

Alles, was recht ist: KI in Recht und Justiz

Jan Mönikes, Rechtsanwalt und Partner bei Schalast Rechtsanwälte in Berlin, mit Spezialisierung auf Medien-, Datenschutz- und Vereinsrecht, sowie Sprecher des Managerkreises Baden-Württemberg.

Seit der Öffnung von ChatGPT Ende 2022 ist der Fortschritt in Anwendungen von künstlicher Intelligenz (KI) offensichtlich. Im **Rechts- und Justizbereich** lassen sich die neuen Möglichkeiten, mit einer Maschine über Sprache, Texte und Bilder zu kommunizieren und ohne aufwändige eigene Vorstrukturierung selbst große und komplexe Datenmengen verarbeiten zu können, besonders leicht nutzen. Denn Jura ist eine Textwissenschaft auf Grundlage der Hermeneutik. Ihre Erkenntnisse über die Bedingungen der Möglichkeit von Sinnverstehen werden als juristische Methode auf die Auslegung von Gesetzen und anderer juristischer Texte angewendet. Gerade diese Systematik erleichtert es, sinnvolle Ergebnisse mit Hilfe von KI zu erzielen.

Kreativität unerwünscht, rasante Ausbreitung

Das Ziel der Rechtswissenschaft, bestmöglich mit strukturierten Gedanken sprachlich erfassbare (Lebens-) Sachverhalte in einem logischen System zu analysieren, muss im praktischen Einsatz im Rechtsstaat allerdings den Verbindlichkeitsanspruchs von Gesetzen und den Deutungsvorrang gerichtlicher Urteile respektieren. Die KI ist somit **weit weniger frei** als in anderen Anwendungsbereichen, bei denen kreative Ergebnisse sogar gewollt sind. Da KI zudem noch lange an normativen Fragen scheitern dürfte, müssen sich Jurist_innen auch nicht von akuter Arbeitslosigkeit bedroht fühlen – anders als andere Berufsgruppen.

KI in Recht und Justiz ist nicht gänzlich neu: Sie wird bereits genutzt, um große Informationsmengen in komplexen Verfahren zu bewältigen. Etwa bei der Suche nach relevanter Kommunikation in Akten und Datenspeichern für Prozesse im sog. „Dieselskandal“. Seit ChatGPT breitet sich ihr Einsatz zur Unterstützung in der Anpassung von Standardverträgen, der Individualisierung und Aktualisierung von Vertragsmustern, der Zusammenfassung von Urteilen, der Korrektur von Schriftsätzen oder bei rechtlich stimmiger Übersetzung in andere Sprachen **in rasantem Tempo** aus. Die Steigerung von Effizienz und Qualität ist so hoch, dass selbst unter den technischen Neuerungen gegenüber eigentlich sehr skeptischen Juristinnen und Juristen der Einsatz von ChatGPT & Co. kaum ernsthaft hinterfragt wird. Der Fachkräftemangel macht KI besonders willkommen, ihre Integration in Office-Produkte etabliert – jedenfalls außerhalb von Behörden – neue Standards für die praktische juristische Arbeit.



Predictive Analytics und nötige Regulierung

Die Fähigkeit der KI, Muster in Daten zu erkennen und Zusammenhänge herzustellen, die für menschliche Bearbeiter zum Teil nur schwer ersichtlich wären, führt bei anderen Anwendungen dagegen zu erheblicher Besorgnis: *Predictive Analytics*, etwa bei der Vorhersage von Straftaten oder voraussichtlichen Wertungen eines Gerichts, wird **zu Recht kontrovers diskutiert**. Gleiches gilt beim Einsatz von KI in der Bildanalyse, die das Thema Videoüberwachung auf ein ganz neues Niveau heben könnte oder ihr Einsatz für eine „Chatkontrolle“, deren (ungewolltes) Ergebnis eine nahezu totale Überwachung der Kommunikation im Internet sein dürfte. Da es selbst in diesen heiklen Fragen jedoch gute Begründungen gibt, die den Einsatz von KI mindestens für Teilbereiche rechtfertigen wollen, sind die politischen und regulatorischen Herausforderungen für KI im Rechtsbereich nicht zu unterschätzen.

Die Regulierung muss einen Balanceakt vollziehen: Einerseits soll sie Innovation und technologischen Fortschritt fördern, andererseits aber auch Missbrauch verhindern und ethische Standards sicherstellen. Dies erfordert eine möglichst klare Gesetzgebung, die sowohl die Haftung bei Fehlern von KI-Systemen regelt als auch den Datenschutz und die Transparenz der Entscheidungsprozesse gewährleistet.

