

Ricarda Aschauer, Linz

ChatGPT, Gemini & Co: Anwendungsfelder von Large Language Models im rechtlichen Kontext

Übersicht:

- I. Hintergrund
- II. LLM, Sprache und Recht
- III. Technische Limitationen und notwendige Kompetenzen
 - A. Halluzinationen
 - B. (Legal) Prompt Engineering
- IV. Anwendungsfelder von LLM-basierten (KI-)Systemen
 - A. LLM als Unterstützungstools in juristischen Bereichen
 - B. Einfacherer Zugang zum Recht durch LLM-basierte (KI-)Systemen
 - C. LLM im kernjuristischen Bereich
- V. Conclusio und Ausblick

I. Hintergrund

Die Entwicklung von intelligenten Systemen schreitet voran. Mit Large Language Models (LLM) wurden universell einsetzbare und enorm leistungsfähige Systeme entwickelt, die aufgrund ihrer sprachlichen Fähigkeiten verschiedene Anwendungsmöglichkeiten im Bereich der Rechtswissenschaften und der Rechtspraxis bieten. Dieser Beitrag beleuchtet diese Technologie, ihr Potential und ihre Limitationen im rechtlichen Kontext. Dabei soll ein Überblick über die aktuellen und die zukünftig denkbaren Einsatzfelder von LLM vermittelt werden.

GPT-4,¹ Gemini², oder LLaMA³ – seit dem Release von ChatGPT⁴ am 30. November 2022 sind Large Language Models (im Folgenden kurz LLM) und generative Künstliche Intelligenz (KI) im Allgemeinen aus der täglichen Medienberichterstattung nicht mehr wegzudenken. Die Einfachheit der Mensch-Maschine-Interaktion sowie die Universalität ihrer möglichen Einsatzfelder haben sie vermehrt in das gesamtgesellschaftliche Bewusstsein gerückt. LLM sind mithilfe von statistischen Berechnungen in der Lage, sprachliche Leistungen zu erbringen, die menschengemacht anmuten und kaum von humaner Spracharbeit zu unterscheiden sind. Vor allem im Bereich der Rechtswissenschaften ist ein gutes Sprachverständnis essentiell, was den Einsatz großer Sprachmodelle (in verschiedenster Ausprägung) lohnend erscheinen lässt. Zwar fehlt diesen Modellen ein Bewusstsein oder auch ein Verständnis, wie wir es aus der Tier- oder Menschenwelt kennen. Sie können jedoch die auf Semantik und dem Verständnis von Kausalität beruhenden sprachlichen Leistungen von Menschen mithilfe von Wahrscheinlichkeitsrechnungen überzeugend gut nachahmen. So erkennen LLM Muster in unserer Sprache beziehungsweise in Textdaten und verwerten diese,⁵ woraus sich authentisch wirkende Ergebnisse generieren lassen – und das in Sekundenschnelle. LLM bzw auf deren Basis entwickelte (KI-)Systeme können zum Erstellen von Texten aller Art verwendet werden, wobei dies auch unter Berücksichtigung bestimmter Stile oder Redeweisen möglich ist.⁶ Auch die menschliche Stimme und Gespräche⁷ können gänzlich künstlich verarbeitet und imitiert werden. Die Bedeutung von großen Sprachmodellen und artverwandten Technologien sowie ihrer universellen Verwendungsmöglichkeit spiegelt sich in diversen Untersuchungen wider: Rund ein Viertel aller Arbeitstätigkeiten in Europa und in den USA soll durch generative KI automatisiert werden können. Dem Rechtswesen wird dabei ein Automatisierungspotential von bis zu 44 % zugeschrieben.⁸ Auch die einzelne Person kann ihre Arbeitsleistung mithilfe generativer KI enorm verbessern:

- 1 GPT-4 gehört zur Generative-Pre-trained-Transformer-Serie von OpenAI. OpenAI, OpenAI's most advanced system, producing safer and more useful responses, <https://openai.com/gpt-4> (abgefragt 13. 2. 2024).
- 2 Entwickelt von Google; vgl *Pichai/Hassabis*, Introducing Gemini: our largest and most capable AI model, <https://blog.google/technology/ai/google-gemini-ai/#sundar-note> (Stand 6. 12. 2023).
- 3 LLaMa stammt vom Konzern Meta; vgl *Meta*, Introducing LLaMA: A foundational, 65-billion-parameter large language model, <https://ai.meta.com/blog/large-language-model-llama-meta-ai> (Stand 24. 2. 2023).
- 4 ChatGPT stellt die wohl bekannteste Anwendung eines Sprachmodells dar. *OpenAI*, Introducing ChatGPT, <https://openai.com/blog/chatgpt> (Stand 30. 11. 2022).
- 5 *Floridi*, AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models, *Philosophy & Technology* 2023, 1 f und 4, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4358789.
- 6 *Schwarcz/Choi*, AI Tools for Lawyers: A Practical Guide, *SSRN Electronic Journal* 2023, 6.
- 7 Vgl Google Soundstorm, <https://ai.googleblog.com/2023/06/soundstorm-efficient-parallel-audio.html> (Stand 22. 6. 2023).
- 8 Nur der Sektor des Office Support und der Administration wäre mit bis zu 46 % der Arbeitstätigkeiten demnach noch stärker von der Automatisierung betroffen; vgl *Briggs et al*, Global Economics Analyst The Potentially Large Effects of Artificial Intelligence on Economic Growth (2023) 6, <https://www.gspublishing.com/content/research/en/reports/2023/03/27/d64e052b-0f6e-45d7-967b-d7be35fabd16.html>.

