noch der Begriff KI einheitlich definiert sind, ist die Auseinandersetzung mit diesen Konzepten mitunter schwierig.

563

2.1. Wissensbasierte Systeme

In den 1960er / 1970er Jahre standen in der KI-Forschung logikbasierte Expertensysteme, die mit explizitem formalisiertem Wissen aufgebaut wurden, im Vordergrund. In solchen Systemen wird u.a. über Regeln menschliches Wissen (Zusammenhänge in der Welt) für Computer verständlich dargestellt (Wissensbasis).

564

Neben der Wissenserwerbskomponente für Aufbau und Verbesserung der Wissensbasis besteht eine Problemlösungs- und eine Erklärungskomponente.⁵ Die Systeme benötigen manuell eingegebenes Wissen sowie Wenn-dann-Regeln, die für eine Anfrage automatisiert genutzt und kombiniert werden. Da es sich als schwierig herausgestellt hat, größere Wissensbasen widerspruchsfrei zu erweitern, hat sich auch die Skalierbarkeit solcher Systeme als enttäuschend erwiesen. Ende der 1980er setzte daher der (zweite) KI-Winter ein.⁶ Beachten sollte man dabei, dass praktisch jedes Computerprogramm über einfache Formen von Wissen verfügt, wenn z.B. Regeln implementiert sind, bei denen auf der Basis einer Werteabfrage eine Entscheidung in die eine oder andere Richtung erfolgt (Steuerung des Kontrollflusses eines Programms). Dies ist nicht als KI zu werten.

2.2. Maschinelles Lernen, Deep Learning

Maschinelles Lernen ist der Oberbegriff für die „künstliche“ Generierung von „Wissen“ aus Erfahrung: Ein System lernt aus Beispielen und kann sie nach Beendigung der Lernphase verallgemeinern. Das heißt, es werden nicht Beispiele auswendig gelernt, sondern das System „erkennt“ Muster und Gesetzmäßigkeiten in den Lerndaten. *Deep Learning* (deutsch: *mehrschichtiges Lernen*, *tiefes Lernen* oder *tiefgehendes Lernen*) bezeichnet eine Methode des maschinellen Lernens, die große künstliche neuronale Netze (KNN) mit zahlreichen Zwischenschichten (englisch *hidden layers*) einsetzt und dadurch eine umfangreiche innere Struktur herausbildet (siehe auch unten).⁷ Gegenüber den schon länger untersuchten Formen neuronaler Netzwerke zeichnen sich die *Deep Learning*-Verfahren durch größere Komplexität und neue Aufbauvarianten

565

aus. Diese Verfahren zeigen vor allem Erfolge bei der Objekterkennung in Bildern sowie im Bereich der automatischen Sprachverarbeitung (maschinelle Übersetzung, Textgenerierung). Letzteres ist für das textlastige Rechtswesen naturgemäß besonders relevant.⁸

Eine Herausforderung ist, dass für ein erfolgreiches Training sehr große Datenmengen erforderlich sind. Ändert sich die Datengrundlage, ist zudem erneutes bzw. weiteres Training notwendig. Probleme stellen die Nachvollziehbarkeit, die gerade für den juristischen Kontext nachteilig ist, sowie Verzerrungen in den Daten (*data bias*⁹) dar, die zu fehlerhaften Ergebnissen führen können. Daneben besteht eine Anfälligkeit für Falschberechnungen: „Künstliche Intelligenzen machen die absurdesten Klassifikationsfehler. Wenn man mit diesen Fehlern weiterrechnet, entsteht noch mehr Unsinn.“¹⁰

566

3. Was ist ChatGPT?

Seit Veröffentlichung des KI-ChatBots ChatGPT durch OpenAI Ende November 2022 hat die gesellschaftliche Debatte zur Leistungsfähigkeit dieses Systems und „intelligenter“ Systeme im Allgemeinen dramatisch an Fahrt aufgenommen. Was zuvor vor allem in Fachkreisen diskutiert wurde – das Sprachmodell GPT 3, auf dem ChatGPT beruht, wurde bereits Mitte 2020 veröffentlicht¹¹ – hat nun einen breiten gesellschaftlichen Diskurs angefangen, der neben den hier betrachteten Fragen des Rechtswesens viele weitere gesellschaftliche Bereiche berührt wie z.B. Fragen der Arbeitsmarktentwicklung¹² oder Fragen des Bildungssystems und des Hochschulwesens¹³. Zahlreiche führende KI-Forscher (u.a. die Turing-Preisträger *Geoffrey Hinton* und *Joshua Bengio*) und IT-Unternehmer (u.a. *Sam Altman*, OpenAI, *Bill Gates*, Microsoft) haben kürzlich einen öffentlichen Appell unterzeichnet, in dem dazu aufgerufen wird, Gefahren durch KI ernst zu nehmen.¹⁴ Dies hat am 12.6.2023 UN-Generalsekretär *Guterres* zum Anlass genommen, auf die Risiken von KI hinzuweisen und eine internationale Regulierungsbehörde nach dem Vorbild der Internationalen Atomenergie-Organisation (IAEO) zu fordern.¹⁵

567

3.1. Entwicklung künstlicher neuronaler Netze seit fast 80 Jahren

Künstliche neuronale Netze entstehen bereits seit knapp achtzig Jahren, sie sind damit ähnlich alt wie die moderne Informationstechnik. Auch die jüngsten Entwicklungen (GPT, ChatGPT) sind als eher evolutionär zu werten, da sie auf Konzepten zur strukturellen Weiterentwicklung künstlicher neuronaler Netze aus den 1990er Jahren sowie auf verbesserten Rechenmöglichkeiten der Hardware (GPUs, Parallelisierung) aufbauen: Bekannte Weiterentwicklungen der Architekturen künstlicher neuronaler Netze umfassen *Convolutional Neural Networks* (CNN, Faltungsnetzwerke), *Recurrent Neural Networks* (RNN, rekurrente neuronale Netze) und *Long Short-Term Memory*-Netze (LSTM, Netze mit lokalem Speicher bzw. „Gedächtnis“, um z.B. sprachliche Abhängigkeiten über große Distanzen verarbeiten zu können)¹⁶.

568

3.2. Vom *Deep Learning* zu den heutigen großen Sprachmodellen (*Large Language Models*, LLM)

Aufbauend auf der Entstehung der *Deep Learning*-Verfahren, die in den 1990er Jahren begann, setzt seit etwa zehn Jahren die Entwicklung der aktuellen großen Sprachmodelle ein. Diese erfolgte grob in den folgenden Schritten:

569

- Mit dem Konzept der **Worteinbettungen** (*word embeddings*) ist es gelungen, den inhaltlichen bzw. semantischen Kontext von Wörtern in ihrem konkreten Textumfeld besser zu erfassen und für das maschinelle Lernen einzusetzen.¹⁷
- Die **Transformer-Architektur**, die konzeptuell ein Zusammenspiel zwischen einer Kodierungs- und einer Dekodierungseinheit vorsieht,¹⁸ hat sich als besonders geeignet für die Verarbeitung sprachlicher Daten herausgestellt.
- Mit der Idee eines „**Aufmerksamkeits-Mechanismus**“ („*attention is all you need*“)¹⁹ konnte zusammen mit der Transformer-Architektur das zugrundeliegende Netzwerkkonzept vereinfacht und damit die Skalierbarkeit deutlich verbessert werden.
- Erst auf dieser Basis wurden Modelle möglich, die **mit praktisch jedem verfügbaren Text vortrainiert** (*pretrained*) werden konnten; letztlich ein äußerst

leistungsfähiges Verfahren des nicht-überwachten maschinellen Lernens (*unsupervised machine learning*), das für das Vortraining im Verhältnis zur verarbeiteten Datenmenge vergleichsweise wenig menschliches Wissen und menschliche Intervention benötigt. Durch die extrem große Menge an vortrainiertem Material können derartige Sprachmodelle sinnvolle Antworten auf eine breite Vielfalt von Fragestellungen liefern, ohne gezielt nachtrainiert zu werden („*zero-shot learning*“).²⁰

- Weitere Trainingsläufe kalibrieren das Modell nach der Methode des **bestärkenden Lernens** (*reinforcement learning*²¹). Dabei erfolgen anfangs Bewertungen durch Menschen, während in späteren Phasen das System sich durch ein Belohnungsmodell selbst optimiert.
- Da die verwendete Menge an Trainingstext deutlich schneller wächst als neuer von Menschen geschriebener Text hinzukommt, wird bereits eine Datenkrise für die kommenden Jahre vorhergesagt.²² lassen sich die Modelle allerdings mit selbsterzeugtem Text weitertrainieren („Textinzest“²³).

