

Dominik Brodowski

Datengestützte Prognose justizieller Entscheidungen

Entwicklungsstufen, Gefahren, Potentiale

Abstract: Tools trying to predict the outcome of judicial decisions can achieve promising results. In this article, I will discuss their risks and potentials. Such systems can, if designed carefully to avoid perpetuating biases and petrifying the law, improve our legal systems and our understanding of the law.

I Hinführung

Angesichts der Datafizierung von Rechtsprechung und Rechtswissenschaft stehen immer größere Datenmengen zur Verfügung, die sich mit quantitativen Methoden auswerten lassen. Damit wird häufig die Erwartung verbunden, in Gerichtsentscheidungen enthaltene Muster herausdestillieren und für neue Sachverhalte den Ausgang des Verfahrens prognostizieren zu können. Die an Beliebigkeit grenzende Begriffsvielfalt, unter der datengestützte Systeme zur Prognose justizieller Entscheidungen beschrieben werden – von „quantitative legal analysis“ über „predictive justice“ hin zu „jurimetrics“¹ –, legt jedoch nahe, dass die Funktionsweise, die Einsatzfelder und daher auch die Gefahren und Potentiale derartiger Systeme noch nicht hinreichend ausdifferenziert werden.

Dieser Beitrag möchte dem entgegenwirken und hierzu zunächst verschiedene Arten solcher Systeme unterscheiden; insbesondere sind Prognosesysteme von Entscheidungssystemen, die eine richterliche Entscheidung ersetzen sollen,² a

1 Instrukтив *Lelieur/Blount/Cherqaoui/Bampasika*, General Report, 94(2) RIDP 29 f. (2023).

2 Zur Automatisierung justizieller Entscheidungen und damit verbundenen (Grundsatz-)Fragen siehe, statt vieler, *Bernzen*, Roboter als Richter? Zur Automatisierung der Rechtsprechung, RD 2023, 132; *Cassuto*, La justice à l'épreuve de sa prédictibilité, AJ Pénal 2017, 334; *Effer-Uhe*, Überlegungen zur Automatisierbarkeit der Rechtsanwendung, JZ 2023, 833; *Gless/Wohlers*, Subsumtionsautomat 2.0 – Künstliche Intelligenz statt menschlicher Richter?, in: Böse/Schumann/Toepel (Hrsg.), Festschrift für Kindhäuser, 2019, S. 147; *Greco*, Richterliche Macht ohne richterliche Verantwortung: Warum es den Roboter-Richter nicht geben darf, RW 2020, 29; *Lelieur/Blount/Cherqaoui/Bampasika* (Fn. 1), 94(2) RIDP 47 f. (2023); *Nink*, Justiz und Algorithmen – Über die Schwächen menschlicher Entscheidungsfindung und die Möglichkeiten neuer Technologien in der Rechtsprechung, 2021, insb. S. 177 ff., S. 260 ff.; *Rostalski*, Judex ex machina? Zum Einsatz neuer Technologien in der Rechtsfindung, in:

priori zu differenzieren (II.). Hierauf aufbauend lassen sich die Hürden beim Einsatz solcher Systeme, mit diesen verbundene Gefahren, aber auch deren Potentiale analysieren (III.). In diesem Beitrag nicht näher betrachtet werden sollen Systeme, die gelegentlich unter dieselben Schlagworte gefasst werden, jedoch andere Zwecke verfolgen oder lediglich Ausgangspunkt oder Hilfsmittel für die hier in den Blick genommenen Prognosesysteme sind. Damit meine ich etwa juristische Rechtersysteme,³ Berechnungsprogramme für Gefährdungsprognosen und Rückfallwahrscheinlichkeiten,⁴ Datenbanken juristischer Entscheidungen⁵ oder Programme zur Analyse digitaler Spuren.⁶

II Phänomenologie datengestützter Systeme zur Prognose justizieller Entscheidungen

Um verschiedene Entwicklungsstufen von datengestützten Systemen zur Prognose justizieller Entscheidungen (unten II.2.) und rechtlich bedeutsame Unterschiede zwischen solchen Systemen herauszuarbeiten, möchte ich zunächst vier Differenzierungskriterien aufzeigen (sogleich II.1.).