Der AI-Act der EU ist ein wichtiger Schritt in Richtung einer solchen Regulierung.

AI Act: Faire Standards nötig, aber kein „One-size-fits-all“

Klare Regelungen zur Haftung, zum Schutz der Grundrechte und zur Förderung von Transparenz sind wichtig, die Regulierung muss aber flexibel genug gestaltet sein, um mit der Entwicklung der KI-Technologie Schritt halten zu können. Ihre Implementierung ins deutsche Recht und die Weiterentwicklung in der EU darf daher nicht nur „Bedenken“ spiegeln, sondern muss auch **faire Standards** setzen. Denn gute Trainingsdaten sind für qualitativ hochwertige KI-Ergebnisse schlicht essenziell. Konflikte mit dem heutigen Urheberrecht, Datenschutz, Persönlichkeitsrechten, Jugendschutz und anderen restriktiven Regeln müssen daher gelöst werden, will man eine einseitige Prägung der KI allein auf andere Sprach-, Rechts- und Kulturräume verhindern. Überzogene Wünsche etwa an Transparenz, Compliance

oder der Nachvollziehbarkeit von Grundmodellen generativer KI können daher ebenso **kontraproduktiv** sein. Ein „one-size-fits-all“ Ansatz, wie man ihn bei der DSGVO gewählt hat, verbietet sich daher. Eine Risiko- und Anwendungsbezogene Regulierung, wie sie die EU im AI-Act vorsieht, dürfte grundsätzlich der bessere Ansatz sein.

Grenzen und Bürgerrechte

Da der Einsatz von *Predictive Analytics* im Bereich von Recht und Justiz die weitreichendsten Fragen ethischer Vertretbarkeit und möglicher Voreingenommenheit aufwirft, sollte die Debatte über Grenzen des Einsatzes solcher KI zu Zwecken von Strafverfolgung, Prävention oder Justiz sicherlich vorrangig geführt werden. Denn im Gegensatz etwa zum Urheberrecht können verlorene Bürgerrechte nicht nachträglich finanziell kompensiert werden und können Daten- und Verbraucherschutz die Bürgerinnen und Bürger zwar vor Unternehmen, nicht aber vor einem übergriffigen Staat schützen.

Dreiklang digitaler Transformation

Lukas von Kohout ist Alumni der Friedrich-Ebert-Stiftung Studienförderung und aktuell Vice President Digital Performance im Technologiebereich der Deutschen Telekom, hier arbeitet er daran, in der Organisation Wertschöpfung, Fähigkeiten und Haltung an digitalen Chancen auszurichten. Der Beitrag gibt seine persönliche Meinung wieder.

Wie nutzt die Deutsche Telekom Künstliche Intelligenz?

Weltweit nutzt die Telekom AI explorativ in drei Bereichen mit dem Ziel, ihre Stakeholder zu überzeugen:

1. in der **individuellen Nutzung offener Tools** durch Mitarbeitende in der alltäglichen Arbeit, ob für Kreativität (z.B. Bildwelten mit Midjourney) oder Produktivität (z.B. Programmieren und Recherchieren mit ChatGPT)
2. in der maßgeschneiderten **Einbettung gekaufter Lösungen**
 - für ihre Geschäftskunden (z.B. komplett abgeschottete private Instanzen von ChatGPT) oder
 - intern für betriebliche Optimierung durch Mustererkennung in Massendaten (z.B. in der Cyberabwehr >1 Mrd. Datenpunkte pro Tag) oder
 - bei Chatbots für Kundendialoge und in der Personalverwaltung
3. durch **echte Eigenentwicklung**, um die Wertschöpfung auf die nächste Ebene zu entwickeln

Beispiele für Einbettung und Eigenentwicklung

a) Kunden ansprechen durch UserExperience „à la Messenger“ und 24/7 Verfügbarkeit

„Frag Magenta“ ist eine Chatanwendung, die mit Hilfe von AI Kunden im Dialog hilft. Sie ist in der Lage, Fragen zu Rechnungen, Verträgen, Tarifen und vielem mehr zu beantworten. Darüber hinaus kann sie Kunden bei der Lösung von Problemen zu unterstützen, indem sie ihnen Schritt-für-Schritt-Anleitungen gibt. Die Anwendung wird seit 2016 stetig weiterentwickelt und ist über diverse Kanäle zugänglich: Website, Magenta App, WhatsApp und Telefon.¹



b) Wertschöpfung disrupten durch eigenes Large Language Model mit internationalen Peers