Auch wenn sich die aktuelle Diskussion auf ChatGPT mit seiner leicht zugänglichen Dialogfunktion (Chatbot) konzentriert, ist darauf hinzuweisen, dass in den letzten Jahren eine große Vielfalt vergleichbarer Modelle mit unterschiedlichen Schwerpunkten und Eigenschaften entwickelt wurde. Praktisch alle großen Tech-Konzerne beteiligen sich daran, es gibt zudem mittelständische Unternehmen, u.a. in Deutschland, die auf diesem Gebiet erfolgreich arbeiten, z.B. der Übersetzungsdienst **DeepL** ([deepl.com/de/translator](https://www.deepl.com/de/translator)). Zudem liegen Modelle vor, die offen verfügbar sind und nach dem *open source*-Prinzip entwickelt werden.²⁴

570

3.3. Wesentliche Merkmale der aktuellen großen Sprachmodelle

LLMs beruhen auf **statistischen Verfahren**; die Trainingsergebnisse lassen sich nicht ohne weiteres erklären; es erfolgt **keine explizite Wissensrepräsentation**.²⁵ Die bekannteste Modellserie GPT (*generative pretrained transformer* – vortrainiertes generatives Transformermodell) wird von der US-Firma OpenAI entwickelt, an der mittlerweile Microsoft maßgeblich beteiligt ist.²⁶ Die derzeit größten Trainingsläufe erfolgten mit substantiellen Teilen des World Wide Web und seinen Inhalten (z.B. Wikipedia) sowie mit den verfügbaren Sammlungen elektronischer Texte.²⁷ Für die neuesten Varianten (GPT-4 mit einer erneuten Größensteigerung auf mehrere Billionen Knoten im Modell) wurden Informationen zum Training allerdings nicht mehr offengelegt.²⁸

571

Die **Trainingszyklen der aktuellen großen Modelle sind lang (bis zu sechs Monate) und kostspielig**, die Kosten für einzelne Trainingsläufe können mehrere Millionen Euro betragen; der Betrieb des Systems soll pro Tag mehrere 100.000 Euro kosten.²⁹ Bei der Vorbereitung der Modelle auf den Dialog mit Menschen und den dabei auftretenden Fragestellungen bzw. bei der Nachbearbeitung der Trainingsergebnisse kommt in erheblichem Umfang menschliches Wissen zum Einsatz.³⁰ Verfahren, die aktuellen sehr großen Transformer-Modelle mit kleinen Datenbeständen nachzutrainieren, stehen noch am Anfang.³¹ Sie bieten aber interessante Perspektiven, allgemein trainierte Modelle auf bestimmte Wissensfelder oder Dokumenttypen anzupassen.

572

Die Modelle haben kein explizites Wissensmodell; **sie wissen nichts** und **sie verstehen nichts**. Sie sind aber in der Lage, plausible und sprachlich korrekte Texte zu produzieren, weil es ihnen gelingt, für zunehmend lange Texte die wahrscheinlichsten oder besten Nachfolger eines Wortes vorherzusagen bzw. auszuwählen. Es treten vielfältige Arten von Fehlern auf: Die Modelle **„halluzinieren“**, d.h. sie erfinden Fakten wie z.B. nicht existente wissenschaftliche Quellen oder sie reproduzieren Texte, mit denen sie trainiert wurden (oder auch die Eingaben (Prompts)), in nahezu identischer Form („*regurgitation*“ – „Auskotzen“). Beides wirft datenschutz- und urheberrechtliche Fragen auf. Aktuelle Studien legen nahe, dass komplexe Aufgaben (z.B. das Multiplizieren von Zahlen) nur dann von großen Sprachmodellen wie ChatGPT gelöst werden können, wenn die richtigen Antworten Teil des Trainingsmaterials gewesen sind. Mit anderen Worten: Eine Möglichkeit zum eigenständigen „Nachdenken“ (*reasoning*) haben die Modelle nicht. Dies beschränkt ihre Möglichkeiten grundsätzlich.³²

573

Weitere Beispiele für die Anwendung aktueller großer Sprachmodelle können Systeme sein, die komplexe Arbeitsprozesse automatisieren (Workflow Automatisierung). Neben

574

der automatisierten Generierung von Texten könnte für unterschiedliche Arten von Tätigkeiten Unterstützung bei eher repetitiven Arbeitsformen entstehen.

4. Künstliche Intelligenz im Rechtswesen

Nachdem im juristischen Kontext – entsprechend der allgemeinen Entwicklung zur künstlichen Intelligenz – zunächst logik- und wissensbasierte Expertensysteme im Fokus der Aufmerksamkeit standen (vgl. etwa das Projekt LEX aus den 1980er Jahren³³), die an den oben genannten Problemen scheiterten, setzt man nunmehr vor allem auf den Einsatz von Methoden des maschinellen Lernens. 575

Die Analyse großer Datenmengen ist dabei ein wichtiges praktisches Einsatzfeld, etwa in Wirtschaftsstrafsachen, wenn große Datenbestände beschlagnahmt werden und eine computerunterstützte Auswertung erfolgen soll. Solche Systeme sind bereits im Einsatz, z.B. bei der Zentralen Staatsanwaltschaft zur Verfolgung von Wirtschaftsstrafsachen und Korruption in Wien. Sie können u.a. verschiedene Dokumenttypen, wie Rechnungen, Protokolle von Vorstandssitzungen etc. erkennen und ausfiltern oder verschiedene Schreibweisen des Datums, der Kontonummer oder des Firmennamens vereinheitlichen, nachdem sie vorher mit Methoden des maschinellen Lernens darauf trainiert wurden.³⁴ 576

Maschinelle Lernverfahren kommen auch bei den Programmen OLGA³⁵ und FRAUKE³⁶ zum Einsatz. Mit OLGA werden beim OLG Stuttgart die angefochtenen erstinstanzlichen Urteile sowie Berufungsbegründungen und -erwiderungen der anhängigen Dieselverfahren im Hinblick auf die Parameter analysiert, nach denen die entscheidenden Richter Fallgruppen bilden (z.B. Motortyp, Abgasnorm, Rückrufbetroffenheit). So können die Verfahren sortiert werden, um gleich gelagerte Fälle gemeinsam bearbeiten zu können.³⁷ FRAUKE, das am Amtsgericht Frankfurt a.M. zur Bewältigung der Verfahren zu den Fluggastrechten entwickelt wurde, extrahiert aus den Schriftsätzen automatisch relevante Falldaten (Start- und Zielflughafen, Flugentfernung), um lästige *copy-* und *paste-*Arbeiten zu vermeiden. Die Analyse erfolgt dabei aufgrund eines vortrainierten Basismodells und einer weiteren Anpassung an den speziellen Sprachgebrauch in diesen Verfahren. Bei beiden Systemen kommen Methoden des maschinellen Trainings zur Extraktion von Entitäten mittels der Watson-Technologie der Firma IBM zum Einsatz, also Verfahren, die zur KI gezählt werden können.³⁸ FRAUKE schlägt zudem regelbasiert passende Urteilsbausteine vor, nachdem die Richter ihre Entscheidung aufgrund des vorliegenden Sachverhalts getroffen haben (z.B. Klageabweisung aufgrund schlechter Wetterbedingungen). Nach der oben genannten Unterscheidung handelt es sich insofern aber nicht um KI, sondern um eine einfache regel- bzw. algorithmenbasierte Anwendung. 577