Hoven/Kudlich (Hrsg.), Digitalisierung und Strafverfahren, 2020, S. 263; *Zalnieriute/Bell*, Technology and the Judicial Role, in: Appleby/Lynch (Hrsg.), The Judge, the Judiciary and the Court: Individual, Collegial and Institutional Judicial Dynamics in Australia, 2021, S. 116, 120 f.; *Završnik*, Criminal justice, artificial intelligence systems, and human rights, 20 ERA Forum 568, 571 f. (2020).

³ Exemplarisch für aktuelle Entwicklungen: „Wolters Kluwer integriert GPT-generierte Zusammenfassungen in die juristische Recherche auf Wolters Kluwer Online“, <https://www.wolterskluwer.com/de-de/news/gpt-zusammenfassungen> (17.12.2023).

⁴ Siehe hierzu, statt vieler, *Eisbach/Heghmanns/Hertel*, Künstliche Intelligenz im Strafverfahren am Beispiel von Kriminalprognosen, *ZfStw* 2022, 489; *Kovalchuk/Karpinski/Banakh/Kasianchuk/Shevchuk/Zagorodna*, Prediction Machine Learning Models on Propensity Convicts to Criminal Recidivism, 14 Information 161 (2023).

⁵ Zur Diskussion über die Einführung von Strafzumessungsdatenbanken *Kaspar*, Digitalisierung als Chance für die Strafzumessung?, *KriPoZ* 2023, 1; *Kaspar/Höffler/Harrendorf*, Datenbanken, Online-Votings und künstliche Intelligenz – Perspektiven evidenzbasierter Strafzumessung im Zeitalter von „Legal Tech“, *NJ* 2020, 35; jeweils m.w.N.

⁶ Exemplarisch *Brodowski/Hartmann/Sorge*, Automatisierung in der Strafrechtspflege. Legal Tech, KI und eine „hybride Cloud“ im Einsatz gegen Kindesmissbrauch, *NJW* 2023, 583.

1 Differenzierungskriterien

a) Datengrundlage

Ein erstes Differenzierungskriterium sind die zur Konstruktion („Programmierung“ und ggf. „Training“) und bei Anwendung des Systems verwendeten Daten. Hinsichtlich dieser Datengrundlage ist eine große Datenvielfalt denkbar, die von der sensorgestützten Beobachtung der Außenwelt (etwa in Gestalt einer Videokamera zur Beobachtung des nonverbalen Aussageverhaltens eines Belastungszeugens) über die „elektronische Akte“ hin zu Rechtsprechungsdatenbanken reicht.

b) Technologie: Regelbasierte Systeme oder maschinelles Lernen

Systeme, die zur Prognose einen vordefinierten Algorithmus anwenden, lassen sich unterscheiden von Systemen, die Technologien des maschinellen Lernens in sogenannten neuronalen Netzen nutzen, um Muster in (Trainings-)Daten zu erkennen, und diese Muster(-erkennung) sodann auf neue Daten anwenden können.⁷ In diese letztgenannte Kategorie, die häufig als „Künstliche Intelligenz“ bezeichnet wird, fallen auch Systeme, die maschinell darauf trainiert sind, Sprache zu verarbeiten oder – basierend auf Eingabedaten („Prompts“) – sprachliche Rückmeldungen zu generieren (sogenannte generative Künstliche Intelligenz).

Diese Differenzierung ist für die rechtliche Würdigung solcher Systeme höchst bedeutsam: Nach aktuellem Forschungsstand lassen sich nämlich nur bei erstgenannten regelbasierten Systemen diejenigen Kriterien, die zu einem bestimmten Ergebnis geführt haben, anhand des Algorithmus hinreichend klar nachvollziehen. Bei Systemen, die auf mit maschinell trainierten neuronalen Netzen beruhen, ist dies – trotz aller Forschungsansätze im Hinblick auf eine „erklärbare Künstliche Intelligenz“⁸ – nicht oder zumindest nicht in gleichem Maße der Fall. Zu beachten ist allerdings, dass es neben diesen Extrema, etwa bei nicht auf neuronalen Netzen beruhendem maschinellen Lernen, auch Mischformen gibt.

⁷ Hierauf fokussiert *Sorge/Krüger*, Die Vorhersage von Gerichtsentscheidungen – Methodische und inhaltliche Ausgangsüberlegungen, BRJ Sonderausgabe 1/2021, 13, 13f. Zu technischen Grundlagen instruktiv *Alpaydin*, Maschinelles Lernen, 2019, insb. S. 3f.; *Bauckhage/Hübner/Hug/Paaß/Rüping*, Grundlagen des Maschinellen Lernens, in: Görz/Schmid/Braun (Hrsg.), Handbuch der Künstlichen Intelligenz, 2021, Kap. 11; *Bauckhage/Hübner/Hug/Paaß*, Tiefe neuronale Netze, in: Görz/Schmid/Braun (Hrsg.): Handbuch der Künstlichen Intelligenz, 2021, Kap. 12.