Gemeinsam mit asiatischen Telekommunikationsunternehmen wurde eine Allianz gegründet, die durch die Entwicklung eines branchenspezifischen LLM innovative und tiefgreifende AI-Anwendungen möglich machen wird. Der Ansatz ist von dem Glauben getragen, dass mit höherer sprachlicher Spezifität, internationaler Zusammenarbeit und Skalierung durch Partnerressourcen die Wettbewerbsfähigkeit als führendes europäisches Telekommunikationsunternehmen weiter ausgebaut werden kann.²

1 <https://www.telekom.de/hilfe/frag-magenta?samChecked=true#:~:text=Der%20digitale%20Assistent%20Frag%20Magenta,%2C%20Auftrag%2C%20Vertrag%20oder%20Umzug> [03.01.24]

2 <https://www.telekom.com/de/medien/medieninformationen/detail/global-telco-ai-alliance-gegruendet-1045150> [03.01.24]

Schützen – verantwortlich und transparent

Die Deutsche Telekom hat eine eigene Digital Ethics Policy, die den Umgang mit AI in Entwicklung, Betrieb und Nutzung regelt. Die Policy basiert auf drei Säulen: **Verantwortungsbewusstsein, Transparenz und Kontrolle**.³ Die Policy ergänzt die übergreifende Governance-Struktur für Datenschutz und Sicherheit.⁴ Der Vorstand wird vom unabhängigen Datenschutzbeirat, der von führenden Datenschutzexperten aus Wissenschaft, Wirtschaft, Politik und unabhängigen Organisationen gebildet wird, beraten.⁵ Neben Normen, Strukturen und Prozessen sind kontinuierliche Awareness- und Skill-Trainings für die Mitarbeitenden zentraler Bestandteil für die effektive Verantwortungskultur.

Regulieren – iterativ und dialogisch

Der AI Act ist ein guter Startpunkt auf der langen Reise die Chancen und Risiken von AI für unsere Gesellschaft optimal

zu balancieren. Positiv ist, dass der Rahmen a) relativ früh gesetzt wird, b) auf Anwendungsfälle abzielt und c) eine Einzelfallprüfung möglich macht. Für die weitere Entwicklung und Umsetzung sind eine Schärfung der Definitionen und eine effiziente Ausgestaltung der Einzelfallprüfung im Dialog mit der Wirtschaft nötig. Um die globale Wettbewerbsfähigkeit der EU in diesem Zukunftsfeld zu erhalten, ist, nachdem man nun den klarsten und strengsten Rahmen gesetzt hat, die nächste Aufgabe, diesen mit den USA und Asien so zu harmonisieren, dass mindestens ein gemeinsames Verständnis der Begrifflichkeiten und Gegenstände herrscht.

3 <https://www.telekom.com/en/company/digital-responsibility/details/artificial-intelligence-ai-guideline-524366> [04.01.24]

4 <https://www.telekom.com/en/company/data-privacy-and-security> [04.01.24]

5 <https://www.telekom.com/de/konzern/datenschutz-und-sicherheit/governance-datenschutz/datenschutzbeirat> [04.01.24]

Faire KI – 5 goldene Regeln gegen Diskriminierung

Dr. Nathalie Sellmeijer, heute bei VW im Bereich Autonomes Fahren tätig, hat sich in ihrer Promotion mit der Problematik der Diskriminierung in KI-Modellen auseinandergesetzt. In ihrer Forschung identifizierte sie Ursachen für Bias und erforschte programmatische Ansätze, um Diskriminierung aus den Modellen zu eliminieren. Hier teilt sie ihre Erkenntnisse in Bezug auf diskriminierungsfreie KI-Modelle.

Die Zahl der Industriezweige und Anwendungen, in denen KI-Modelle verwendet werden, steigt stetig. Das Europäische Parlament will nun im Rahmen des EU AI Act u.a. sicherstellen, dass KI-Systeme nicht diskriminierend sind. Was auf den ersten Blick klar erscheint, ist bei näherem Hinsehen nicht einfach zu bewerten.

Im Herzen der Maschine: Wie kann ein Modell diskriminieren?