Das Grundlagenpapier der Präsidentinnen und Präsidenten der Oberlandesgerichte, des Kammergerichts, des Bayerischen Obersten Landesgerichts und des Bundesgerichtshofs „Einsatz von KI und algorithmischen Systemen in der Justiz“ von 2022 listet im Anhang neben OLGA und FRAUKE alle Vorhaben der Justiz auf diesem Gebiet auf, etwa auch ein Pilotprojekt am LG Ingolstadt, das ebenfalls eine bessere Aufbereitung von Massenverfahren zum Ziel hat.³⁹ Aufgrund der Abgrenzungsschwierigkeiten zwischen „einfachen“ algorithmischen Systemen und KI-Systemen hat die Arbeitsgruppe bewusst entschieden, sich nicht auf KI-Anwendungen zu beschränken, sondern alle algorithmischen Systeme zu erfassen, die der Richterassistenz dienen.⁴⁰ 578

Ein an der Universität zu Köln unter der Leitung der Strafrechtlerin Frauke Rostalski durchgeführtes Projekt will mit Methoden des maschinellen Lernens strafzumessungsrelevante Informationen aus Urteilen ermitteln. Das Ziel ist, eine Datenbank mit den extrahierten Strafzumessungserwägungen und der jeweiligen Strafhöhe aufzubauen, um dadurch zur Transparenz hinsichtlich der bundesweit von den verschiedenen Gerichten verhängten Strafen beizutragen.⁴¹ Hintergrund ist ein seit Jahren beobachteter regionaler Unterschied in den Strafen bei vergleichbaren Fällen.⁴² 579

Ein weiteres viel diskutiertes Einsatzgebiet von KI im juristischen Kontext ist der Bereich der *Predictive Analytics*. Dabei werden historische Daten herangezogen, um zukünftige Ereignisse vorherzusagen. Denkbar erscheinen Systeme zur Entscheidungsvorhersage oder zur Risikobewertung, z.B. Informationssysteme, die Fälle nach bestimmten 580

Eigenschaften, der Rechtsprechung eines bestimmten Spruchkörpers oder zur Kriminalitätsprognose, etwa der Betrugswahrscheinlichkeit von Versicherungsfällen, auswerten. Ein Beispiel, das zur Vorhersage des Rückfallrisikos von Straftätern in den USA dient, ist das umstrittene Programm COMPAS.⁴³ Soweit Algorithmen dazu dienen sollen, Vorhersagen dazu zu machen, mit welcher Wahrscheinlichkeit eine bestimmte Gerichtsentscheidung zu erwarten ist, wird damit lediglich das Ergebnis einer richterlichen Entscheidung vorhergesagt, ohne die Argumentation in irgendeiner Form nachzuzeichnen. Solche Entscheidungsvorhersagesysteme stecken noch in den Kinderschuhen.⁴⁴ Ob sie jemals in der Breite möglich sind, ist fraglich – nicht nur aufgrund des Datenmangels im juristischen Kontext, da vergleichsweise wenige und fast nur obergerichtliche Entscheidungen veröffentlicht werden.⁴⁵ Um hier eine größere Datenbasis zu erhalten, wären Anonymisierungstools notwendig, die ebenfalls Gegenstand der Forschung zum Einsatz von KI im juristischen Kontext sind.⁴⁶

Relativ klar ist nach derzeitiger Lage, dass ein Richterautomat, Entscheidungsroboter oder Robo-Richter, der eigenständig Gesetze auslegt, Sachverhalte subsumiert und Entscheidungen wertender Art trifft, technisch in weiter Ferne liegt, auch wenn er Gegenstand zahlreicher wissenschaftlicher Beiträge ist.⁴⁷ So hat sich beispielsweise ein von den wissenschaftlichen Diensten des Deutschen Bundestages⁴⁸ aufgelistetes angebliches Projekt in Estland, wonach bei Streitwerten von bis zu 7.000 Euro ohne menschliche Mitwirkung eine künstliche Intelligenz die Entscheidung trafe, als Falschmeldung herausgestellt: Auf Nachfrage teilte das estnische Justizministerium mit, dass es lediglich darum geht, ob Informations- und Kommunikationstechnologien zur Unterstützung der richterlichen Tätigkeit nutzbar gemacht werden können, dass die Entscheidung im Streitfall aber weiterhin ein Mensch trifft.⁴⁹

581

5. Einsatz von ChatGPT im Rechtswesen

Seit der Veröffentlichung von ChatGPT durch OpenAI Ende 2022 stellen viele sich die Frage, ob sich durch die neuen Möglichkeiten von Sprachmodellen wie ChatGPT nunmehr eine Entwicklung zumindest in technischer Hinsicht in Richtung Robo-Richter abzeichnet. Dazu gibt es eine breite Diskussion. So hatte der Rechtsausschuss des Landtags Nordrhein-Westfalen am 22.3.2023 den Einsatz von ChatGPT im Justizbereich auf der Tagesordnung. Am 13.6.2023 erfolgte dazu eine Anhörung von Sachverständigen. Der entsprechende Bericht zur Sitzung des Rechtsausschusses sowie die vorbereitenden Stellungnahmen der Sachverständigen sind veröffentlicht.⁵⁰

582

Übereinstimmend gehen die Stellungnahmen zunächst davon aus, dass ein den Richter ersetzender Einsatz von ChatGPT schon aus verfassungsrechtlichen Gründen nicht möglich ist.⁵¹ Lediglich als Assistenzsystem kämen LLMs wie ChatGPT bei gleichzeitiger Sicherstellung der richterlichen Unabhängigkeit in Frage.⁵² Die Stellungnahme des EDV-Gerichtstags hält es dabei sogar für unzulässig, „dass ein Richter oder eine Richterin ChatGPT eine für die Entscheidung eines Rechtsstreits relevante **Rechtsfrage** stellt und die Antwort dann in eine Entscheidung einfließen lässt.“ Die Berücksichtigung einer solchen Antwort berge, „auch wenn sie kritisch hinterfragt wird, die Gefahr einer unerwünschten Beeinflussung des Richters“. Im Bericht des Ministeriums der Justiz des Landes Nordrhein-Westfalen vom 20.3.2023 an den Präsidenten des Landtags zum Tagesordnungspunkt „Einsatz von ChatGPT im Justizbereich“ heißt es ebenfalls unter Hinweis auf Intransparenz und mangelnde Nachvollziehbarkeit solcher Systeme: „Vor diesem Hintergrund lehnt das Ministerium der Justiz die Anwendung von ChatGPT im Justizdienst insgesamt ab“.⁵⁴

583

5.1. Mögliche Einsatzszenarien

Als mögliche Einsatzszenarien werden von den Sachverständigen, die im Rechtsausschuss des Landtags Nordrhein-Westfalen angehört wurden, verschiedene Anwendungsgebiete genannt. Ein wichtiger Aspekt bezieht sich auf die große Stärke solcher Modelle, nämlich die Textgenerierung. Einsatzmöglichkeiten werden im Verfassen einfacher Schreiben gesehen wie Ladungen, Fristverlängerungen, Anforderungen von Akten, Streitwertfestsetzungen⁵⁵ oder Hilfe bei der Umsetzung richterlicher Verfügungen in Texte durch die Geschäftsstellen⁵⁶. Die Nutzung von ChatGPT für komplexere Dokumente, also Beschluss- und Urteilsentwürfe, wird aufgrund der Unzuverlässigkeit von Modellen wie ChatGPT (zumindest derzeit) abgelehnt.⁵⁷ Betont wird zudem, dass ChatGPT keine Suchmaschine ist und deshalb als

584

Recherchewerkzeug ausscheidet.⁵⁸

Denkbar erscheint den Sachverständigen, ChatGPT als Assistenz einzusetzen, etwa zur sprachlichen Glättung von Schreiben, die durch eine Zusammenstellung von Textbausteinen entstanden sind,⁵⁹ oder zum Verfassen von Texten aus Daten in strukturierter Form, beispielsweise im Bereich des Strafrechts bei Ladendiebstählen, geringfügiger Betäubungsmittel-Kriminalität oder einfachen Verkehrsdelikten⁶⁰. 585