Wissensarbeiter:innen sollen mithilfe dieser Technologie rund 12 % mehr Aufgaben erledigen können, dabei um bis zu 25 % schneller sein und die Qualität ihrer Arbeitsergebnisse um ca 40 % steigern können.⁹

Die gegenwärtige Leistungsfähigkeit von LLM wurde insbesondere durch die technologischen Fortschritte der jüngsten Vergangenheit ermöglicht: Erste Erfolge im Language Modelling¹⁰ konnten zwar bereits in den 1990er-Jahren durch statistische Language Models (LM) erzielt werden. Der heutige Stand der Technik konnte jedoch erst durch die Entwicklung von neuronalen Netzwerken sowie von vortrainierten Modellen erreicht werden.¹¹ Das Vortrainieren (Pre-Training) als das grundlegende Konzept, das die Entwicklung von Systemen wie ChatGPT und ähnlichen erst ermöglicht hat, setzt eine entsprechend große Verfügbarkeit von Rechenkapazitäten voraus. Die Sprachmodelle werden anhand von gigantischen (Text-)Datensätzen (vor)trainiert, sodass sie eine universelle (kontextuelle) Sprachrepräsentationen erlernen können. Insbesondere ist hierbei auf die jüngste Generation der Transformer-Technologie hinzuweisen. Sie ermöglicht es nicht nur, die semantische Bedeutung von Wörtern, sondern auch die kontextuelle Wortbedeutung zu erfassen. Aufgrund dieser Fähigkeiten gelten vortrainierte Modelle (Pre-Trained-Models, PTM) als vielseitig und für unterschiedlichste Anwendungsfällen in der natürlichen Sprachverarbeitung (Natural Language Processing, NLP) einsetzbar. Die ressourcenintensive Entwicklung eines LLM für nur einen bestimmten Aufgabenzweck wird daher nicht mehr zwingend notwendig sein.¹² Nunmehr kann es für den Einsatz für einen spezifischen Task ausreichen, ein (universell verwendbares) PTM durch sogenanntes Fine-Tuning zu spezialisieren.

II. LLM, Sprache und Recht

Das enge Verhältnis von Recht und Sprache ist unbestritten. Die teils beeindruckenden Ergebnisse von LLM-basierten (KI-)Systemen im juristischen Bereich, wie das Bestehen des US-amerikanischen Bar-Exam,¹³ können unter anderem wohl durchaus als in dieser wechselseitigen Beziehung begründet angesehen werden. Für die juristische Arbeit im Allgemeinen ist die Sprache in zweierlei Hinsicht maßgebend:

- 9 *Dell'Acqua et al*, Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality, Harvard Business School: Technology & Operations Management Unit Working Paper Series 2023, 11 f und 15, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4573321. Vgl auch *Pettinato Oltz*, ChatGPT, Professor of Law, SSRN Electronic Journal 2023 (passim), <https://ssrn.com/abstract=4347630>.
- 10 Vereinfacht ausgedrückt, zielt Language Modelling auf die Berechnung der Wahrscheinlichkeit einer Sprach- bzw Textsequenz ab, um die am wahrscheinlichsten folgenden Worte, Wortteile oder Satzteile (Tokens) vorherzusagen; vgl *Zhao et al*, A Survey of Large Language Models 1, <https://arxiv.org/pdf/2303.18223.pdf> (Stand 29. 6. 2023).
- 11 *Zhao et al*, Large Language Models 1 f.
- 12 *Qiu et al*, Pre-trained models for natural language processing: A survey, Science China Technological Sciences 2020, 1872 (1872 f und 1875 f).
- 13 Vgl beispielsweise *Katz et al*, GPT-4 Passes the Bar Exam, SSRN Electronic Journal 2023 (passim), <https://ssrn.com/abstract=4389233>.