⁸ *Körner*, Nachvollziehbarkeit von KI-basierten Entscheidungen, in: Kaulartz/Braegelman (Hrsg.), Rechtshandbuch Artificial Intelligence und Machine Learning, 2020, Kap. 2.4 Rn. 24 f.

c) Vorhersage des Ergebnisses oder der Begründung?

Sodann bietet es sich an, Systeme zur Prognose justizieller Entscheidungen dahingehend zu differenzieren, ob sie darauf ausgerichtet sind, lediglich das Ergebnis einer justiziellen Entscheidung oder auch deren (tragende) Begründung zu prognostizieren.

In die erste Kategorie fallen Systeme, die für jedes mögliche Ergebnis einer juristischen Prüfung einen bestimmten Wahrscheinlichkeitswert berechnen, etwa „Abweisung der Klage: 30 %“ oder „Schuldig im Sinne der Anklage: 3 %“. Wird im ersten Fall dem Klageantrag entsprochen und im zweiten Fall der Angeklagte freigesprochen, so lässt sich zwar davon sprechen, dass das System das Ergebnis zutreffend prognostiziert hat. Dies bezieht sich jedoch ausschließlich auf eine Ergebnisrichtigkeit in demjenigen, eng begrenzten Sinne, dass das tatsächliche dem vorhergesagten Ergebnis entspricht, also auf eine Korrelation. Auf eine Kausalität oder, allgemeiner gesprochen, auf eine Begründungsrichtigkeit darf man hieraus jedoch nicht schließen.⁹ Weder eine Übereinstimmung zwischen Vorhersage und tatsächlicher Entscheidung noch eine Divergenz vermögen über die Tragfähigkeit, Stimmigkeit, Binnenrationalität oder gar Rechtmäßigkeit der Entscheidung etwas auszusagen. Dies ist – bzw. wäre – anders bei Systemen, welche Tatsachenfeststellung, argumentative Schritte und Subsumtionen nachzuahmen suchen, bis hin zum Denkmodell des „Subsumtionsautomaten“.

d) Bloße Vorhersage oder bereits Automatisierung?

Eine grundlegende Differenzierung ist schließlich dahingehend vorzunehmen, welche Folgerungen aus der Prognose gezogen werden: Am einen Ende des Spektrums stehen Systeme, die im Dienst der Wissenschaft lediglich eine Vorhersage treffen, ohne dass hieraus für das konkrete Verfahren rechtliche Schlüsse gezogen werden. Am anderen Ende sind Systeme zu verordnen, deren Prognoseergebnis unmittelbar in eine justizielle Entscheidung münden, also Systeme, die „automated judicial decision-making“ bzw. „algorithmic justice“ betreiben. Wenngleich der Fokus dieses Beitrags auf Prognose- und nicht auf Entscheidungssystemen liegt: Zwischen diesen Extrema befinden sich Systeme, die zwar nicht anstelle eines

⁹ Statt vieler *Lelieur/Blount/Cherqaoui/Bampasika* (Fn. 1), 94(2) RIDP 48 f. (2023); *Rühl/Reiß*, Daten-gestützte Entscheidungsvorhersagesysteme vor deutschen Gerichten? Zum Einsatz Künstlicher Intelligenz in der Ziviljustiz, BRJ Sonderausgabe 1/2021, 18; *Sorge/Krüger* (Fn. 7), BRJ Sonderausgabe 1/2021, 13, 14; *Timmermann*, Legal Tech-Anwendungen, 2020, S. 273 f.

Menschen die justizielle Entscheidung treffen, aber eine menschliche Entscheidung vorbereiten bzw. einen Entscheidungsvorschlag unterbreiten sollen.