Ein sinnvoller Anwendungsbereich wird weiterhin in der Erschließung von Daten/Akten und Beweismitteln gesehen.⁶¹ Konkret wird an die Visualisierung von Sachverhalten⁶², die Extrahierung von Metadaten⁶³, die automatische Klassifizierung nach Dokumenttypen und automatisiert geordnete Akten⁶⁴, die tabellarische Aufarbeitung von Dokumentinhalten⁶⁵, das Erkennen von Zusammenhängen und das Erstellen von Textvergleichen⁶⁶ gedacht. Vorstellbar erscheint auch die Hilfestellung bei der Berechnung von Prozesskostenhilfe oder des Schmerzensgelds, ebenso ein automatisierter Fristenkalendar.⁶⁷ Erwähnt wird immer wieder die Nutzung zur Anonymisierung⁶⁸ von Texten, die z.T. dem Bereich der einfachen Assistenzaufgaben zugeordnet wird.⁶⁹ Im Einzelnen wird allerdings nicht deutlich, ob diese funktionalen Zuschreibungen von ChatGPT auch tatsächlich erfüllt werden können. 586

Ein weiterer Aspekt betrifft die Verbesserung der Barrierefreiheit bei der Justiz, etwa durch eine automatisierte Beschreibung von Bildern, die Generierung von Texten in einfacher Sprache oder die Übertragung des gesprochenen Worts in Gebärdensprache.⁷⁰ Auch könnte ChatGPT als Grundlage intelligenter Chatbots⁷¹ dienen. Projekte zur Entwicklung von Chatbots (ohne Nutzung von ChatGPT) bestehen bereits, z.B. zur Unterstützung der Rechtsantragsstelle.⁷² 587

Ein interessanter Ansatz ist die geäußerte Vorstellung, Sprachmodelle zur Verringerung menschlicher Rationalitätsschwächen, wie kognitiver Verzerrungen und Denkfehler, einzusetzen oder auch als Risikoprognosewerkzeug⁷³. Möglich erscheint hier eine Unterstützung, indem ähnliche Urteile ausgegeben werden oder – ähnlich der Vorstellung des Projekts zur Erstellung einer Datenbank zur Strafzumessung – Vergleichsfälle zum Strafmaß. Dadurch könnten die Richter ihre juristische Argumentation mit denen der Vergleichsfälle abgleichen und so ggf. auf neue Argumente kommen oder die Entscheidung noch einmal überdenken.⁷⁴ 588

5.2. Gefahren / Herausforderungen beim Einsatz von Sprachmodellen wie ChatGPT

Zunächst wird von den Sachverständigen betont, dass datenschutzrechtliche Bedenken⁷⁵ einem direkten Einsatz von ChatGPT, etwa durch die Eingabe konkreter Verfahrensdaten, entgegenstehen. Zwar sei es zumindest bei der kostenpflichtigen Nutzung von ChatGPT möglich, die Verwendung der Eingaben für das weitere Training zu untersagen; es würde aber dennoch eine Datenverarbeitung von personenbezogenen Justizdaten durch OpenAI erfolgen, was nicht zulässig sei.⁷⁶ Die Nutzung für die hoheitliche Entscheidungsfindung sei zudem nach den Nutzungsbedingungen von ChatGPT in sensiblen Bereichen, z.B. im Strafrecht, sowie generell für die Rechtsberatung untersagt.⁷⁷ 589

Als Gefahren werden neben der Unzuverlässigkeit hinsichtlich der inhaltlichen Richtigkeit die mangelnde Nachvollziehbarkeit sowie die Intransparenz identifiziert,⁷⁸ wobei dies teilweise als unproblematisch angesehen wird, da ohnehin eine Überprüfung auf Richtigkeit erfolgen müsse.⁷⁹ 590

Auch die Abhängigkeit der Qualität und Nutzbarkeit eines generierten Textes von den Systemeingaben erscheint als Herausforderung. Da die Fähigkeit zur Erstellung guter Systemeingaben, der „Prompts“, zeitaufwendig ist und Erfahrung erfordert, sei es Richter:innen und Rechtspfleger:innen derzeit nicht zuzumuten, „Experten in der Prompt-basierten Bedienung von generativen Sprachmodellen zu werden, um die Vorteile dieser Technologie nutzen zu können“.⁸⁰ Sinnvoller sei es, dass für gängige Aufgaben eine „auf die Rechtspraxis zugeschnittene Software“ die Interaktion mit dem Sprachmodell übernehme, wie etwas das Produkt *CoCounsel* der amerikanischen Firma Casetext.⁸¹ 591

Von allen Sachverständigen wird der *Automation Bias* als Problempunkt

herauskristallisiert, also die Befürchtung, dass die Anwender zu großes Vertrauen in die Ergebnisse des Systems haben und falsche Informationen als richtig akzeptieren könnten.⁸² Zur Eindämmung dieses Phänomens wird vorgeschlagen, die generierten Ausgaben ausdrücklich als „Assistenz“, „Unterstützung“ oder „unterstützende Information“ zu kennzeichnen; hilfreich könne auch ein spezielles Training mit dem System sein, das (absichtlich) eingebaute Fehler enthält.⁸³

592

Die Hoffnung zur Vermeidung der Gefahren durch die Nutzung von ChatGPT liegt auf dem Aufbau eines Justiz-LLM.⁸⁴ Dadurch würden die datenschutzrechtlichen Probleme entschärft⁸⁵, zudem hätte man die Kontrolle über die Trainingsdaten, so dass Urheberrechtsverletzungen ebenso wie diskriminierende Inhalte in den Daten verhindert werden könnten.⁸⁶ Da bislang ChatGPT nicht auf die Anwendung des Rechts hin optimiert sei⁸⁷, könnte durch ein Finetuning auf das deutsche Recht⁸⁸ ein deutlich leistungsfähigeres *customized large language model*⁸⁹ entstehen. Zum einen könnte dies durch eine Beteiligung der Fachdatenbanken wie beck-online oder juris erfolgen oder durch den Aufbau eines justizeigenen LLM, vgl. auch das auf dem 2. Digitalgipfel am 25.5.2023 beschlossene Forschungsprojekt von Bayern und Nordrhein-Westfalen („Generatives Sprachmodell der Justiz“).⁹⁰

593

5.3. Nutzen von ChatGPT in den vorgeschlagenen Einsatzszenarien

Hinsichtlich des möglichen Einsatzes von ChatGPT in den verschiedenen Szenarien stellt sich die Frage, wie hoch der Nutzen einzuschätzen ist. So wird von den Sachverständigen des Landtags Nordrhein-Westfalen die Frage aufgeworfen, inwieweit der Einsatz bei der Textgenerierung Vorteile gegenüber der heute üblichen Verwendung von Mustertexten oder Textbausteinen habe.⁹¹ Dies erscheint zumindest für Standardtexte in der Tat fraglich. Lediglich bei der Umsetzung von individuellen richterlichen Verfügungen durch die Geschäftsstellen oder bei der Textverbesserung ist hier ein Vorteil zu erwarten.

