www.jusletter-it.eu

Chiara Zengerer

Nachvollziehbarkeit von Entscheiden – eine Gegenüberstellung von menschlichen Richterinnen und Richtern und künstlicher Intelligenz

Die Nachvollziehbarkeit von Ergebnissen stellt eine der grössten Herausforderungen beim Einsatz von künstlicher Intelligenz (KI) in der Justiz dar. Der vorliegende Beitrag untersucht, wie Entscheide einer KI im Vergleich zu Entscheiden von menschlichen Richterinnen und Richtern hinsichtlich ihrer Nachvollziehbarkeit zu beurteilen sind. Die Analyse diverser Beiträge zeigt, dass menschliche Entscheide nicht per se eine bessere Nachvollziehbarkeit bieten und dass den Gefahren einer KI teilweise mit Schutzmassnahmen und Kontrollen entgegengewirkt werden kann. Die grössten Hürden für eine KI bilden demgegenüber das fehlende Kontextwissen, die Unfähigkeit zur kritischen Reflexion sowie die Angabe der richtigen Gründe für einen Entscheid.

Beitragsart: Beiträge
Rechtsgebiete: LegalTech

Zitiervorschlag: Chiara Zengerer, Nachvollziehbarkeit von Entscheiden – eine Gegenüberstellung von menschlichen Richterinnen und Richtern und künstlicher Intelligenz, in: Jusletter IT 20. Juli 2023

ISSN 1664-848X, jusletter-it.weblaw.ch, Weblaw AG, info@weblaw.ch, T +41 31 380 57 77 EDITIONS WEBLAW

Inhaltsübersicht

1. Einleitung

[1] Gemäss den Leitlinien «Künstliche Intelligenz» für den Bund stellt die Nachvollziehbarkeit von Ergebnissen eine der grössten Herausforderungen beim Einsatz von künstlicher Intelligenz (KI) dar. Das eidgenössische Departement für Wirtschaft, Bildung und Forschung legt fest, dass die Funktionsweise, der Zweck sowie die verwendeten Datensätze zum Training oder zur Entwicklung der KI in verantwortungsvoller und rechtskonformer Weise offengelegt werden sollen, damit die Entscheidungsprozesse für direkt und indirekt Betroffene nachvollziehbar und die Wirkungsweise für Fachleute überprüfbar sind.[1] Die inhaltliche Nachvollziehbarkeit hat in der Justiz eine besonders wichtige Bedeutung, da hoheitliche Entscheide den rechtsstaatlichen Grundsätzen genügen müssen.[2]

[2] Angesichts dieser fundamentalen Bedeutung und den dazu häufig vorgebrachten Einwänden gegen den Einsatz von KI drängt sich die Frage auf, wie berechtigt diese Einwände im Vergleich zu menschlichen Entscheiden sind. Zu diesem Zweck wird die Frage nach der technischen Machbarkeit von automatisierten Entscheiden sowie die Frage nach der positivrechtlichen Zulässigkeit beiseitegelassen. Der vorliegende Beitrag befasst sich allein mit der Nachvollziehbarkeit und stellt dabei konkret die Frage: Wie sind Entscheide einer KI im Vergleich zu Entscheiden von menschlichen Richterinnen und Richtern hinsichtlich ihrer Nachvollziebarkeit zu beurteilen?

[3] Zur Einführung in die Thematik wird der Einsatz von KI zur Automatisierung von rechtlichen Entscheiden erläutert. Anschliessend werden im dritten Kapitel die Problembereiche bei Entscheiden von KI aufgezeigt, bevor im vierten Kapitel die untersuchten Problembereiche mit menschlichen Entscheiden verglichen werden. Abschliessend werden die Ergebnisse zusammengefasst und diskutiert.

2. Automatisierung von rechtlichen Entscheiden

[4] Der vorliegende Beitrag legt den Fokus auf den Einsatz von KI zur Automatisierung von rechtlichen Entscheiden. Eine Automatisierung der Entscheidfindung bedeutet, dass die Entscheide

1 Eidgenössisches Departement für Wirtschaft, Bildung und Forschung WBF, Leitlinien «Künstliche Intelligenz» für den Bund, 2020, https://www.sbfi.admin.ch/dam/sbfi/de/dokumente/2020/11/leitlinie_ki.pdf.download.pdf.