2 Entwicklungsstufen

Ausgehend von diesen Differenzierungskriterien zeigen sich bislang im Wesentlichen drei Entwicklungsstufen bzw. Generationen von datengestützten Systemen zur Prognose justizieller Entscheidungen.¹⁰

a) Erste Generation: Regelbasierte Systeme auf Grundlage von Metadaten

Archetypisch für eine erste Generation derartiger Systeme steht ein von *Katz* u. a. entwickeltes Modell zur Vorhersage von Entscheidungen des U.S. Supreme Court, einschließlich des Abstimmungsverhaltens der Richter*innen.¹¹ Zur Berechnung wurde die 140 Variablen umfassende Supreme Court Database (SCDB) herangezogen, in der Metadaten zum Verfahren – etwa Rechtsgebiet, Instanzgericht und Monat der mündlichen Verhandlung – erfasst sind. Die Gewichtung dieser Parameter wurde durch einen zeitvariablen Zufallswald-Algorithmus (*time-evolving random forest classifier*) mittels maschinellen Lernens ermittelt. Im Unterschied zu einem neuronalen Netz lässt sich bei einem solchen Algorithmus der Entscheidungsprozess und die Gewichtung der einzelnen Faktoren besser nachvollziehen.¹² Die Autoren geben eine Treffgenauigkeit von 70,2 % bezogen auf den – historischen – Verfahrensausgang an.¹³

¹⁰ Instruktiv zu den verschiedenen Erscheinungsformen ist der rechtsvergleichende Querschnitt von *Lelieur/Blount/Cherqaoui/Bampasika* (Fn. 1), 94(2) RIDP 33 ff. (2023).

¹¹ *Katz/Bommarito/Blackman*, A general approach for predicting the behavior of the Supreme Court of the United States, 12 PLOS ONE e0174698 (2017).

¹² Exemplarisch *Petkovic/Altman/Wong/Vigil*, Improving the explainability of Random Forest classifier – user centered approach, in: Altman/Dunker/Hunter/Ritchie/Murray/Klein (Hrsg.), Pacific Symposium on Biocomputing, 2018, S. 204.

¹³ Noch treffsicherer, aber weniger innovativ wären schließlich Systeme, die allein basierend auf der Variable „Einlegung des Rechtsmittels“ als Ergebnis *immer* dessen Erfolglosigkeit zu prognostizieren suchten: Bei Verfassungsbeschwerden erreichte man so eine Treffgenauigkeit von etwa 97 % (Beispiel nach *Sorge/Krüger* [Fn. 7], BRJ Sonderausgabe 1/2021, 13, 13f.), bei strafrechtlichen Revisionen immerhin von 80 % (unter Zugrundelegung der bei *Basdorf*, Aufhebungspraxis der Strafsenate des BGH 2015 bis 2019, NSTZ 2022, 399, 401 berichteten Zahlen). Dies unterstreicht die Notwendigkeit, solche Treffgenauigkeiten kritisch zu hinterfragen.

Systeme der ersten Generation sind somit dergestalt konstruiert, dass sie die justizielle Entscheidung basierend auf Metadaten über das Verfahren – und gerade nicht basierend auf Verfahrensakten, Schriftsätzen oder Plädoyers – prognostizieren.¹⁴ Für derartige, auf Metadaten fußende Prognosen lassen sich durch maschinelles Lernen trainierte, aber auch regelbasierte Algorithmen nutzen. Die Prognosegenauigkeit und damit der Aussagegehalt für einen konkreten, noch nicht gerichtlich entschiedenen Fall ist bei solchen Systemen zwar gering. Deren Stärke liegt indessen darin, empirische Zusammenhänge im Rechtssystem offenlegen zu können.

b) Zweite Generation: Maschinelles Lernen auf Grundlage fallbezogener Feststellungen

Herausragendes Beispiel für eine zweite Generation datengestützter Systeme zur Prognose justizieller Entscheidungen ist die von *Aletras* u. a. publizierte Studie zu Entscheidungen des *EGMR*: Die Feststellungen des *EGMR* zum Streitgegenständlichen Sachverhalt und zum Vortrag der Parteien wurden hier als Grundlage verwendet, um den (historischen) Verfahrensausgang mit 79 %iger Wahrscheinlichkeit zu prognostizieren.¹⁵ Spätere Studien vermögen dies zu replizieren¹⁶ und in Bezug auf italienische steuerrechtliche Entscheidungen mit ca. 70 %iger Trefferquote zu übertragen.¹⁷

Systeme, welche die vom Gericht selbst formulierte Entscheidungsgrundlage heranziehen, um den Ausgang des Verfahrens vorherzusagen, sind offensichtlich untauglich zur Vorhersage zukünftiger Entscheidungen: Gerichte veröffentlichen nämlich die Entscheidungsgründe, die aus diesen Feststellungen und der rechtlichen Würdigung bestehen, erst mit oder sogar erst *nach* dem Entscheidungstenor. Es wäre auch naiv anzunehmen, dass Berichterstatter*innen oder Spruchkörper zunächst die Feststellungen formulieren und sich erst in einem zweiten Arbeits-

¹⁴ Ebenso *Rühl*, KI in der gerichtlichen Streitbeilegung, in: Kaulartz/Braegelmann (Fn. 8), Kap. 14.1 Rn. 7f.