Sie kann einerseits als das Medium verstanden werden, mit dem das Recht vergegenständlicht wird. Andererseits bildet sie aber auch das Werkzeug der Jurist:innen.¹⁴ Im Schrifttum wird die Auffassung vertreten, dass die Sprache und das Recht nicht ein- und dasselbe sind:¹⁵ Die positivistische Rechtsordnung besteht aus Rechtsnormen, die in Rechtstexten verschriftlicht werden. Die Rechtsnormen regulieren das Verhalten der Rechtsunterworfenen. Dabei stellen sie die (juristische) Bedeutung beziehungsweise den Gehalt einer Regelung dar. Die Rechtstexte (Normtexte) hingegen dienen als Vektoren: Sie können als der bloß schriftliche Ausdruck von Rechtsnormen angesehen werden, um die Normen effektiv kommunizieren zu können.¹⁶

Folgt man dieser Unterscheidung zwischen Sprache und Recht, liegt zunächst die Schlussfolgerung nahe, dass LLM und darauf aufbauende Systeme lediglich mit Texten und somit (nur) mit dem Rechtstext arbeiten könnten, nicht aber mit der Rechtsnorm als Gehalt einer Regelung. Denn letzterer kann durch die Verschriftlichung in Form eines Rechtstextes gegebenenfalls nur verkürzt dargestellt werden. Weitestgehend behoben würde dieses Problem wohl durch den Zugriff auf die (ebenfalls verschriftlichte) Literatur und Judikatur. Beides erlaubt regelmäßig, den Gehalt von Rechtsnormen festzumachen. Die Entwicklung eines solchermaßen „Legal Large Language Models“ könnte versucht werden.¹⁷ Denn schließlich hat auch ein:e Jurist:in bloß Texte zur Verfügung, anhand derer er:sie versucht, den Gehalt einer Rechtsnorm zu ermitteln.

Was bleibt, ist unter anderem die Frage, ob eine rein statistische Analyse aufbauend auf historischen Daten ausreichend für die juristische inhaltliche Arbeit ist – insbesondere wenn man bedenkt, dass sich an die konkrete, kontextuell mitunter unterschiedliche Wortbedeutung regelmäßig Rechtsfolgen knüpfen. Ein LLM gibt grundsätzlich nicht mehr und nicht weniger wieder als die wahrscheinlichste Wortabfolge entsprechend der Trainingsdaten – ohne konkret die dahinterliegende Sachmaterie zu begreifen.¹⁸ Ein LLM-basiertes (KI-)System kann somit, stark vereinfacht dargestellt, so exakt Auskunft geben, wie sich diese (statistisch) aus den Trainingsdaten ableiten lässt. Allein aufgrund dieser statistischen Textanalyse soll es LLMs möglich sein, scheinbare Logikschlussfolgerungen zu ziehen, die in Wahrheit aber nicht mehr sind als Muster des Trainingsdatenkorpus.¹⁹ Dabei ist es sogar möglich, den Stil bestimmter (notwendigerweise dem System bekannter) Disziplinen oder Autor:innen

14 *Heinrich Klang* sagte wohl in einer Vorlesung „Die Sprache ist das Werkzeug des Juristen. Jeder anständige Handwerker pflegt sein Werkzeug. Pflegen Sie daher Ihre Sprache.“ Vgl. *Schönherr*, Sprache und Recht (1985) 81; vgl. auch *Großfeld*, Sprache, Recht, Demokratie, NJW 1985, 1577 (1577).

15 *Kelsen* spricht beispielsweise von der Wortbedeutung und dem Gegenstand, den das Wort bezeichnet; vgl. *Kelsen*, Allgemeine Theorie der Normen (1979) 26. Umfassend zum Meinungsstand der Wissenschaft *Werni*, Die Rechtsnorm als Wille und sprachlicher Ausdruck, in *Burger et al.* (Hrsg.), Recht und Sprache: Tagung der Österreichischen Assistentinnen und Assistenten Öffentliches Recht (2019) 43 (52 ff.).

16 *Adamovich/Funk/Holzinger/Frank*, Österreichisches Staatsrecht I² (2011) 30.

17 Vgl. *Cui et al.*, ChatLaw: Open-Source Legal Large Language Model with Integrated External Knowledge Bases (2023) passim, <https://arxiv.org/abs/2306.16092> (Stand 28. 6. 2023).

18 *Floridi*, Philosophy & Technology 2023 1 f und 4.

