

Den Abschluss bildet ein vierter Teil, der einen Ausblick auf neue und zukünftige unionale Rechtsakte eröffnet.

Pflichtlektüre

„There's no chance that the iPhone is going to get any significant market share. No chance.“ (Steve Ballmer, ehemaliger CEO Microsoft Corp., 2007) Solche Fehlprognosen über die umwälzende Relevanz der digitalen Transformation darf sich die juristische Ausbildung in Zeiten zurückgehender Studierendenzahlen und von Fachkräftemangel nicht leisten, um im Wettbewerb um die klügsten Köpfe attraktiv zu bleiben. Das fulminante Lehrbuch zum „Recht der Digitalisierung“ leistet hierzu einen kaum zu überschätzenden Beitrag. Ein gesamtes Rechtsgebiet wird erstmals für Ausbildungs- und Prüfungszwecke vermessen. Das Werk präsentiert sich nicht nur für die einschlägigen Schwerpunktbereiche sowie Bachelor- und Masterstudiengänge als Lehr- und Lernfundament. Auch im Pflichtfachbereich führt kein Weg am neuen Martini/Möslein/Rostalski vorbei.

auch Recht

Wiss. Mitarbeiter Christian Braun, Wiss. Mitarbeiterin Sarah Großkopf und Wiss. Mitarbeiter Simon A. Nonn, Passau

Projekt DeepWrite: Kann ChatGPT Lehrende an der Universität ersetzen?

Mit dem vom Bundesministerium für Bildung und Forschung geförderten Projekt DeepWrite an der Universität Passau wird genau diese Frage untersucht – vor allem mit Blick auf die Vermittlung der juristischen Argumentationsfähigkeiten und speziell des Gutachtenstils anhand von KI-Feedback.

Die Corona-Pandemie und die damit einhergehenden immer rasanter werdenden Entwicklungen und Fortschritte im Bereich Digitalisierung haben in den letzten Jahren gezeigt, dass sich Didaktik im Hochschulbereich aktuell im Wandel befindet und dass dieser Prozess aktiv beeinflusst werden kann und sollte, um die Zukunfts- und Wettbewerbsfähigkeit der Universitäten zu erhalten. Ein großer Bestandteil dessen ist die Nutzung innovativer Technik und Tools, wie zB Künstliche Intelligenz (KI), insbesondere *large language models* (LLM) und *natural language processing* (NLP), um digitale Lehr- und Lernräume für Studierende zukünftiger Generationen zu schaffen.

Bildung im Wandel?

Eine „Hochschule der Zukunft“ bietet den Studierenden mehr Flexibilität und Individualisierung. Digitale Lernangebote können angepasst an das eigene Lerntempo und -verhalten eingesetzt werden und sind standortunabhängig nutzbar. Zudem kann das Lernverhalten analysiert und infolgedessen können personalisierte Empfehlungen gegeben werden, um das Lernen effektiver zu gestalten. Dies hilft dabei, den (emotionalen) Druck im Studium zu redu-

zieren. Der Wunsch nach einer solchen Entlastung wurde kürzlich auch in der größten bisher durchgeführten iur.reform-Studie zur juristischen Ausbildung dargelegt.

Daneben hat die gerade genannte Studie gezeigt, dass vielfach eine bessere Berücksichtigung alternativer Lösungswege gefordert wird. Auch insofern verspricht KI einen Lösungsansatz, da auf digitalem Wege Musterlösungen in Form von Baumstrukturen erstellt und so verschiedene Lösungswege umfassend dargestellt werden können. Dabei steigt die Länge und Komplexität von Musterlösungen auf ein Maß, das für menschliche Korrektoren nur mit einigem Aufwand bewältigt werden kann; für KI sollte ein Abgleich hingegen vergleichsweise einfach sein.