Um zu verstehen, wie man faire KI-Modelle baut und evaluiert steht zu Beginn die Frage, warum diese Art von Modellen überhaupt das Risiko der Diskriminierung mit sich bringen. Hier gibt es verschiedene Ursachen (u.A.):

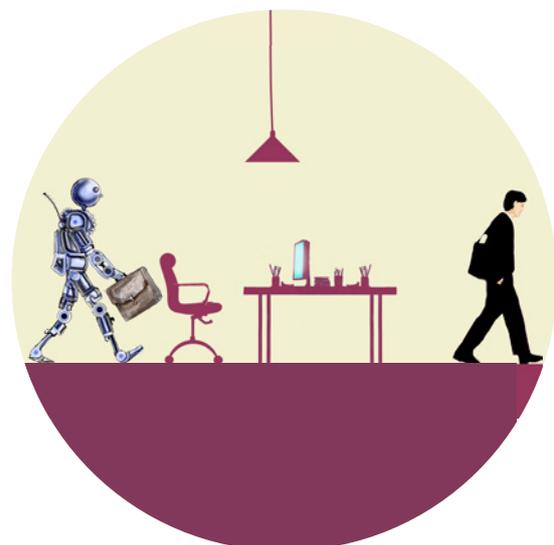
- der sogenannte Historical Bias. Der Begriff beschreibt die Problematik, dass unsere Realität bereits diskriminierend ist. Daten aus der Realität enthalten diskriminierende Strukturen, die dann beim Training eines Modells übernommen werden.
- Ähnliche Problematiken bestehen, wenn keine Daten von Minderheiten vorliegen und nur mit Daten der Mehrheitsgesellschaft trainiert wird (Representational Bias).
- Andere Ursachen für diskriminierende Modelle sind der Einsatz von Modellen in einem unpassenden Kontext (Deployment Bias)
- oder der Einsatz von one-size-fits-all-Lösungen, die für verschiedene Anwendungsfälle genutzt werden und dabei für keinen so richtig passen (Aggregation Bias).

Die Macht der Codes: Konsequenzen eines diskriminierenden KI-Modells

- Die Konsequenzen von diskriminierenden KI-Modellen können weitreichend sein:

- ein KI-gesteuertes Kredit scoring gibt Migrant_innen schlechtere Scores;
- eine KI-gesteuerte Vorauswahl von Bewerbungen sortiert Frauen für Führungspositionen aus
- oder die Gesichtserkennung einer App erkennt nur helle Hautfarben.

Erschwerend kommt hinzu, dass die Entscheidungen von Maschinen weniger hinterfragt werden als die von Menschen. Insbesondere, wenn Anwender_innen nicht die Ersteller des Modells sind, sondern das Modell beispielsweise von einem Zulieferer gekauft haben, werden die Entscheidungen des Systems nicht in Frage gestellt.



Fair oder nicht: Wie misst man Diskriminierung?

Für die Feststellung, ob ein KI-System diskriminierend ist, gibt es keine einheitliche Messgröße. Fairness-Metriken hingegen gibt es einige – alle haben ihre Daseinsberechtigung, und doch ist die Uneinheitlichkeit dabei problematisch: bei der Anwendung verschiedener Fairness-Metriken können unterschiedliche Ergebnisse herauskommen. Manchmal gilt ein Modell als diskriminierend, manchmal gilt dasselbe Modell als fair. Um herauszufinden, welche Metrik für den Anwendungsfall genutzt werden kann, muss erstmal verstanden werden, wie die Berechnung der Metrik funktioniert – anschließend kann man entscheiden.

5 goldene Regeln für faire KI-Modelle

1. Kritisches Hinterfragen der verwendeten Daten

Die Trainingsdaten von Modellen sollten immer hinsichtlich ihres Diskriminierungspotentials und ihrer Aussagekraft bewertet werden. Woher kommen die Daten? Welche Gesellschaft bilde ich mit meinen Daten ab? Sind alle demographischen Gruppen vertreten, oder sind nur Daten der Mehrheitsgesellschaft enthalten?

2. Gezielte Auswahl der Metrik

Bei der Evaluation eines Modells ist es wichtig, nicht mit der erstbesten Metrik oder den schönsten Ergebnissen zufrieden zu sein. Die Vielzahl der Fairness-Metriken bilden verschiedene Sichtweisen auf das Diskriminierungspotential ab und sollten sorgfältig ausgewählt werden. Um zu verstehen, welche Metrik zu meinem Anwendungsfall passt, hilft nur eines: die Berechnung dahinter verstehen.