19 Vgl. auch *Yuan*, Justiz GPT: Möglichkeiten und Grenzen des Einsatzes generativer Sprach-

imitieren zu lassen.²⁰ Für einen Einsatz in der kernjuristischen Tätigkeit kann einerseits sprechen, dass auch der Mensch in seiner Entscheidungsfindung wohl nicht perfekt und mitunter eine (fehlerbehaftete) Blackbox ist.²¹ Andererseits kann in der rein statistischen Antwortgeneration auch ein Potential von LLM gesehen werden, diese in Zukunft für den Einsatz als Legal Tech anhand von juristischen Textdaten zu trainieren oder mittels Fine-Tuning zu spezialisieren.²² So könnte ein Mensch unmöglich eine vergleichbare Informationsmenge verarbeiten. Die Antwort auf die Frage des Ob des Einsatzes von LLM-basierten (KI-)Systemen im kernjuristischen Arbeitsbereich wird wohl häufig schlicht davon abhängen, wie gut die vom System erzielten Ergebnisse im Einzelnen ausfallen. Für einige Bereiche scheint die Ergebnisqualität für den Einsatz zu juristischen Aufgaben bereits ausreichend zu sein.²³

Ein Problem, das sich bei einem rein anhand von historischen Daten trainierten Modell schon aus jetziger Sicht erahnen lässt, ist das einer mangelnden Rechtsfortentwicklung: Sollen LLM-basierte (KI-)Systeme maßgeblich in die Entscheidungsfindung oder gar in die Erstellung von Rechtstexten involviert werden,²⁴ gründen die Ergebnisse auf vergangenheitsbezogenen Sachverhalten und Texten. Das umfasst auch die darin enthaltenen Verzerrungen, gegebenenfalls Diskriminierungen, mittlerweile widerlegte wissenschaftliche Erkenntnisse, nunmehr sozialinadäquate Inhalte usw.²⁵ Im Weiteren kann es auch das Einfrieren der Werte einer Gesellschaft zum Zeitpunkt der Trainingsdaten bedeuten. Die Möglichkeit eines Lock-In auf Verzerrungen und Diskriminierungen sowie auf ein bestimmtes Wertekonstrukt scheint dadurch gegeben.

III. Technische Limitationen und notwendige Kompetenzen

A. Halluzinationen

Die nähere Betrachtung von LLM-basierten (KI-)Systemen und ihren Einsatzmöglichkeiten verlangt auch eine Auseinandersetzung mit ihren Limitationen. Trotz ihrer Potentiale bestehen auch für LLM Leistungsgrenzen. Das bekannteste Problem stellen

modelle bei gerichtlichen Entscheidungen, LTZ 2023, 195. Dazu, dass eine solche statistische Textberechnung (auch) fehleranfällig ist, siehe sogleich unter III.A.

20 *Szwarcz/Choi*, SSRN Electronic Journal 6.

21 Vgl. *Roth-Isigkeit*, DÖV 2020, 1018 (1024). So mag eine Entscheidung zwar logisch begründet werden, die genauen Vorgänge im Gehirn der entscheidenden Person werden dadurch aber ebenso wenig offengelegt wie mögliche entscheidungsirrelevante Motive, die die Person bei der Entscheidungsfindung tatsächlich geleitet haben mögen.

22 Vgl. *Cui et al*, ChatLaw, <https://arxiv.org/abs/2306.16092>.

23 Siehe IV.

24 Eine Verordnung der brasilianischen Stadt Porto Alegre zum Austausch gestohlener Wasserzähler soll vollständig von ChatGPT verfasst worden sein; vgl. *Knobloch*, Brasilien: Erste vollständig von KI geschriebene Verordnung, <https://www.heise.de/news> (Stand 4. 12. 2023).

25 Ausführlich hierzu *Eisenberger*, ChatGPT: Brauchen wir 2025 noch Jurist:innen? FS Merli (2023) 803 (810f), die insbesondere auch auf eine mögliche Gefährdung des Pluralismus hinweist.

sogenannte „Halluzinationen“ dar. Große Sprachmodelle können faktisch unwahre Aussagen generieren, die regelmäßig wie Tatsachen wirken und von Nutzer:innen nicht ohne Weiteres als unwahr erkennbar sind.²⁶ Halluzinationen können durch verschiedene (technische) Ursachen hervorgerufen werden.²⁷ Sie treten bei den verschiedensten Anwendungen auf, zum Beispiel bei Zusammenfassungen oder bei Chat-Konversationen. Weil es LLM immanent ist, Texte (bloß) anhand von Wahrscheinlichkeiten zu generieren, geht damit zwangsweise das Erschaffen inhaltlicher Informationen einher. Letztere müssen aber nicht notwendigerweise objektiv richtig sein. Trotz der laufenden Forschung an Abhilfemaßnahmen (beispielsweise Methoden zur Detektion von Halluzination²⁸) kann ihr Auftreten nicht ausgeschlossen werden. Die Verwendung von LLM-basierten (KI-)Systemen im juristischen Kontext verlangt daher besondere Vorsicht, Aufmerksamkeit, und insbesondere ein entsprechendes Problembewusstsein seitens der Nutzer:innen.²⁹ Das Potential von Sprachmodellen sollte schon allein aufgrund dieses Phänomens nicht in einer Art Wissensdatenbank gesehen werden, sondern vielmehr darin, menschliche Sprache maschinell zu verstehen – und somit darin, signifikante Unterstützungsleistungen im Arbeitsalltag zu erbringen, die in sensiblen Bereichen regelmäßig zumindest einer fachkundigen Nachkontrolle bedürfen.