Vielversprechend ist dabei die Verknüpfung von digitalen Elementen mit studierendenzentrierten Ansätzen des Lernens. So entwickelt das Projekt DeepWrite ein KI-basiertes Tool, das den Erwerb von Schreib- und Argumentationskompetenzen in den Fachbereichen Jura und Wirtschaft fördern soll, indem es den Studierenden unmittelbar individuelles Feedback auf deren Antworten gibt. Wie wichtig Feedback für die Studierenden ist, zeigen nicht nur persönliche Erfahrungen, sondern auch verschiedene Studien. Lerneffekte bleiben bei schlechtem oder fehlendem Feedback zu weiten Teilen auf der Strecke. Da menschliches Feedback oft weniger verfügbar oder zumindest kostspieliger ist, lohnt es sich gerade bei Studiengängen mit einem überproportionalen Verhältnis von Studierenden zu Dozentinnen und Dozenten umso mehr, auf alternative Methoden auszuweichen und neue Wege zu gehen – einer davon könnte KI sein.

ChatGPT für oder vs. Studierende

Bislang scheitern KI-Modelle noch daran, komplexe juristische Gutachten zu erstellen und Falllösungen selbstständig zu bearbeiten; insbesondere stellen die verschiedenen ineinander geschachtelten Ebenen bei der juristischen Argumentation und Prüfung eine große Herausforderung dar. Es scheint jedoch lediglich eine Frage der Zeit zu sein, bis die KI hier aufholt und zumindest kürzere, einfach gelagerte Fälle zufriedenstellend lösen kann. So hat sich das Sprachmodell GPT-4 von OpenAI von seiner Vorgängerversion GPT-3.5 schon innerhalb weniger Monate deutlich abgesetzt.

Dass die (juristischen) Fähigkeiten auch von der Rechtsordnung und den damit oft einhergehenden Prüfungssettings abhängen, lässt sich in den USA sehen. Dort hat ChatGPT-4 das *Uniform Bar Exam* erfolgreich gemeistert; die Vorgängerversion war noch an der Bestehensquote von 120 der 200 gestellten Multiple-Choice-Fragen gescheitert. Zudem konnte GPT-4 auch den „Essay“ und den „Performance Test“ des *Uniform Bar Exam* bestehen;¹ GPT-3.5 war bei diesen beiden Teilen noch gar nicht angetreten. Der Fortschritt lässt sich allerdings im Hinblick auf das deutsche

¹ https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389233, zuletzt aufgerufen am 9.1.2024.

Rechtssystem etwas relativieren. So ging es um englischsprachige Texte und die KI wurde größtenteils auch in Englisch trainiert. Auch beruht das US-amerikanische Rechtssystem vorwiegend auf sog. *precedents*, also richtungsweisenden richterlichen Entscheidungen, die im Trainingsdatensatz des Sprachmodells mutmaßlich in erheblichem Maße vorhanden waren. Vor allem an der juristisch hinreichend detaillierten Datenbasis fehlt es der KI hierzulande zumindest derzeit noch. Auch an diesem Beispiel zeigt sich jedoch, wie rasant sich die Systeme entwickeln; Beiträge dazu lassen sich in jeder Veröffentlichungsform finden.

Für den deutschsprachigen Raum ist deshalb eine logische Konsequenz, die KI hinsichtlich des Feedbacks zu juristischen Sachverhalten und Lösungen zu trainieren und an die spezielle juristische Argumentationstechnik heranzuführen, um die Studierenden unterstützen zu können.

Das Projekt DeepWrite untersucht dabei unter anderem, wie Antworten von Studierenden durch verschiedene KI-Modelle und studentische Korrektoren bewertet werden und wie das Feedback angenommen wird. So soll untersucht werden, wie verlässlich das Feedback von KI im Moment ist und wie es durch Dozentinnen und Dozenten sowie Studierende gewinnbringend eingesetzt werden kann. Dabei spielt vor allem die Interrater-Reliabilität eine Rolle, mit der die Übereinstimmung des unterschiedlichen Feedbacks angegeben und anschließend bewertet werden soll. Außerdem findet eine qualitative Untersuchung des KI-Feedbacks statt; dabei wird auch der Einfluss auf den Lernfortschritt und die Motivation der Studierenden untersucht.