¹⁵ *Aletras/Tsarapatsanis/Preoŋiuc-Pietro/Lamos*, Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective, 2 PeerJ Computer Science e93 (2016).

¹⁶ *Medvedeva/Vols/Wieling*, Using Machine Learning to Predict Decisions of the European Court of Human Rights, 28 Artif Intell Law 237 (2020).

¹⁷ *Galli/Grundler/Fidelangeli/Galassi/Lagioia/Palmieri/Ruggeri/Sartor/Torroni*, Predicting Outcomes of Italian VAT Decisions, in: Francesconi/Borges/Sorge (Hrsg.), Legal Knowledge and Information Systems. (35. JURIX-Tagung), 2022, S. 188.

schritt mit der rechtlichen Würdigung befassen. Vielmehr beziehen sich Sachbericht und rechtliche Würdigung aufeinander und werden mit Blick auf die jeweils andere Komponente verfasst. Damit wird allerdings deutlich, worin die Stärke solcher Modelle der zweiten Generation liegt: in der Analyse der Binnenkohärenz von Entscheidungen, wofür die Korrelation zu Präjudizien als Maßstab herangezogen wird.

c) Dritte Generation: Maschinelles Lernen auf Grundlage von Verfahrensakten und Schriftsätzen

Die neueste Generation an Prognosesystemen zieht Verfahrensakten – sei es in unstrukturierter, sei es in strukturierter Form – heran, um diese in neuronale Netze einzuspeisen. So hat das an der Universität Cambridge entwickelte System *Case Cruncher Alpha* den Ausgang von Verfahren zu Restschuldversicherungen mit einer Treffgenauigkeit von 86,6 % besser vorhergesagt als eine Vergleichsgruppe von Rechtsanwält*innen (66,3 %).¹⁸ Noch höhere Werte von bis zu 95 % Treffgenauigkeit wird dem argentinischen System *Prometea* zugeschrieben, das dort bereits in Zivilverfahren, z.B. Miet- und Verkehrsunfallsachen, Verwendung finden soll.¹⁹ Ebenso zu dieser Generation zählen Entscheidungsunterstützungssysteme wie FraUKe, Codefy²⁰ und MAKI²¹, die an deutschen Zivilgerichten bereits Verwendung finden.

Zu betonen ist, dass derartige Genauigkeitswerte sich nicht auf den sachlich wie rechtlich *richtigen* Ausgang des Verfahrens beziehen, sondern auf die Übereinstimmung von Testläufen mit Trainingsdaten. Nicht nur hinsichtlich solcher *Ex-ante*-Evaluationen, sondern auch im praktischen Einsatz vermögen derartige Prognosesysteme nur Korrelation mit Präzedenzfällen aufzuzeigen.²²

Damit wird zugleich deutlich, dass derartige Systeme konzeptionell eine Ähnlichkeit zu *case-law*-basierten Rechtssystemen aufweisen, in denen vergleichbare Sachverhalte bei unveränderter Gesetzeslage nach dem Vorbild von Präjudizien zu

18 Vgl. Rühl, in: Kaulartz/Braegelman (Fn. 8), Kap. 14.1 Rn. 10 unter Verweis auf <https://www.case-crunch.com/#challenge>.

19 Corvalá, *Prometea: Artificial Intelligence to Transform Justice and Public Organizations*, 6 International Journal of Digital and Data Law 89 (2020).

20 Vgl. <https://justizministerium.hessen.de/presse/ki-projekt-codefy-am-landgericht-frankfurt-ge-startet> (17.12.2023).

21 Vgl. https://www.linkedin.com/posts/olg-celle_ki-k%C3%BCnstlicheintelligenz-legaltech-activity-711850503659053056-FtFU/.