B. (Legal) Prompt Engineering

Bekanntlich wird der Wert und die Effektivität einer Technologie maßgeblich von den Fähigkeiten und dem Wissen ihrer Nutzer:innen bestimmt. Insbesondere die Fähigkeit, präzise und effektive Eingabeanweisungen – sogenannte Prompts³⁰ – zu

26 *Feldman/Foulds/Pan*, Trapping LLM Hallucinations Using Tagged Context Prompts, <https://arxiv.org/abs/2306.06085> (Stand 9. 6. 2023).

27 Vgl *Ji et al*, Survey of Hallucination in Natural Language Generation, ACM Computing Surveys 3 u 31, <https://arxiv.org/abs/2202.03629> (Stand 7. 11. 2022).

28 Vgl beispielsweise *Manakul/Liusie/Gales*, SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models (passim), <https://arxiv.org/abs/2303.08896> (Stand 11. 10. 2023).

29 Für Halluzinationen, die bei der Verwendung von LLM im juristischen Kontext auftraten, können folgende Beispiele genannt werden: Zwei Anwälte wurden abgestraft, nachdem sie ChatGPT beim Verfassen eines Schriftsatzes verwendeten und dabei die als Nachweise generierten fiktiven Judikate nicht überprüft beziehungsweise dem Gericht so vorgelegt hatten. Vgl *Milmo*, Two US lawyers fined for submitting fake court citations from ChatGPT, <https://www.theguardian.com/technology/2023/jun/23/two-us-lawyers-fined-submitting-fake-court-citations-chatgpt> (Stand 23. 6. 2023). In Australien wurde ein LLM beim Verfassen eines Gutachtens einer Forschungseinheit im Rahmen einer parlamentarischen Untersuchung verwendet. Die Forschungsarbeit zu Wirtschaftsprüfungspraktiken der „Big Four“ (KPMG, Ernst & Young [EY], Deloitte und PwC) beinhaltete (unwahre) Anschuldigungen, in illegale Praktiken verwickelt zu sein. Vgl *Belot*, Australian academics apologise for false AI-generated allegations against big four consultancy firms, <https://www.theguardian.com/business/2023/nov/02/australian-academics-apologise-for-false-ai-generated-allegations-against-big-four-consultancy-firms> (Stand 2. 11. 2023).

30 Als Prompt ist eine Anweisung oder ein Eingabebefehl zu verstehen, die einem LLM zum Beispiel durch Texteingabe zur Verfügung gestellt werden und die das Modell beziehungsweise dessen Kontextwissen anpassen und dadurch dessen Fähigkeiten aufgaben- oder

formulieren, spielt eine entscheidende Rolle für die Qualität der Ergebnisse eines LLM-basierten (KI-)Systems.³¹ Ihre Formulierung ist maßgeblich, um genaue, relevante und schlüssige Antworten zu erhalten. Dementsprechend ist grundsätzlich auf das Verfassen klarer und spezifischer Eingabebefehle zu achten, sodass wenig Interpretationsspielraum für das System besteht. Daneben kann das Bereitstellen von einem (one-shot prompting) oder mehreren Beispielen (few-shot prompting) beziehungsweise des Lösungswegs die Ergebnisse stark verbessern.³²

Die Arbeit mit juristischen Texten stellt sich für Sprachmodelle und darauf aufbauende (KI-)Systeme sehr herausfordernd dar. Die regelmäßig langen Textkörper, komplexen Formulierungen und das spezifische Fachvokabular können die Ergebnisqualität stark beeinträchtigen. Legal Prompt Engineering umfasst verschiedene Methoden, um diese Limitationen zu kompensieren. Beim sogenannten Prompt Chaining handelt es sich um eine davon.³³ Es werden hierbei die für den Lösungsweg notwendigen Arbeitsschritte in einzelne Prompts heruntergebrochen. Diese einzelnen Eingabebefehle sowie deren (Zwischen-)Ergebnisse bilden die Grundlage für die nachfolgenden Prompts, wodurch eine sich aufbauende Kaskade an Eingaben und Ausgaben entsteht, die die Lösung komplexerer Probleme und die Bearbeitung längerer Texte ermöglicht beziehungsweise verbessert.³⁴

IV. Anwendungsfelder von LLM-basierten (KI-)Systemen

Aufgrund des universalen Charakters von LLM ergeben sich die verschiedensten Anwendungsfelder im rechtlichen Kontext, die sich angesichts der rasanten technologischen Entwicklung und immer besserer Systeme voraussichtlich noch (stark) erweitern werden. Großen Sprachmodellen konnte bereits ein verblüffend überzeugendes Potential in der juristischen Arbeit nachgewiesen werden: Nachdem das Modell GPT-3.5 im US-amerikanischen Bar-Exam³⁵ eine positive Bewertung noch

themenspezifisch verfeinern. *White et al*, A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT 1, <https://arxiv.org/abs/2302.11382> (Stand 21. 2. 2023).