Prompt engineering als Accelerator und Begrenzer

Für den Einsatz von bestehenden KI-Modellen in der Lehre ist vor allem die Entwicklung sog. *prompts* relevant, also des Inputs, zu dem das KI-Modell anschließend einen Output erzeugt – vereinfacht gesagt: des Arbeitsauftrags. *Prompts* steuern also die Herangehens- und Arbeitsweise der KI und sind für die Qualität des Feedbacks von zentraler Bedeutung. Dabei bestehen die *prompts* des Projekts DeepWrite aus fallbezogenen Elementen, insbesondere dem Sachverhalt und der Musterlösung für den konkreten Fall, und aus allgemeinen Elementen, die etwa den Stil und die Struktur des KI-Feedbacks betreffen.

Durch entsprechende *prompts* lässt sich etwa erreichen, dass die KI den Stil und Inhalt studentischer Arbeiten bewertet oder dass sie verschiedene Abschnitte eines Lösungsvorschlags separat analysiert.

Ein großer Vorteil ist, dass vertiefte technische Vorkenntnisse nicht zwingend notwendig sind. Stattdessen ist es wichtig, dass sich der Entwickler genaue Vorstellungen davon macht, welches Ziel er konkret verfolgt, und dieses sprachlich präzise formuliert. Außerdem empfiehlt sich eine iterative Herangehensweise, bei der das Feedback eines *prompts* bewertet und dieser anschließend modifiziert wird, wobei die Anzahl der Wiederholungen mit zunehmender Komplexität der gewünschten Aufgabe ansteigt.

Dabei ist auch zu berücksichtigen, dass selbst erprobte *prompts* nicht immer zuverlässige Ergebnisse liefern. Dies liegt einerseits in der Natur von LLMs: Bei ChatGPT etwa lässt sich anhand verschiedener Parameter (etwa Temperatur oder top-P) einstellen, wie „kreativ“ bzw. wie vorhersehbar ein Ergebnis ist. Erfahrungsgemäß steigt die durchschnittliche Qualität des Feedbacks, wenn ein gewisses Maß an „Kreativität“ zugelassen wird, aber auch das Risiko von Ausreißern mit sehr schlechter Qualität.

Andererseits ist, zumindest bei der Nutzung von „geschlossenen“ Sprachmodellen wie denen von OpenAI, für die Nutzenden nicht transparent, wie sich das Sprachmodell entwickelt, wie sich größere Updates auswirken und ob etwa die Qualität für den konkreten Einsatzzweck durch ein Update abnimmt. Auch das führt dazu, dass die Effektivität eines *prompts* ständig schwankt und *prompts* regelmäßig neu angepasst werden müssen.

Die zweite Seite der Medaille

Der Einsatz von KI in der universitären Lehre bietet jedoch nicht nur Chancen, sondern stellt auch eine nicht zu vernachlässigende Herausforderung dar. Bei jedem Einsatz von KI an einer Universität muss sich die sie nutzende Person ihrer Verantwortung gegenüber den Studierenden bewusst sein. Diese ist umso höher, je größer der Eingriff in den Lebensbereich der Studierenden ist, was man auch anhand des aktuellen Entwurfs zur anstehenden KI-Verordnung² sehen kann (Erwägungsgrund 35). Hiernach werden KI-Systeme, die unter anderem Studierende in Prüfungen bewerten, als hochriskant (Art. 6 iVm Anhang III Ziff. 3 KI-VO-E) eingestuft.