2 Nink David, Justiz und Algorithmen, Über die Schwächen menschlicher Entscheidungsfindung und die Möglichkeiten neuer Technologien in der Rechtsprechung, Berlin 2021, S. 334.

autonom und somit ausschliesslich durch ein auf KI basierendes System getroffen werden.[3] Hierfür besonders geeignet sind Verfahren, welche die Rechtsfolge an eindeutig messbare Kriterien binden.[4] Bei Entscheiden, die im Einzelfall abgewogen werden müssen, besteht die Herausforderung vor allem in der Erkennung vorhandener Spielräume und ihrer ordnungsgemässen Nutzung.[5]

[5] Für die Automatisierung von rechtlichen Entscheiden sind vor allem Systeme maschinellen Lernens von Bedeutung. Die Funktionsweise solcher Systeme besteht darin, durch das Trainieren mit grösseren Mengen bestehender gerichtlicher Entscheide die relevanten Entscheidungsfaktoren zu erkennen und zu erlernen. Im Idealfall soll nach der Trainingsphase die Anwendung auch auf unbekannte Fälle möglich sein, indem das System auf das trainierte Entscheidungsmodell zurückgreift.[6]

[6] Die nachfolgend diskutierten Aspekte betreffen nur den Einsatz von KI in zivilrechtlichen Prozessen. In Strafprozessen wird die Datenerhebung durch Aussage- und Mitwirkungsverweigerungsrechte zusätzlich erschwert und die Festlegung eines Strafmasses beinhaltet weitere Faktoren wie Gerechtigkeitsempfinden und Resozialisierungschancen, die nur schwer algorithmisch erfasst werden können.[7] Des Weiteren ist davon auszugehen, dass menschliche Richterinnen und Richter mindestens zur Feststellung des Sachverhalts unabdingbar bleiben, da technische Systeme bislang auf den Daten-Input durch Menschen angewiesen sind.[8] Die Wahrheitsfindung durch eine KI ist ungleich schwieriger als die blosse Interpretation von Gesetzen auf der Basis von als wahr angenommenen Fakten.[9]

3. Problembereiche bei Entscheiden von künstlicher Intelligenz

[7] Das dritte Kapitel dieses Beitrags befasst sich mit den Einwänden hinsichtlich Nachvollziehbarkeit von Entscheiden, die durch eine KI gefällt wurden. Zu diesem Zweck werden die am häufigsten erwähnten Problembereiche beleuchtet.

3.1. Wertungen und Abwägungen

[8] Gerichtliche Entscheide beruhen regelmässig auf einer Abwägung der betroffenen Rechtsgüter auf dem Wege praktischer Konkordanz sowie auf Wertungs-, Beurteilungs- und Ermessensspielräumen. In diesen Fällen existieren verschiedene Begründungs- und Entscheidungsoptionen, die aus Sicht des Rechts alle «richtig» sind, aber trotzdem unterschiedlich gut vertretbar sein kön-

3 Von Lucke Jörn/Etscheid Jan, Wie Ansätze künstlicher Intelligenz die öffentliche Verwaltung und die Justiz verändern könnten, in: Jusletter IT 21. Dezember 2020, S. 258.

4 Von Lucke/Etscheid, S. 263.

5 Von Lucke/Etscheid, S. 259 f.

6 Zum Ganzen: Dreyer Stephan/Schmees Johannes, Künstliche Intelligenz als Richter?, Wo keine Trainingsdaten, da kein Richter – Hindernisse, Risiken und Chancen der Automatisierung gerichtlicher Entscheidungen, Computer und Recht 11/35 2019, N1.

7 Puppe Frank, Gesellschaftliche Perspektiven einer fachspezifischen KI für automatisierte Entscheidungen, Informatik Spektrum 2/45 2022, S. 92.

8 Nink, S. 178.

9 Puppe, S. 92.

nen.[10] Bei solchen Entscheidungsspielräumen ist die Prüfung der Rationalität des Entscheidungsverfahrens sowie der Wissensgrundlagen für die Herleitung eines Entscheids besonders wichtig, was eine systemische Grenze für automatisierte Entscheide darstellen kann.[11]

[9] Greco sieht hingegen keinen Grund, der bereits im Vorhinein ausschliesst, einer künstlichen Intelligenz das juristische Werten beizubringen. Zu diesem Zweck könnte nach einer ersten Trainingsrunde mit bisherigen gerichtlichen Entscheiden in einer zweiten Kontrollrunde überprüft werden, ob die Ergebnisse der KI denen der Menschen entsprechen. Anschliessend würden die begangenen Fehler als Lernmaterial eingearbeitet werden. Nicht einmal Rechtsfortbildung liegt zwingend ausserhalb des Möglichen, da Rechtsfortbildung praktisch immer darin besteht, aus in anderen Rechtsgebieten bereits anerkannten Prämissen Folgen für einen anderen Sachverhalt abzuleiten und nicht darin, etwas wahrhaft Neues zu erschaffen.[12] Ist eine Rechtsfortbildung allerdings aufgrund eines Wertewandels in der Gesellschaft angezeigt, fehlt es am menschlichen Faktor, der die Anpassung der Rechtsprechung auslösen könnte.[13]