594

Auch bei der Überführung strukturierter Daten in Text, etwa zur automatisierten Erstellung von Strafbefehlsanträgen durch die Staatsanwaltschaft, stellt sich die Frage nach dem Vorteil gegenüber einer regelbasierten Überführung der strukturierten Daten, wie Tatzeit, Tatort, Wert des gestohlenen Diebesgutes etc., die sich ebenfalls relativ einfach realisieren ließe. So bedarf es auch nach *Biallaß* einer genaueren wissenschaftlichen und praktischen Erprobung, in welchen Bereichen der Einsatz von LLM und in welchen Bereichen der Einsatz von regelbasierten Systemen und intelligenten Abfragemasken sinnvoller ist.⁹²

595

Beim Einsatz von Sprachmodellen wie ChatGPT ist generell zu fragen, wie groß der Nutzen ist, wenn aufgrund der bekannten und modellbedingten inhaltlichen Unzuverlässigkeit jeweils eine genaue Überprüfung durch die handelnden Akteure notwendig ist. So führt *Nink* im Hinblick auf die Frage, ob im Kernbereich der richterlichen Tätigkeit ChatGPT von Nutzen ist, aus: „Inbesondere mit Blick auf eine mögliche Zeitersparnis und das Erfordernis der Nachprüfung der Quellen, aber auch der Antworten insgesamt, dürfte dies im Ergebnis zu verneinen sein.“⁹³

596

In diesem Zusammenhang zitiert Grabmair den Informatiker Richard Socher, dass geeignete Nutzungsszenarien dann gegeben seien, wenn „es eine lange Zeit bräuhete um ein Artefakt zu erschaffen, jedoch sehr wenig Zeit um seine Korrektheit zu verifizieren“.⁹⁴

597

6. Exkurs: Nutzung großer KI-Modelle bei Anwäl:innen

Auch aus der Perspektive der Anwäl:innen ist der Einsatz großer Sprachmodelle schnell aufgegriffen worden: Erste Legal Tech-Startups wie Justin-Legal⁹⁵ nutzen ChatGPT im Rahmen einer KI-basierten „Mandanten-Anamnese“, bei der potenzielle Klägerinnen oder Kläger auf einer browserbasierten Anwendung im Web strukturierte Angaben zu ihrem Fall machen können (insoweit dem elektronischen Basisdokument⁹⁶ nicht unähnlich). Die Firma JUPUS GmbH bietet einen Anwalts-Chatbot auf der Basis von GPT-4 an, mit dem die Mandantenakquise unterstützt werden kann.⁹⁷

598

Für die aktive Nutzung von KI-ChatBots durch Anwäl:innen hat der Anwalt und KI-Experte *Tom Braegelmann* Empfehlungen publiziert⁹⁸ und ist zu diesem Thema als Dozent etwa für Online-Schulungen aktiv.⁹⁹ Auch in der wissenschaftliche Literatur

599

werden entsprechende Empfehlungen für ein erfolgreiches *prompt engineering* durch Anwält:innen entwickelt.¹⁰⁰ Dabei werden sehr konkrete Formulierungsempfehlungen für den Dialog mit Systemen wie ChatGPT gegeben, insbesondere zur sprachlichen und inhaltlichen Gestaltung der Prompts bzw. zu den unterschiedlichen denkbaren Aufgaben, Fragen und Recherchetypen.

Die *New Technologies Commission* der *European Bar Association* hat im Juni 2023 Richtlinien und Empfehlungen zum Einsatz von KI-Werkzeugen durch Anwält:innen veröffentlicht.¹⁰¹ Die insgesamt sieben Empfehlungen sind recht allgemein gehalten: Sie empfehlen die Erarbeitung eines soliden Verständnisses für generative KI, das Anerkennen der Grenzen und des Kontextbezugs der Technologie, die Befolgung der geltenden Regeln für den Umgang mit KI, die Nutzung als Ergänzung eigener Rechtsexpertise, ohne sich zu sehr darauf zu verlassen, den Schutz des Verhältnisses zwischen Anwalt und Mandant, die Sicherung von Datenschutz und Privatheit sowie die Forderung nach Transparenz und Verantwortlichkeit z.B. in der Kommunikation mit Mandanten und der Aufklärung über die Nutzung von KI-Systemen.

600

7. Fazit und Ausblick

Der Begriff KI wird nach wie vor sehr heterogen verwendet, und es findet nicht selten und gleichzeitig unzutreffend eine Gleichstellung von Automatisierung mit KI statt. Das erschwert die Kommunikation und führt zu Fehlvorstellungen. Auch angesichts der Möglichkeiten großer Sprachmodelle wie ChatGPT ist man von einer KI, die Richter:innen ersetzen könnte, technisch wie rechtlich weit entfernt. Es ist eine Utopie, zu meinen, dass KI-Entwicklungen schnell substanzielle Hilfe bei der richterlichen Arbeit ermöglichen könnten. Dies gelang bislang nur in kleineren Pilotprojekten mit eingeschränktem Anwendungsbereich und einer sehr hohen Anzahl vergleichsweise einfacher und ähnlich strukturierter Fälle (Dieselverfahren, Fluggastrechte).

601

Der Einsatz von KI-Systemen hat dennoch erhebliches Potential zur Optimierung der Arbeit im Rechtswesen im Allgemeinen und der Justiz im Besonderen; allerdings ist die Entwicklung aufwändig. In diesem Zusammenhang können auch große Sprachmodelle mit ihren vielfältigen, sich sukzessive entwickelnden Anwendungsmöglichkeiten Verwendung finden. Neben der Generierung von Texten werden weitere Funktionalitäten, etwa die Befragung des Aktenbestands eines Prozesses, dann verfügbar werden, wenn es gelingt, solche Modelle entsprechend nachzutrainieren. Auch die Generierung von Anschreiben oder die Erstellung von Zusammenfassungen größerer Textbestände erscheinen machbar.¹⁰²

602

¹ Teile des Textes in Kap. 1 und 3 greifen einen Beitrag der Autorin *Mielke* im Mitteilungsblatt des Bayerischen Richtervereins e.V. auf (*Mielke*, Künstliche Intelligenz in der Justiz, BRV-Nachrichten, 1/2023, S. 4 ff.). Die Ausführungen zu großen Sprachmodellen und ChatGPT basieren auf einer Stellungnahme des Autors *Wolff* anlässlich einer Anhörung im Wissenschaftsausschuss des Bayerischen Landtags (*Wolff*, Stellungnahme zur Anhörung „Chancen und Risiken von KI im Wissenschaftsbetrieb“ des Ausschusses für Wissenschaft und Kunst des Bayerischen Landtags, 18. Wahlperiode, 88. Sitzung, 23.6.2023, Universität Regensburg (unveröffentlicht).

² Vgl. *Mitchell*, Artificial Intelligence. A Guide for Thinking Humans, 2019, S. 3 f.; formal.stanford.edu/jmc/history/dartmouth/dartmouth.html, zuletzt aufgerufen am 2.7.2023.

³ *Mitchell* (Fn. 2), Kap. 1-3, S. 3 ff.

⁴ *Mielke/Wolff*, Der strukturierte Parteivortrag im Zivilprozess, Jusletter IT 31.5.2022, online: doi.org/10.38023/14550731-9195-4c76-a1f7-ba57b5f4914b.

⁵ *Pickover*, Künstliche Intelligenz. Eine illustrierte Geschichte. Von mittelalterlichen Robotern zu neuronalen Netzen, 2021, S. 111; de.wikipedia.org/wiki/Expertensystem, zuletzt aufgerufen am 2.7.2023.

⁶ *Fraunhofer-Allianz Big Data* (Hrsg.), Künstliche Intelligenz in Deutschland. Ein systematischer Katalog von Anwendungen des maschinellen Lernens, S. 6, Sankt Augustin 2018.

⁷ Vgl. den Wikipedia-Eintrag zu Deep Learning, de.wikipedia.org/wiki/Deep_Learning, zuletzt aufgerufen am 3.7.2023; *Kelleher*, Deep Learning, 2019.

⁸ Zum Einsatz von Texttechnologie allgemein siehe *Mielke/Wolff*, Möglichkeiten und Perspektiven der Korpuslinguistik für die Analyse von Rechtstexten, *Der Wirtschaftsführer. Sprache und Recht. Auf dem Weg zur Digitalisierung*, 2022, S. 13 ff.

⁹ *Kelleher/Tierney*, 2018, *Data Science*, S. 46 ff.