Diese Einstufung löst eine Reihe von Anforderungen (Art. 8–15 KI-VO-E) aus, die unter anderem von einem Risikomanagementsystem über Dokumentations-, Aufzeichnungs- und Transparenzpflichten bis zu menschlicher Aufsicht reichen. Begründet wird dies für den Bildungsbereich damit, dass hier nicht unerheblich in den Bildungsverlauf und damit auch das Berufsleben eingegriffen wird, was wiederum maßgeblich die künftige Sicherung des Lebensunterhalts beeinflusst.

Es besteht auch die Befürchtung, dass bei einer unreglementierten Nutzung das Recht auf allgemeine und berufliche Bildung sowie auf Nichtdiskriminierung verletzt werden könnte und historische Diskriminierungsmuster fortgeschrieben und verfestigt werden. Bei Letzterem handelt es sich um das Problem des sog. *social bias*, einen systematischen, selbsterhaltenden Fehler in einem System zur Entscheidungsfindung. Die Ursachen können hierbei vielfältig sein: Sie reichen von den Entwickelnden des Modells selbst über die Zielsetzung bis zur Datenauswahl. Ist der zum Training verwendete Datensatz zu klein oder bildet die menschliche Diversität nicht hinreichend ab, kommt es zu einem verzerrten Datensatz auf Grund von überrepräsentierten Gruppen (White, Educated, Industrialized, Rich, and

² <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52021PC0206>, zuletzt aufgerufen am 8.1.2024.

Democratic, sog. *WEIRD-sample*). Wohnt der KI erstmal ein solcher *bias* inne, ist es kaum mehr möglich, diesen in den zugrundeliegenden Algorithmen zu identifizieren (sog. *Blackbox-Problematik*).

Umgestaltung des Prüfungssystems erforderlich?

Insgesamt kann festgehalten werden, dass KI für die juristische Tätigkeit viel Potenzial bereithält: Schriftsätze können entworfen und zusammengefasst, rechtliche Probleme im „Dialog“ mit der KI herausgearbeitet werden. Dies erzeugt jedoch für Lehrende und Studierende einen Zielkonflikt. Zwar bereitet ein Lernen mit der KI möglicherweise gut auf zukünftige juristische Tätigkeiten und die sog. *21st Century Skills* (Kreativität, kritisches Denken, Kommunikation und Kollaboration) vor; das aktuelle Hochschul- und Prüfungsrecht sowie die meisten Studien- und Prüfungsordnungen sind hingegen noch nicht bzgl. des Einsatzes von KI überarbeitet worden, wodurch für Studierende, Lehrende und Mitarbeitende in der Universitätsverwaltung eine gewisse Rechtsunsicherheit besteht. Gemäß Art. 84 III 1 BayHIG regeln die jeweiligen Prüfungsordnungen wesentliche Fragen im Hinblick auf Prüfungsanforderungen und -verfahren (s. zu weiteren Regelungen auch S. 2). Hier müssen also an den Universitäten Regelungen geschaffen werden, die einerseits einen fairen und reflektierten Einsatz von KI gewährleisten, andererseits praxisferne Überregulierung vermeiden. Zudem muss sichergestellt werden, dass alle Studierenden gleichberechtigten, kostengünstigen Zugang zu KI-Modellen haben. Dies kann beispielsweise über eine Hochschullizensierung sichergestellt werden.

Diese Hürden bedeuten jedoch nicht, dass die derzeitigen technischen Entwicklungen ausgeblendet und eine Anpassung des Prüfungssystems vermieden werden sollten. Gerade im Hinblick auf Studien- bzw. Seminararbeiten muss genau überlegt werden, welche Kompetenzen von Studierenden in Zukunft verlangt werden sollten und wie man die sinnvolle Nutzung von KI und Co. als Prüfungsleistung einbinden kann; eine breitere Diskussion über die Zielsetzung von Hausarbeiten ist wünschenswert.³ Unabhängig von technischen Entwicklungen bleibt selbstverständlich auch die generelle Debatte über eine Umstrukturierung des Jura-Studiums auf der Tagesordnung.