31 Vgl. *Trautmann/Petrova/Schilder*, Legal Prompt Engineering for Multilingual Legal Judgment Prediction 2, <https://arxiv.org/abs/2212.02199> (Stand 5. 12. 2022).

32 *Reynolds/McDonell*, Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm, in CHI Conference on Human Factors in Computing Systems 2021, 2, <https://arxiv.org/abs/2102.07350v1> (Stand 15. 2. 2021).

33 *Trautmann/Petrova/Schilder*, Legal Prompt Engineering (passim), <https://arxiv.org/abs/2212.02199>.

34 *Yu/Quarley/Schilder*, Legal Prompting: Teaching a Language Model to Think Like a Lawyer (passim), <https://arxiv.org/pdf/2212.01326.pdf> (Stand 8. 12. 2022).

35 Das US-amerikanische Bar Exam ist die notwendige Prüfung, um in den USA den Anwaltsberuf ausüben zu können und ist dahingehend der österreichischen Rechtsanwaltsprüfung nach dem RAPG (Bundesgesetz vom 12. Dezember 1985, mit dem Bestimmungen über die Rechtsanwaltsprüfung und über sonstige Erfordernisse zur Ausübung der Rechtsanwaltschaft getroffen werden [Rechtsanwaltsprüfungsgesetz] BGBl 1985/556 idF BGBl I 2020/19) ähnlich. Im Bar-Exam werden sowohl Multiple-Choice als auch offene Fragen iSv Prüfungsfällen gestellt. Dass GPT-4 auch letztere überzeugend beantworten konnte, verdeutlicht die enormen Fortschritte, die im Bereich des NLP gemacht werden konnten.

verfehlte,³⁶ erreichte das Folgemodell GPT-4 nicht bloß eine solche, sondern konnte in manchen Fachgebieten über 90 % der Prüfungsfragen richtig beantworten.³⁷ Dabei wurde es nicht auf die Prüfung vorbereitet (zum Beispiel durch kontextverwandte Inputs), sondern erzielte dieses Ergebnis bereits in einem zero-shot³⁸-Verfahren.³⁹ Bei bestimmten Tätigkeiten sollen LLM-basierte (KI-)Systeme bereits mit ihrer Leistung an jene von Jurist:innen heranreichen können.⁴⁰ Im Legal-Tech-Bereich scheinen sie bereits insbesondere zur Vertragsanalyse und -erstellung eingesetzt zu werden,⁴¹ wo schon vergleichsweise länger Effektivierungsvorhaben⁴² und Digitalisierungsbestrebungen bestehen.⁴³ Das Potential von LLM-basierten (KI-)Systemen lediglich auf die (Unterstützung bei der) Vertragsprüfung zu reduzieren, wäre aber wohl äußerst kurzsichtig. Unabhängig von den jeweiligen rechtswissenschaftlichen Disziplinen lassen sich anhand internationaler Beispiele die folgenden Anwendungsbereiche identifizieren, die aus heutiger Sicht bereits adressiert werden oder die denkbar erscheinen, wobei ein Anspruch auf Vollständigkeit weder gestellt werden kann noch sollte.

A. LLM als Unterstützungstools im juristischen Berufsalltag

1. LLM für bloße Hilfstätigkeiten

Administrative Tätigkeiten stellen unter anderem aufgrund ihres repetitiven Charakters jenen Bereich dar, dem das größte Automatisierungspotential unserer gegenwärtigen Arbeitswelt zugeschrieben wird. Bis zu 46 % dieser Aufgaben könnten von generativer KI wie LLM-basierten (KI-)Systemen erfüllt werden.⁴⁴ Entsprechend viele Anwendungen adressieren diese administrativen Tätigkeiten.⁴⁵ Denkbar ist weiters

36 *Bommarito/Katz*, GPT Takes the Bar Exam, SSRN Electronic Journal (2022) 5 f, <https://ssrn.com/abstract=4314839>.

37 *Katz et al*, SSRN Electronic Journal 2023, 10.

38 Zero-Shot Prompting bedeutet die Verwendung von Eingabebefehlen, ohne dem System gleichzeitig Beispiele zur Verfügung zu stellen, die demselben oder einem vergleichbaren Lösungsweg folgen; vgl <https://www.promptingguide.ai/techniques/zeroshot> (abgefragt 17. 7. 2023).

39 *Katz et al*, SSRN Electronic Journal 2023, 9 f.

40 So soll die Leistungsfähigkeit von Sprachmodellen zum Teil an die menschliche bei der Vertragsanalyse heranreichen; vgl *Cui et al*, ChatLaw, <https://arxiv.org/abs/2306.16092>.