¹⁰ *Spiekermann*, Die große Big-Data-Illusion, *FAZ*, 25.4.2018. Vgl. auch *Rodemann*, Tango auf dem

Zebrastreifen, SZ, 18.6.2020: „Als künstliche neuronale Netze immer öfter eingesetzt wurden, um Bilder zu klassifizieren, begannen Informatiker, erstaunliche Lücken zu entdecken: Die Netze hielten Golden Retriever für Flugzeuge, Bananen für Toaster oder eben Stoppschilder für Tempolimits – nur weil einige wenige Details verändert waren.“

¹¹ *Brown et al.*, Language Models are Few-Shot Learners. arXiv preprint. 2020, doi.org/10.48550/arXiv.2005.14165; *Floridi/Chiriacchi*, GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 30(4), 2020, 681 ff., online: doi.org/10.1007/s11023-020-09548-1.

¹² *Zarifhonarvar*, Economics of ChatGPT: A Labor Market View on the Occupational Impact of Artificial Intelligence. *Social Sciences Research Network (SSRN)*, 2023, online: doi.org/10.2139/ssrn.4350925.

¹³ Vgl. *Salden/Leschke* (Hrsg.), *Didaktische und rechtliche Perspektiven auf KI-gestütztes Schreiben in der Hochschulbildung*, Ruhr-Universität Bochum, März 2023, hss-opus.ub.ruhr-uni-bochum.de/opus4/frontdoor/index/index/docId/9734; *Schöppe*, ChatGPT und Prüfungsrecht an Hochschulen: Mehr Fragen als Antworten, 27.3.2023, euroacad.eu/artikel/chatgpt-und-pruefungsrecht-an-hochschulen-mehr-fragen-als-antworten/, zuletzt aufgerufen am 2.7.2023.

¹⁴ Vgl. safe.ai/statement-on-ai-risk#open-letter, zuletzt aufgerufen am 10.7.2023. Der offene Brief selbst ist vergleichsweise kurz: “Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.”

¹⁵ *United Nations*, Secretary-General Urges Broad Engagement from All Stakeholders towards United Nations Code of Conduct for Information Integrity on Digital Platforms, Pressemitteilung, 12.6.2023, online: press.un.org/en/2023/sgsm21832.doc.htm, zuletzt aufgerufen am 10.7.2023.

¹⁶ Einen exzellenten Überblick zu Deep Learning geben die drei Turing-Preisträger des Jahres 2018 *LeCun/Bengio/Hinton*, *Deep Learning*, *Nature*, 521(7553), 436-444, online: doi.org/10.1038/nature14539.

¹⁷ *Mikolov et al.*, Efficient Estimation of Word Representations in Vector Space. arXiv preprint, 2013, online: doi.org/10.48550/arXiv.1301.3781; *Mikolov et al.*, Distributed Representations of Words and Phrases and their Compositionality. In: *Burges et al.* (Hrsg.), *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, 2013, S. 3111-3119, online: proceedings.neurips.cc/paper_files/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf.

¹⁸ *Cho et al.*, Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. arXiv preprint, 2014, online: doi.org/10.48550/arXiv.1406.1078; *Sutskever/Vinyals/Le*, Sequence to Sequence Learning with Neural Networks. In: *Ghahramani et al.* (Hrsg.), *Advances in Neural Information Processing Systems 27 (NIPS 2014)*, S. 3104-3112, online: proceedings.neurips.cc/paper_files/paper/2014/file/a14ac55a4f27472c5d894ec1c3c743d2-Paper.pdf.

¹⁹ *Vaswani et al.*, Attention is All you Need. In: *Guyon et al.* (Hrsg.) *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, S. 5998 ff., online: proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

²⁰ *Brown et al.* (Fn. 11).

²¹ *Plaat*, *Deep Reinforcement Learning*, 2022, online: doi.org/10.1007/978-981-19-0638-1.

²² *Villalobos et al.*, Will we Run out of Data? An Analysis of the Limits of Scaling Datasets in Machine Learning, 2023, arXiv preprint, online: doi.org/10.48550/arXiv.2211.04325.

²³ *Meckel*, KI-Sprachmodelle: Es droht Textinzest. *Handelsblatt*, 14.3.2023, online: handelsblatt.com/meinung/kolumnen/kolumne-kreative-zerstoerung-ki-sprachmodelle-es-droht-textinzest/29034260.html, zuletzt aufgerufen am 9.7.2023.

²⁴ Einen aktuellen Überblick zu großen KI-Modellen, gegliedert nach Anwendungsgebieten und Trainingsformen, geben *Zhou et al.*, *A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT*, arXiv preprint 2023, online: doi.org/10.48550/arXiv.2302.09419.

²⁵ *Floridi*, AI as Agency without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models. *Philosophy & Technology*, 36:15, 2023, S. 1-7, online: doi.org/10.1007/s13347-023-00621-y.

²⁶ *Jahn*, Was Sie über OpenAIs KI wissen sollten, *Handelsblatt*, 19.6.2023, online: handelsblatt.com/technik/it-internet/chatgpt-was-sie-ueber-openais-ki-wissen-sollten-/28941524.html, zuletzt aufgerufen am 3.7.2023.

²⁷ *Brown et al.* (Fn. 11), S. 8.

²⁸ *Vincent*, OpenAI Co-Founder on Company's past Approach to Openly Sharing Research: 'We were Wrong'. *The Verge*, 15.3.2023, online: theverge.com/2023/3/15/23640180/openai-gpt-4-launch-closed-research-ilya-sutskever-interview, zuletzt aufgerufen am 2.7.2023. Einen Überblick zu den Anwendungen und Leistungsdaten auf der Basis der verfügbaren Literatur geben *Liu et al.*, *Summary of ChatGPT/GPT-4 Research and Perspective towards the Future of Large Language Models*, 2023, arXiv preprint, online: doi.org/10.48550/arXiv.2304.01852; 2023); ohne vertieften technischen Hintergrund verständliche Einführungen in die Funktionsweise finden sich bei *Linde*,

So funktioniert ChatGPT. Ein Deep Dive, 2023. Golem.de. IT-News für Profis, online: golem.de/news/kuenstliche-intelligenz-so-funktioniert-chatgpt-2302-171644.html, zuletzt aufgerufen am 2.7.2023; und *Riedl*, A very Gentle Introduction to Large Language Models without the Hype, 2023, mark-riedl.medium.com/a-very-gentle-introduction-to-large-language-models-without-the-hype-5f67941fa59e, zuletzt aufgerufen am 2.7.2023.

²⁹ Vgl. ChatGPT Statistics (2023) — Essential Facts and Figures (stylefactoryproductions.com), zuletzt aufgerufen am 2.7.2023. Allgemein zu Kosten und Ressourcenverbrauch aktueller Verfahren der KI vgl. *Crawford*, Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence, 2021.

³⁰ *Stiennon et al.*, Learning to Summarize with Human Feedback. In: *Larochelle et al.* (Hrsg.), Advances in Neural Information Processing Systems 33 (NeurIPS 2020), S. 3008-3021, online: proceedings.neurips.cc/paper_files/paper/2020/file/1f89885d556929e98d3ef9b86448f951-Paper.pdf.

³¹ Das *Fine Tuning* von Transformermodellen war eine gebräuchliche Anpassungsstrategie, als die Modelle noch kleiner waren, ist aber für Modelle in der Größenordnung von GTP-3 oder-4 kaum mehr praktikabel, weshalb nach neuen Adaptionstrategien gesucht wird, vgl. *Hu et al.*, Lora: Low-Rank Adaptation of Large Language Models, 2021. arXiv preprint, online: doi.org/10.48550/arXiv.2106.09685.

³² Zur systematischen Beschreibung der Fehler solcher Systeme vgl. *Borji*, A Categorical Archive of ChatGPT Failures, 2023 arXiv preprint, online: doi.org/10.48550/arXiv.2302.03494.

³³ Es ging dabei um die Entwicklung eines juristischen Expertensystems zu § 142 StGB zur Beantwortung von Fragen zu juristischen Sachverhaltsbeschreibungen in natürlicher Sprache. Trotz einer Projektdauer von fünf Jahren konnte lediglich eine Fallbeschreibung bearbeitet werden, auf die die Wissensbasis speziell zugeschnitten wurde, vgl. *Mielke*, Bewertung juristischer Informationssysteme, 2000, S. 8.