3. Prüfen, welche Gruppen von Diskriminierung betroffen sein könnten

Weiterhin ist es wichtig zu überprüfen, auf welche demographischen Gruppen hin das Modell auf seine Diskriminierung evaluiert werden soll. Oftmals fällt sofort die Diskriminierung nach Geschlecht oder Ethnie ein, andere demographische Gruppen können jedoch genauso von Diskriminierung betroffen sein. Ein Beispiel dafür sind KI-gesteuerte Lernanwendungen in der Schule. Neben dem Geschlecht und des Migrationshintergrunds ist in der Schule für allem die soziale Herkunft ein Prädiktor für Erfolg. Ein KI-Modell muss auch dahingehend evaluiert werden, ob es Schüler_innen aus bestimmten Stadtvierteln benachteiligt. Für die Evaluierung eines Modells ist es daher wichtig zu sehen, welche Menschen durch das System diskriminiert werden könnten.

4. Transparenz schaffen durch Explainable AI

Die Frage wie ein KI-Modell zu seinem Ergebnis kommt, kann durch Explainable AI beantwortet werden. Explainable AI bietet die Möglichkeit das „Innere“ eines Modells sichtbar zu machen und die Blackbox zu durchleuchten. Das gibt die Möglichkeit zur Auswertung und schafft Transparenz über die Ergebniserstellung.

5. Fachgerechte Verwendung von Modellen

Bei der Verwendung von KI-Systemen muss immer vorab geprüft werden, ob das Modell zu den Zwecken passt, für die es eingesetzt werden soll. Dazu reicht es nicht aus, ein Modell von einem Zulieferer zu kaufen und einzubinden. Es muss klar sein, wie es eingesetzt werden soll, welche Fragestellungen es beantworten kann und welche nicht, und in welchem Kontext es trainiert worden ist.

Das Bedürfnis nach Regulierung von Künstlicher Intelligenz ist verständlich. Auch der Wunsch nach evaluierten, transparenten und diskriminierungsfreien KI-Systemen ist nachvollziehbar. Gleichwohl zeigt sich bei der näheren Betrachtung der Handlungsempfehlungen, wie komplex das Thema der diskriminierungsfreien KI-Modellen werden kann. Hier gibt es kein Patentrezept, denn die Unterschiedlichkeit und Anwendungsfälle von Künstlicher Intelligenz sind enorm. Eine Regulierung von KI-Systemen sollte das berücksichtigen.

Weitere Literatur:

Suresh, H., & Guttag, J. (2021). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. In EAAMO '21, Equity and Access in Algorithms, Mechanisms, and Optimization. Association for Computing Machinery. doi.org/10.1145/3465416.3483305

Verma, S., & Rubin, J. (2018). Fairness definitions explained. In Proceedings of the International Workshop on Software Fairness. ACM. doi.org/10.1145/3194770.3194776

Rzepka, N., Simbeck, K., Pinkwart, N. (2023). Learning Analytics & Diskriminierung. In Schiefner-Rohs, M., Hofhues, S., & Breiter, A. Digitale Gesellschaft Band 59. Transcript Verlag · 1. Okt. 2023

Bildquellen: Titel: Marei John-Ohnesorg, Runway, S. 3 Marei John-Ohnesorg, ChatGPT, S.4 picture alliance // Westend61 / Gray Waters

Alle Veröffentlichungen finden Sie unter:
<https://www.managerkreis.de/publikationen>

Impressum: © Friedrich-Ebert-Stiftung | Herausgeberin: Friedrich-Ebert-Stiftung e.V. | Godesberger Allee 149 | 53175 Bonn | Deutschland

Verantwortlich: Managerkreis der Friedrich-Ebert-Stiftung | Hiroshimstraße 17 | 10785 Berlin

www.managerkreis.de | Februar 2024

Inhaltliche Verantwortung: Marei John-Ohnesorg | Redaktion: Nicole Castillo | Kontakt: managerkreis@fes.de, 030 26 935 7051

Die in dieser Publikation zum Ausdruck gebrachten Ansichten sind nicht notwendigerweise die der Friedrich-Ebert-Stiftung e.V. Eine gewerbliche Nutzung der von der Friedrich-Ebert-Stiftung (FES) herausgegebenen Medien ist ohne schriftliche Zustimmung durch die FES nicht gestattet. Publikationen der Friedrich-Ebert-Stiftung dürfen nicht für Wahlkampfzwecke verwendet werden.