41 Eine Liste von LLM integrierenden LegalTech-Tools beziehungsweise ihrer Anbieter ist abrufbar unter <https://www.legaltechnologyhub.com/contents/the-use-of-large-language-models-in-legaltech/> (Stand 18. 2. 2023).

42 Man denke an die unter anderem bei Verlagen abrufbaren Vertragsschablonen.

43 Vgl beispielsweise *Woschnak*, Künstliche und emotionale Intelligenz: Gedanken zur Digitalisierung des Rechtslebens, in FS Bittner (2018) 825 (826).

44 Vgl *Briggs et al*, Effects of Artificial Intelligence on Economic Growth 7.

45 So wurde unter anderem der Microsoft 365 Copilot entwickelt, welcher angebunden an die von Microsoft angebotenen Anwendungen (Word, Excel, Teams usw.) administrative Tätigkeiten wahrnehmen soll. *Microsoft*, Introducing Microsoft 365 Copilot – your copilot for work, <https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/> (Stand 16. 3. 2023). Auch Google versucht die Implementierung generativer KI in Google Workspace und Google Cloud insbesondere mit dem Ziel, Alltagsstätigkeiten zu vereinfachen und zu automatisieren. *Google Workspace*, The

eine (bloße) Hilfsleistung durch LLM-basierte (KI-)Systeme bei juristisch-inhaltlichen Arbeiten, wie eine grammatikalische, kreative und stilistische (nicht aber inhaltliche) Unterstützung beim Verfassen von Texten wie Schriftsätzen, Gutachten usw. Die aktuellen Einsatzzwecke von KI zu solcherart unterstützenden Tätigkeiten reichen bis hin zur Anonymisierung von Gerichtsentscheidungen.⁴⁶ Verwaltungsbehörden, Regierungen und Gerichte setzen vermehrt auf LLM-basierte (KI-)Systeme für administrative oder unterstützende Tätigkeiten. Sprachmodelle werden von der Verwaltung insbesondere für die Textverarbeitung und -zusammenfassung eingesetzt. Das System Ion soll Anliegen der rumänischen Bürger:innen zusammenfassen und die Regierung in Echtzeit über diese informieren können.⁴⁷ Die Beamtenschaft von Singapur plant den Einsatz von ChatGPT, um Entwürfe von Schriftstücken und Reden zu generieren und um Recherchetätigkeiten zu unterstützen.⁴⁸ Im Vereinigten Königreich soll die Nutzung von LLM im Bereich der Verwaltung intensiviert werden.⁴⁹ Auch die österreichische Gerichtsbarkeit ist bemüht, sich die Vorteile neuer Technologien bei administrativen Tätigkeiten zu Nutze zu machen. So wurde in einem Pilotprojekt die Anonymisierung von Gerichtsentscheidungen erfolgreich automatisiert.⁵⁰

2. Recherche mithilfe von LLM-basierten (KI-)Systemen

Die bereits erläuterte Thematik der Halluzination kann bei der Verwendung eines LLM zu überzeugenden, jedoch faktisch unrichtigen Outputs führen. Problematisch kann dies vor allem dann sein, wenn LLM-basierte (KI-)Systeme zur rechtswissenschaftlichen Recherche genutzt werden, obwohl sie nicht als Recherchetool

next chapter of our Gemini era, <https://blog.google/technology/ai/google-gemini-update-sundar-pichai-2024/> (8. 2. 2024).

- 46 Bundesrechenzentrum GmbH, BMJ und BRZ gewinnen eAward für Einsatz von KI zur Anonymisierung von Gerichtsentscheidungen, <https://www.brz.gv.at/presse/eaward2022.html> (Stand 11. 10. 2022).
- 47 Sharma, Meet Ion, The “world’s First” AI Government Adviser Unveiled In Romania, <https://www.republicworld.com/world-news/europe/meet-ion-the-worlds-first-ai-government-adviser-unveiled-in-romania-articleshow.html> (2. 3. 2023); Dumitrescu, Romanian government to include ION, the first AI honorary advisor, <https://www.romania-insider.com/romanian-government-first-ai-honorary-advisor-2023> (Stand 1. 3. 2023).
- 48 PS News, SINGAPORE: ChatGPT to help civil servants, <https://psnews.com.au/2023/02/20/singapore-chatgpt-to-help-civil-servants/> (20. 2. 2023); Ang et al, Commentary: The good, bad and unknowns of letting Singapore’s civil servants use ChatGPT, <https://www.todayonline.com/commentary/commentary-good-bad-and-unknowns-letting-singapores-civil-servants-use-chatgpt-2119016> (1. 3. 2023); Smart Nation Singapore, Implication if civil servants use ChatGPT (PQ reply by SMS Janil Puthucherry), <https://www.smartnation.gov.sg/media-hub/parliament/28022023/> (Stand 28. 2. 2023).
- 49 Zunächst soll die Informationswebsite des Staates (Gov.uk) durch die Möglichkeit einer semantischen Suchfunktion sowie einer Chatbot-Funktion basierend auf einem großen Sprachmodell insbesondere für die Bürger:innen einfacher in der Bedienung werden, was die Anzahl von individuell zu beantwortenden Bürger:innenanfragen reduzieren soll. Vgl Wodecki, UK Taps Large Language Models to Reinvent Government, <https://aibusiness.com/nlp/uk-wants-to-tap-large-language-models-to-reinvent-government> (8. 3. 2023).
- 50 S FN 46.