³⁴ *Mielke*, Legal Tech im Strafrecht: eDiscovery, Legal Prediction und weitere Einsatzmöglichkeiten, 2023, online: legal-tech.de/legal-tech-im-strafrecht/, zuletzt aufgerufen am 3.7.2023.

³⁵ OLGA steht für *Oberlandesgerichts-Assistent*, *Spoenle*, Künstliche Intelligenz im Einsatz bei Dieselverfahren, DRiZ 2023, 68 f.

³⁶ FRAUKE steht für *Frankfurter Urteils-Konfigurator, elektronisch*; *Sesing-Wagenpfeil et al.*, Legal Tech im Richterzimmer? Streiflichter aus Wissenschaft und Praxis zum KI-Einsatz bei Fluggastverspätungen, Jusletter IT, 30.3.2023, online: jusletter-it.weblaw.ch/issues/2023/30-maerz-2023/legal-tech-im-richte_ac6e81c7c2.html.

³⁷ *Spoenle* (Fn. 35).

³⁸ *Sesing-Wagenpfeil et al.* (Fn. 36); s. auch: Auf Augenhöhe mit Legal-Tech, vdz.org/moderne-digitale-justiz/auf-augenhoehe-mit-legal-tech, zuletzt aufgerufen am 10.7.2023.

³⁹ Grundlagenpapier zur 74. Jahrestagung der Präsidentinnen und Präsidenten der Oberlandesgerichte, des Kammergerichts, des Bayerischen Obersten Landesgerichts und des Bundesgerichtshofs vom 23. bis 25. Mai 2022 in Rostock, online: oberlandesgericht-celle.niedersachsen.de/download/184478

/Grundlagenpapier_der_Arbeitsgruppe_zum_Einsatz_von_KI_und_algorithmischen_Systemen_in_der_Justiz.pdf, Anhang, S. I ff., zuletzt aufgerufen am 3.7.2023.

⁴⁰ Grundlagenpapier zum Einsatz von KI und algorithmischen Systemen (Fn. 39), S. 3.

⁴¹ Vgl. *Rostalski*, Smart Sentencing: Legal Tech in der Strafzumessung, 2019, online: legal-tech.de/smart-sentencing-legal-tech-in-der-strafzumessung, zuletzt aufgerufen am 3.7.2023; *Mielke* (Fn. 34).

⁴² Grundlagenpapier zum Einsatz von KI und algorithmischen Systemen (Fn. 39), S. 32.

⁴³ Siehe dazu *Mielke* (Fn. 34).

⁴⁴ Das immer wieder beispielhaft herangezogene System zur Vorhersage von Entscheidungen des US Supreme Court, das mit einer Genauigkeit von 70,2 % vorhersagen kann, ob der Supreme Court die instanzgerichtliche Entscheidung bestätigen oder aufheben wird, nutzt zur Vorhersage vor allem formale Daten bzw. Metadaten, z.B. die Tendenz des Richters zur Aufhebung oder Bestätigung der instanzgerichtlichen Entscheidungen oder die politische Gesinnung, s. *Hoch*, Big Data und Predictive Analytics im Gerichtsprozess. Chancen und Grenzen der Urteilsprognose, MMR 2020, 295 ff., auch zu den Schwachpunkten eines weiteren häufig genannten Systems zur Vorhersage von Entscheidungen des EGMR (*Aletras et al.*, Predicting Judicial Decisions of the European Court of Human Rights: a Natural Language Processing Perspective, PeerJ Comput. Sci 2:e93, 2016; doi.org/10.7717/peerj-cs.93).

⁴⁵ Siehe *Mielke* (Fn. 34).

⁴⁶ Vgl. etwa *Adrian et al.*, Automatische Anonymisierung von Gerichtsurteilen – Eine Vision scheint realisierbar, Jusletter IT, 30.3.2023, online: jusletter-it.weblaw.ch/issues/2023/30-maerz-2023/automatische-anonymi_afe7683d74.html.

⁴⁷ Grundlagenpapier zum Einsatz von KI und algorithmischen Systemen (Fn. 39), S. 42.

- ⁴⁸ *Wissenschaftliche Dienste des Deutschen Bundestags*, Künstliche Intelligenz in der Justiz, Internationaler Überblick, 1.3.2021.
- ⁴⁹ *Herberger*, „Can AI Be a Fair Judge in Court?“ Denkt Estland so?, NJW-aktuell 37/2021, S. 19.
- ⁵⁰ Die Tagesordnung und die Stellungnahmen finden sich auf der Website des Landtags NRW, landtag.nrw.de/home/der-landtag/tagesordnungen/WP18/300/E18-362.html, zuletzt aufgerufen am 9.7.2023.
- ⁵¹ *Biallaß*, Stellungnahme des Deutschen EDV-Gerichtstags e.V. zur Anhörung „Einsatz von ChatGPT im Justizbereich“ im Landtag NRW, 13.6.2023, Stellungnahme 18/575, S. 2, online: landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMST18-575.pdf; *Hartmann*, Stellungnahme der Generalstaatsanwaltschaft Köln zur Anhörung „Einsatz von ChatGPT im Justizbereich“ im Landtag NRW, 13.6.2023, Stellungnahme 18/571, S. 3, online: landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMST18-571.pdf; *Johannisbauer*, Stellungnahme zur Anhörung „Einsatz von ChatGPT im Justizbereich“ im Landtag NRW, 13.6.2023, Hochschule für Polizei und öffentliche Verwaltung Nordrhein-Westfalen, Stellungnahme 18/590, S. 3, 12, online: landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMST18-590.pdf; *Nink*, Stellungnahme zur Anhörung „Einsatz von ChatGPT im Justizbereich“ im Landtag NRW, 13.6.2023, Stellungnahme 18/592, S. 2 ff., online: landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMST18-592.pdf; *Schlicht/Heetkamp*, Stellungnahme zur Anhörung „Einsatz von ChatGPT im Justizbereich“ im Landtag NRW, 13.6.2023, Stellungnahme 18/574, S. 14 f. online: landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMST18-574.pdf; s. ausführlich zu den rechtlichen und ethischen Anforderungen sowie den Grenzen des Einsatzes von KI das Grundlagenpapier zum Einsatz von KI und algorithmischen Systemen (Fn. 39), S. 6 ff.
- ⁵² *Hartmann* (Fn. 51), S. 3; *Johannisbauer* (Fn. 51), S. 3 ff.; *Nink* (Fn. 51), S. 4; *Schlicht/Heetkamp* (Fn. 51), S. 15.
- ⁵³ *Biallaß* (Fn. 51), S. 4.
- ⁵⁴ *Ministerium der Justiz des Landes Nordrhein-Westfalen*, Stellungnahme zur Anhörung „Einsatz von ChatGPT im Justizbereich“ im Landtag NRW, 13.6.2023, Vorlage 18/1022, S. 2, landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMV18-1022.pdf, zuletzt aufgerufen am 9.7.2023, wobei diese Aussage vor der Sachverständigenanhörung im Juni 2023 getroffen wurde.
- ⁵⁵ *Johannisbauer* (Fn. 51), S. 3 ff.
- ⁵⁶ *Schlicht/Heetkamp* (Fn. 51), S. 12.
- ⁵⁷ *Johannisbauer* (Fn. 51), S. 7 f.; *Schlicht/Heetkamp* (Fn. 51), S. 4 f.
- ⁵⁸ *Biallaß* (Fn. 51), S. 4; s. auch *Nink* (Fn. 51), S. 8.
- ⁵⁹ *Hartmann* (Fn. 51), S. 5 f.
- ⁶⁰ *Hartmann* (Fn. 51), S. 6.
- ⁶¹ *Biallaß* (Fn. 51), S. 5; *Grabmair* Stellungnahme zur Anhörung „Einsatz von ChatGPT im Justizbereich“ im Landtag NRW, Stellungnahme 18/588, S. 3, online: landtag.nrw.de/portal/WWW/dokumentenarchiv/Dokument/MMST18-588.pdf, zuletzt aufgerufen am 9.7.2023; *Hartmann* (Fn. 51), S. 6, 11, 13; *Nink* (Fn. 51), S. 23; *Schlicht/Heetkamp* (Fn. 51), S. 8.
- ⁶² *Biallaß* (Fn. 51), S. 12; *Nink* (Fn. 51), S. 23.
- ⁶³ *Biallaß* (Fn. 51), S. 12; *Schlicht/Heetkamp* (Fn. 51), S. 8.
- ⁶⁴ *Nink* (Fn. 51), S. 23.
- ⁶⁵ *Grabmair* (Fn. 61), S. 3; *Schlicht/Heetkamp* (Fn. 51), S. 8.
- ⁶⁶ *Biallaß* (Fn. 51), S. 12; *Schlicht/Heetkamp* (Fn. 51), S. 8; auch *Grabmair* (Fn. 61), S. 3 f.
- ⁶⁷ *Nink* (Fn. 51), S. 23.
- ⁶⁸ *Johannisbauer* (Fn. 51), S. 4; *Schlicht/Heetkamp* (Fn. 51), S. 8.
- ⁶⁹ *Johannisbauer* (Fn. 51), S. 4.
- ⁷⁰ *Biallaß* (Fn. 51), S. 17; *Schlicht/Heetkamp* (Fn. 51), S. 8.
- ⁷¹ *Hartmann* (Fn. 51), S. 14.
- ⁷² Grundlagenpapier zum Einsatz von KI und algorithmischen Systemen (Fn. 39), Anhang S. III, X.
- ⁷³ *Nink* (Fn. 51), S. 23 f., 25.
- ⁷⁴ *Nink* (Fn. 51), S. 24 f.
- ⁷⁵ *Biallaß* (Fn. 51), S. 6 f.; *Schlicht/Heetkamp* (Fn. 51), S. 2; siehe auch *Johannisbauer* (Fn. 51), S. 5.
- ⁷⁶ *Biallaß* (Fn. 51), S. 7.
- ⁷⁷ *Biallaß* (Fn. 51), S. 5; *Nink* (Fn. 51), S. 16.
- ⁷⁸ *Biallaß* (Fn. 51) S. 3; *Grabmair* (Fn. 61), S. 9 f.; *Johannisbauer* (Fn. 51), S. 9; *Schlicht/Heetkamp* (Fn. 51), S. 21.
- ⁷⁹ *Schlicht/Heetkamp* (Fn. 51), S. 21.
- ⁸⁰ *Grabmair* (Fn. 61), S. 8.
- ⁸¹ *Grabmair* (Fn. 61), S. 8 m.w.N.; s. auch *Johannisbauer* (Fn. 51), S. 10, der eine Standardisierung der Prompts fordert und/oder eine Schulung zu den Auswirkungen verschiedener Eingaben. Kürzlich hat der internationale Informationskonzern Thomson Reuters Casetext für 650 Mio.