3.2. Unzureichende Begründungen

[10] Die Begründung ist im Sinne einer rationalitätsorientierten Rechtsauffassung für die Rechtmässigkeit des Entscheids konstitutiv. Die Begründung muss einerseits an sich nachvollziehbar sein und andererseits im Verhältnis zum Ergebnis nachvollziehbar sein. Bereits die erste Qualitätsanforderung erscheint nach dem heutigen Stand der Technik schwierig zu erreichen, da Algorithmen häufig Ergebnisse hervorbringen, ohne Auskunft darüber geben zu können, wie diese zustande gekommen sind. Einer KI die Fähigkeit beizubringen, nicht nur den richtigen Entscheid zu treffen, sondern auch die richtigen Gründe dafür anzugeben, erscheint jedoch nicht im Vorhinein als unmöglich, zumal juristische Texte häufig Textbausteine enthalten.[14]

[11] Für das Verhältnis der Begründung zum Ergebnis muss zunächst zwischen aufrichtigen Begründungen und Rationalisierungen unterschieden werden. Aufrichtige Begründungen stellen das Ideal dar, bei dem die Gründe für einen Entscheid auch die Motive für den Entscheid bilden. Als Rationalisierungen kritisiert werden Entscheide, die aus Motiven gefällt werden, die nicht zu Gründen werden können oder dürfen. Die Aufdeckung von Rationalisierungen ist allerdings bereits bei menschlichen Richterinnen und Richtern kaum möglich, da nicht in ihre Köpfe hineingeschaut werden kann. Die Unterscheidbarkeit ist sogar aussichtsreicher bei einer KI, da Programme entworfen werden könnten, die zusätzlich jeden Arbeitsschritt dokumentieren.[15] Das Nachvollziehen einer solchen Dokumentation dürfte jedoch mit hohem Aufwand verbunden sein und das Ergebnis nur bei Routineentscheiden in typischen Fällen befriedigend ausfallen.[16]

---

10 Zum Ganzen: Dreyer/Schmees, N 17.

11 Dreyer/Schmees, N 18.

12 Zum Ganzen: Greco Luis, Richterliche Macht ohne richterliche Verantwortung – Warum es den Roboter-Richter nicht geben darf, RW Rechtswissenschaft 1/11 2020, S. 37 f.

13 Wagner Jens, Legal Tech und Legal Robots, Wiesbaden 2020, S. 94.

14 Zum Ganzen: Greco, S. 42 f.

15 Zum Ganzen: Greco, S. 44 f.

16 Gless Sabine/Wohlers Wolfgang, Subsumtionsautomat 2.0, Künstliche Intelligenz statt menschlicher Richter?, in: Böse Martin/Schumann Kay H./Toepel Friedrich (Hrsg.), Festschrift für Urs Kindhäuser zum 70. Geburtstag, Baden 2019, S. 159 f.

[12] Erstellt die KI ihren Entscheid auf der Basis einer Vielzahl historischer Entscheide mittels Mustererkennung, wird es im Normalfall keine eigenen, am Gesetz orientierten Begründungsschritte angeben können. Allerdings könnte die KI den historischen Parallelfall finden und die dortigen Begründungen mitliefern.[17] Im Allgemeinen besteht jedoch die Gefahr von deutlich schematischeren und weniger ausdifferenzierten Entscheiden, da menschliche Richterinnen und Richter aufgrund der Lebenserfahrung Nuancen besser erkennen.[18]

3.3. Gefahr von Diskriminierungen

[13] Hinsichtlich der erhofften Objektivität von Entscheiden einer KI muss beachtet werden, dass ihre Entscheide nur so gut respektive so rational oder gerecht sein können wie die zugrundeliegenden Trainingsdaten.[19] Das bedeutet, dass KI nicht von sich aus die notwendigen Schutzrechte für Schwächere oder Minderheiten gewährleistet, sondern dies Aufgabe des Staates bleibt.[20] Dieser Effekt wird zusätzlich verstärkt, wenn die gewonnenen Ergebnisse der KI ihrerseits als Grundlage für künftige Ergebnisse verwendet werden.[21]