konzipiert wurden und/oder ihre Ergebnisse nicht ausreichend überprüft werden.⁵¹ Nichtsdestotrotz ergeben sich verschiedene und sinnvolle Anwendungsmöglichkeiten von LLM im Rechercheprozess.⁵² So können LLM-basierte (KI-)Systeme im ersten Stadium einer Recherchefrage zur Ideenfindung und zum Brainstorming genutzt werden. Die natürliche Kommunikation kann die Kreativität steigern und damit neue Perspektiven eines Themas eröffnen. Ebenso unkritisch erscheint die Unterstützung durch ein LLM-basiertes (KI-)System bei der Eingangsrecherche zu einer (Rechts) Frage, denn sie sind regelmäßig in der Lage, einen niederschweligen Einstieg und besonders raschen Überblick über ein Thema zu vermitteln. Weiters können LLM für eine verbesserte Suche in Datenbanken eingesetzt werden. Sie ermöglichen eine semantische Suche, die eben nicht auf den genauen Wortlaut beziehungsweise die Zeichenabfolge, sondern auf die Wortbedeutung (gegebenenfalls im Kontext) abstellt.⁵³ Daneben wird versucht, möglichst verlässliche Recherchesysteme basierend auf LLM zu entwickeln. Diese sind etwa in der Lage, die für den Output verwendeten Quellen offenzulegen und somit ihre Ergebnisse zumindest nachvollziehbar und überprüfbar darzustellen.⁵⁴ Weiters können Recherchetätigkeiten innerhalb umfangreicher Akten unterstützt werden. So können mithilfe intelligenter Aktensysteme⁵⁵ die verfahrensrelevanten Sachverhaltsinformationen innerhalb kürzester Zeit aus den Akteninhalten herausgefiltert und übersichtlich dargestellt werden.⁵⁶

B. Einfacherer Zugang zum Recht durch LLM-basierte (KI-)Systeme

LLM-basierte (KI-)Systeme bieten nicht nur die Möglichkeit, Jurist:innen bei ihrer beruflichen Tätigkeit zu unterstützen, sondern sind – losgelöst von der aktuellen Rechtslage betrachtet⁵⁷ – grundsätzlich und technisch in der Lage, einfache Erstaus-

51 Ein bekanntes Beispiel ist der Einsatz von ChatGPT durch Anwälte zur Recherche von Präzedenzfällen und die Verwendung dieser KI-generierten Ergebnisse – ohne diese zu überprüfen – in einer Klagschrift; vgl. *Milmo*, Two US lawyers fined for submitting fake court citations from ChatGPT; für weitere Bsp siehe FN 29.

52 Mögliche Einsatzfelder von ChatGPT in der Recherchearbeit (insbesondere zum Thema Finance) untersuchten *Dowling/Lucey*, ChatGPT for (Finance) research: The Bananarama Conjecture, *Finance Research Letters* 53 2023, 2f, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4322651.

53 *Schumacher/Forcher/Tran*, Semantische Suche, in: *Dengel* (Hrsg.), *Semantische Technologien* (2012) 233.

54 Zum Beispiel der Chatbot „Perplexity“ generiert mithilfe eines LLM Antworten auf Fragen von Nutzer:innen und gibt dabei die verwendeten Quellen durch Weblinks an. <https://www.perplexity.ai/> (abgefragt 13. 2. 2024).

55 Vgl. *Aschauer*, *Digitale Assistenzsysteme in der Verwaltung*, ZVG 2024.

56 Das Hechinger Landgericht verwendet für diesen Zweck das System Codefy; <https://codefy.de/de/> (abgefragt 11. 2. 2024). Vgl. *Staatsministerium Baden-Württemberg*, Landgericht Hechingen testet neue Gerichtssoftware, <https://www.baden-wuerttemberg.de/de/service/presse/pressemitteilung/pid/landgericht-hechingen-testet-neue-gerichtssoftware> (Stand 1. 12. 2022).

57 Insb ist hier auf den Rechtsanwaltsvorbehalt der Rechtsanwälte:Rechtsanwältinnen nach § 8 RAO (Rechtsanwaltsordnung [RAO] RGBl 96/1868 idF BGBl I 2023/39) sowie auf § 1 Winkelschreibereiverordnung (Verordnung des Justizministeriums vom 8. Juni 1857,