Dollar erworben; sicher eine der bisher größten Transaktionen im Bereich Legal Tech, vgl. Der Tagesspiegel Background, 28.6.2023, [background.tagesspiegel.de/digitalisierung/thomson-reuters-uebernimmt-ki-start-up-casetext](https://www.tagesspiegel.de/digitalisierung/thomson-reuters-uebernimmt-ki-start-up-casetext), zuletzt aufgerufen am 9.7.2023.

⁸² *Biallaß* (Fn. 51), S. 4; *Grabmair* (Fn. 61), S. 7; *Hartmann* (Fn. 51), S. 14; *Johannsbauer* (Fn. 51), S. 9; *Nink* (Fn. 51), S. 6, 18; *Schlicht/Heetkamp* (Fn. 51), S. 15, 20.

⁸³ *Nink* (Fn. 51), S. 6.

⁸⁴ *Biallaß* (Fn. 51), S. 15; *Schlicht/Heetkamp* (Fn. 51), S. 2.

⁸⁵ *Biallaß* (Fn. 51), S. 15; *Johannsbauer* (Fn. 51), S. 6.

⁸⁶ *Biallaß* (Fn. 51), S. 15.

⁸⁷ *Nink* (Fn. 51), S. 16, 22.

⁸⁸ *Biallaß* (Fn. 51), S. 9.

⁸⁹ *Schlicht/Heetkamp*, (Fn. 51), S. 7 f.

⁹⁰ *Johannsbauer* (Fn. 51), S. 6.

⁹¹ *Johannsbauer* (Fn. 51), S. 5; *Schlicht/Heetkamp* (Fn. 51), S. 4.

⁹² *Biallaß* (Fn. 51), S. 17.

⁹³ *Nink* (Fn. 51), S. 27; s. auch *Biallaß* (Fn. 51), S. 7.

⁹⁴ *Grabmair* (Fn. 61), S. 2.

⁹⁵ Vgl. [justin-legal.com](https://www.justin-legal.com) „die digitale Mandatsannahme“, zuletzt aufgerufen am 2.7.2023.

⁹⁶ Vgl. *Mielke/Wolff* (Fn. 4).

⁹⁷ Vgl. [jupus.de](https://www.jupus.de), zuletzt aufgerufen am 2.7.2023.

⁹⁸ *Braegemann*, Gut geprompted ist halb gewonnen. Die zehn besten Prompts für ChatGPT und BingChat für Kanzleien, Legal Tech-Magazin 2/2023, [legal-tech.de/Magazine/FFI_Legal_Tech-Magazin_02-23.pdf](https://www.legal-tech.de/Magazine/FFI_Legal_Tech-Magazin_02-23.pdf), zuletzt aufgerufen am 2.7.2023.

⁹⁹ U.a. hier: „ChatGPT für Juristen. So gelingt der Einsatz von ChatGPT & Co.“, [fachseminare-von-fuerstenberg.de/fortbildung/chat-gpt?file=files/media/Dokumente/Prospekte/web_ChatGPT.pdf](https://www.fachseminare-von-fuerstenberg.de/fortbildung/chat-gpt?file=files/media/Dokumente/Prospekte/web_ChatGPT.pdf), zuletzt aufgerufen am 2.7.2023.

¹⁰⁰ *Regalia*, ChatGPT and Legal Writing: The Perfect Union?, Social Sciences Research Network (SSRN), Februar 2023, online: [ssrn.com/abstract=4371460](https://www.ssrn.com/abstract=4371460), dx.doi.org/10.2139/ssrn.4371460; *Schwarcz/Choi*, AI Tools for Lawyers: A Practical Guide, Social Sciences Research Network (SSRN), März, online: [ssrn.com/abstract=4404017](https://www.ssrn.com/abstract=4404017), dx.doi.org/10.2139/ssrn.4404017.

¹⁰¹ *The European Bar Association*, European Lawyers in the Era of ChatGPT. Guidelines on how Lawyers should Take Advantage of the Opportunities Offered by Large Language Models and Generative AI, Juni 2023, online: [fbe.org/wp-content/uploads/2023/06/European-lawyers-in-the-era-of-ChatGPT-FBE-Guidelines-on-how-lawyers-should-take-advantage-of-the-opportunities-offered-by-large-language-models-and-gene-kopia.pdf](https://www.fbe.org/wp-content/uploads/2023/06/European-lawyers-in-the-era-of-ChatGPT-FBE-Guidelines-on-how-lawyers-should-take-advantage-of-the-opportunities-offered-by-large-language-models-and-gene-kopia.pdf).

¹⁰² Zu ähnlichen Einsatzmöglichkeiten bei Anwält:innen s. *Braegemann*, Der ChatGPTorische Imperativ: „Schreibe stets so gut, dass du dir sicher bist, dass eine KI nicht genauso gut schreibe“, Legal Tech-Magazin 1/2023, online: [legal-tech.de/legal-tech-magazin-1-23/](https://www.legal-tech.de/legal-tech-magazin-1-23/), S. ff., zuletzt aufgerufen am 9.7.2023.