[14] Das Argument der Diskriminierung kann jedoch lediglich eine Mahnung darstellen, die Möglichkeit der Diskriminierung ernst zu nehmen und ihr mit angemessenen technischen und organisatorischen Schutzmassnahmen und Kontrollen zu begegnen, da die Gefahr der Diskriminierung auch bei menschlichen Richterinnen und Richtern nicht ausgeschlossen werden kann. Es ist eher fraglich, ob sie ihnen nicht in einem noch stärkeren Masse ausgeliefert sind. Bei von einer KI gewonnenen Entscheiden kann zumindest mit Sicherheit ausgeschlossen werden, dass diese Ergebnisse planmässig herbeigeführt wurden.[22]

4. Vergleich zu menschlichen Entscheiden

[15] In diesem Kapitel werden die oben beleuchteten Problembereiche bei Entscheiden einer KI mit den Problembereichen bei menschlichen Entscheiden verglichen. Dazu werden ebenfalls die am häufigsten vorgebrachten potenziellen Problemfelder untersucht und einer KI gegenübergestellt.

4.1. Einstellungen und Befindlichkeiten

[16] Wo Spielraum für Wertungen besteht, tragen menschliche Richterinnen und Richter immer auch persönliche Elemente in die Wahrnehmung des Sachverhalts hinein.[23] Unser Rechtssystem ist offen und bisweilen auch angewiesen auf den Einbezug des gesellschaftlichen Kontexts und

17 Zum Ganzen: Wagner, S. 91.

18 Wagner, S. 93 f.

19 Nink, S. 167.

20 Nink, S. 168.

21 Greco, S. 39.

22 Zum Ganzen: Greco, S. 40 f.

23 Nink, S. 41 f.

der aktuellen Wertmassstäbe in die Auslegung und Anwendung von Normen.[24] Ein Entscheid stützt sich deshalb typischerweise immer auch auf intuitiv-wertende Erkenntnisse. Allerdings dürfen einem Entscheid nicht diese «Gefühle», sondern nur die Fakten sowie ihre Gewichtung zugrunde gelegt werden, da ansonsten die Gefahr der Willkür durch übermässige Subjektivität und strukturell bedingte Ungleichbehandlung gleicher Sachverhalte entsteht.[25]

[17] Im Gegenzug dazu würde eine KI unter gleichen Bedingungen vorhersagbar die gleichen Entscheide treffen, frei von Voreinstellungen und aktuellen Befindlichkeiten. Damit ginge eine grössere Transparenz und Standardisierung einher. Gleichzeitig besteht jedoch der Nachteil, dass von subtilen Besonderheiten jedes Einzelfalles abstrahiert werden muss, da abstrakte Begriffe zwar immer weiter konkretisiert werden können, aber nicht alle Konstellationen vorhersehbar sind. Daher müsste eine Berufung an eine höhere Instanz möglich sein, bei welcher die Entscheide von menschlichen Richterinnen und Richtern gefällt werden.[26]

4.2. Unabhängigkeit und Unparteilichkeit

[18] Die Fähigkeit des Menschen, eine ganzheitliche Sichtweise einzunehmen, öffnet auf der Kehrseite ein mögliches Einfallstor für subjektive Einschätzungen, die sich vielleicht nur aus der individuellen Biografie oder einem informellen Entscheidungsprogramm erklären lassen.[27] In diesem Zusammenhang könnten Maschinen in viel grösserem Umfang unabhängig und unparteilich sein als Menschen, sofern das Programm immer wieder korrigierend kalibriert wird, um den bereits im Trainingsdatensatz enthaltenen Vorurteilen entgegen zu wirken.[28]

[19] Bei menschlichen Richterinnen und Richtern kann nie ausgeschlossen werden, dass sie befangen sind, sich von aussen unter Druck setzen lassen oder sonst einen unsachlichen Entscheid treffen. Beim Einsatz von KI entstehen allerdings dadurch, dass sich schon ein einzelner unzulässiger Eingriff auf eine Vielzahl von Entscheiden auswirken kann, neue Angriffspotentiale. Ebenfalls könnten zentral entwickelte und eingesetzte Algorithmen von politischen Akteuren missbraucht werden.[29]