Ausblick

Gerade im deutschsprachigen juristischen Bereich fehlt der KI noch eine hinreichend breite Datengrundlage. Dies führt einerseits dazu, dass die Sprachmodelle nicht in Bezug auf Analyse und Verfassen solcher Texte trainiert wurden, und andererseits dazu, dass sie auf konkrete Anfragen hin keine Datenbanken durchsuchen können – zumindest, solange die großen Datenbankanbieter (beck-online, juris, lexis-nexis) keinen Zugriff gewähren. Dabei spielt nicht nur das Verständnis von Normen sowie deren Zitierweise und Auf-

bau in Absätze und Sätze eine Rolle, sondern auch der systematische Zusammenhang verschiedener Normen untereinander.

Hinzu kommen die wissenschaftlichen und wertenden Aspekte in der juristischen Ausbildung – besonders in der Ersten Juristischen Prüfung –, die dazu führen, dass es fast nie nur eine Lösung gibt, sondern meistens verschiedene Meinungen miteinander in Wettstreit treten. Hier die richtige Schwerpunktsetzung zu finden, alternative Lösungswege in gleichem Maße zuzulassen und das nötige Feingespür für (vermeintlich) minimale Abweichungen zu etablieren, dürfte noch einige Zeit der technischen Entwicklung und vor allem der Kooperation zwischen Technik, juristischer Wissenschaft und Wirtschaft in Anspruch nehmen. Dabei gilt es, sich früh im Studium mit diesen Entwicklungen zu beschäftigen und offen für weitere sowie sensibilisiert für deren Anwendung zu sein. Das Projekt DeepWrite verbindet genau diese Bereiche und schafft durch seine interdisziplinäre Ausrichtung auch Anwendungsbereiche außerhalb des Juristischen.

Lehrende an der Universität werden durch die Anwendung somit mittelfristig keinesfalls ersetzt. Gerade in Studiengängen wie Rechtswissenschaft kann die KI zunächst vor allem unterstützende Funktion übernehmen und hält die zeitlichen Kapazitäten der Lehrenden für Bereiche frei, in denen sie effektiver genutzt werden können. Dass es auch in anderen Bereichen nicht um das Ersetzen menschlicher Qualifikationen geht, zeigt sich im Übrigen auch in parallelen Diskussionen über die Nutzung von KI in der Anwaltschaft, der Verwaltung oder im Gesundheitswesen.

Es ist aber eine Neuausrichtung bzw. Anpassung unseres Bildungssystems notwendig; die technische Fortentwicklung ist eine Frage der menschlichen Genese – diese einfach unbeachtet zu lassen oder ihr sogar im Weg zu stehen, war auch in der Vergangenheit selten zielführend.

► www.ird.uni-passau.de/kramer/projekt-deep-write

Veranstaltungshinweis

Call for Papers: **10. Prozessrechtstagung: Verfahrensrecht und Rechtsstaat – 30. und 31.8.2024 in Bonn**

Wir freuen uns, die zehnte Prozessrechtstagung im Jahr 2024 an der Rheinischen Friedrich-Wilhelms-Universität Bonn auszurichten. Am 30. und 31. August wollen wir uns mit den Wechselwirkungen zwischen Verfahrensrecht und dem verfassungsrechtlichen Gebot der Rechtsstaatlichkeit auseinandersetzen.

Anknüpfend an das Generalthema könnten prozessrechtlich orientierte Vorträge beleuchten, welche Anforderungen das Rechtsstaatsgebot an das Verfahrensrecht stellt; umgekehrt könnte auch untersucht werden, inwiefern die Ausgestaltung des Verfahrensrechts zur Optimierung der Rechtsstaatlichkeit beiträgt und wo Verbesserungspotenziale bestehen.

³ www.bpb.de/lernen/digitale-bildung/werkstatt/521773/das-ende-der-hausarbeit-ki-an-hochschulen, zuletzt aufgerufen am 8.1.2024.