4.3. Kontextwissen und Reflexion

[20] Von menschlichen Richterinnen und Richtern wird erwartet, dass sie bei einem Entscheid das Ergebnis in einer holistischen Gesamtbewertung hinterfragen und so reflektiert begründen, dass es intersubjektiv nachvollziehbar ist. Dazu ist die Fähigkeit zur offenen und kritischen Reflexion des eigenen Entscheids nötig. Gless/Wohlers sind der Meinung, dass nur Menschen in der Lage sind, ihren Entscheid adäquat zu begründen, Korrekturbedarf zu erkennen und so dem Berufsethos gerecht zu werden.[30]

---

24 Nink, S. 42.

25 Zum Ganzen: Nink, S. 43 f.

26 Zum Ganzen: Puppe, S. 93.

27 Gless/Wohlers, S. 160 f.

28 Gless/Wohlers, S. 162 f.

29 Zum Ganzen: Wagner, S. 93.

30 Zum Ganzen: Gless/Wohlers, S. 159.

[21] Für KI besteht auf der semantischen Ebene zusätzlich die Herausforderung, dass teilweise gleiche Sinngehalte mit unterschiedlichen Wörtern und teilweise unterschiedliche Sinngehalte mit ähnlichen oder gleichen Wörtern beschrieben werden können. Um die Bedeutung dieser unterschiedlichen Sinngehalte und damit auch den Kontext selbst bei simplen sozialen Komplexen zu verstehen, ist ein umfassendes Weltverständnis nötig. Verfahren maschinellen Lernens agieren jedoch bei ihren Wahrscheinlichkeitsrechnungen primär anhand identifizierter Korrelationen von mehreren Variablen. Sie verfügen damit nicht über das erforderliche Kontextwissen oder moralische Bewusstsein.[31]

5. Schlussfolgerungen

[22] Die Analyse diverser Beiträge zeigt, dass hinsichtlich der Nachvollziehbarkeit beim Einsatz von KI zur automatisierten Entscheidfindung im Vergleich zu menschlichen Entscheiden Uneinigkeit herrscht. Bei der Gegenüberstellung von KI und menschlichen Richterinnen und Richtern wird jedoch deutlich, dass menschliche Entscheide nicht per se eine bessere Nachvollziehbarkeit im Bereich der vorgebrachten Einwände gegen KI bieten.

[23] Der Gefahr der Diskriminierung durch eine KI kann mit angemessenen technischen und organisatorischen Schutzmassnahmen und Kontrollen sowie der sorgfältigen Auswahl der Daten begegnet werden. Zudem kann mit Sicherheit ausgeschlossen werden, dass solche Ergebnisse planmässig herbeigeführt wurden, was bei menschlichen Richterinnen und Richtern nur gehofft werden kann. Ebenso könnte eine KI in viel grösserem Masse als Menschen unabhängig und unparteilich sein.

[24] Auf der anderen Seite stehen das fehlende Kontextwissen einer KI und die Unfähigkeit zur kritischen Reflexion des eigenen Entscheids im Sinne einer holistischen Gesamtbewertung. Bei Wertungs-, Beurteilungs- und Ermessensspielräumen ist die Prüfung der Rationalität des Entscheidungsverfahrens und der Wissensgrundlagen für die Herleitung eines Entscheids besonders wichtig. Dies kann eine systemische Grenze für automatisierte Entscheide darstellen, solange einer KI das juristische Werten nicht beigebracht werden kann.

[25] Da die Begründung für die Rechtmässigkeit eines Entscheids konstitutiv ist, muss einer KI die Fähigkeit beigebracht werden, die richtigen Gründe für einen Entscheid anzugeben. Selbst wenn es gelingt, ein Programm zu entwerfen, das jeden Arbeitsschritt dokumentiert, dürfte das Nachvollziehen einer solchen Dokumentation mit hohem Aufwand verbunden sein. Im Allgemeinen besteht die Gefahr von deutlich schematischeren und weniger ausdifferenzierten Entscheiden.

[26] Die fortschreitende Digitalisierung in den Gerichten bietet die Grundlage für eine mögliche technische Umsetzung. Neben der vorliegend beiseitegelassenen Frage nach der technischen Machbarkeit von automatisierten Entscheiden sowie die Frage nach der positivrechtlichen Zulässigkeit ist ein öffentlicher Diskurs hinsichtlich der Wünschbarkeit solcher Systeme sowie die Präzisierung der Anforderungen hinsichtlich der Nachvollziehbarkeit in Bezug auf KI unabdingbar.

31 Zum Ganzen: Dreyer/Schmees, N 15.

Chiara Zengerer ist Bachelor-Studentin der Rechtswissenschaften an der Universität St. Gallen (HSG).