22 Siehe oben II.1.c.

entscheiden sind. Solche Prognosesysteme sind jedoch nicht darauf angelegt, das im *Case-law*-System wichtige *distinguishing* zu leisten: sie sind nicht fähig, rechtlich bedeutsame Abweichungen des Sachverhalts von den zuvor bereits entschiedenen Sachverhalten zu erkennen.

Inhärent problematisch ist die Übertragung in Rechtssysteme wie das deutsche: Hier besteht keine strenge Bindungswirkung an Präjudizien,²³ sodass aus der sachlich richtigen und rechtlich vertretbaren Entscheidung eines Falles gerade nicht darauf geschlossen werden kann, dass ein vergleichbarer, zukünftiger Fall ebenso zu entscheiden ist. Ein Einsatz von Prognosesystemen könnte daher dazu beitragen, dass sich kontinental geprägte Rechtssysteme noch stärker als bisher am *case law* orientieren. Ein erster solcher Effekt ist bereits dahingehend zu verzeichnen, dass Forderungen nach einer vollständigen digitalen Verfügbarkeit *aller* Gerichtsentscheidungen in Deutschland lauter werden.²⁴

III Hürden, Gefahren und Potentiale datengestützter Systeme zur Prognose justizieller Entscheidungen

1 Hürden

a) Trainingsdaten und die Evolution des Rechts

Mit dem letztgenannten Aspekt der digitalen Verfügbarkeit von Gerichtsentscheidungen ist sogleich eine erste technische Hürde für Prognosesysteme aufgerufen: Insbesondere das Training von neuronalen Netzen erfordert große Mengen an qualitativ hochwertigen Daten, aus denen die zu erkennenden Muster maschinell erlernt werden sollen. Zwar liegen die Entscheidungen deutscher Höchstgerichte seit einigen Jahren vollständig und in digitaler Fassung vor. Instanzgerichtliche Entscheidungen werden jedoch nur in geringem Umfang veröffentlicht. Zudem ist nicht allein die Quantität der Gerichtsentscheidungen bedeutsam, sondern auch

²³ Statt aller *Hoch*, Big Data und Predictive Analytics im Gerichtsprozess: Chancen und Grenzen der Urteilsprognose, MMR 2020, 295, 298.

²⁴ „Gerichtsentscheidungen sollen grundsätzlich in anonymisierter Form in einer Datenbank öffentlich und maschinenlesbar verfügbar sein.“ (Koalitionsvertrag zwischen SPD, Bündnis 90/Die Grünen und FDP, Mehr Fortschritt wagen: Bündnis für Freiheit, Gerechtigkeit und Nachhaltigkeit, 2021, 106).

deren Qualität und Konsistenz.²⁵ So müsste ein System zur Prognose von Entscheidungen zu strafrechtlichen Konkurrenzen – um nur zwei Beispiele zu nennen – berücksichtigen, dass Entscheidungen, die eine fortgesetzte Handlung angenommen haben, infolge BGHSt 40, 138 überholt sind, und dass Entscheidungen, die zwischen Wohnungseinbruchdiebstahl und zugleich begangener Sachbeschädigung Konsumtion annehmen, infolge BGHSt 63, 253 keine taugliche Prognosegrundlage mehr sind. Je komplexer die zu prognostizierenden Gerichtsentscheidungen sind, desto mehr müssen Prognosesysteme die fortlaufende Veränderung des Rechts durch Rechtsetzung und Rechtsprechung berücksichtigen.²⁶ Hierfür gibt es, wie das erwähnte Prognosesystem zur Rechtsprechung des U.S. Supreme Court zeigt,²⁷ zwar innovative Ansätze. Die heutzutage in der Diskussion dominierenden neuronalen Netze haben jedoch ihre Stärke bei der Erkennung von Mustern, die im Verlauf der Zeit konstant bleiben.

b) Prognosegenauigkeit, „Distinguishing“ und das Vorandenken des Rechts

Eng damit verbunden ist eine zweite Herausforderung von Prognosesystemen: Weil deren Modelle auf der Vergangenheit und insbesondere auf Präjudizien basieren, ist deren Genauigkeit immer dann hoch, wenn das Verfahren, dessen Ausgang zu prognostizieren ist, eine hohe Übereinstimmung mit den Trainingsdaten aufweist. Somit dürfte die Prognosegenauigkeit insbesondere in „Standardfällen“ hoch sein, nicht hingegen in Fällen, in denen rechtlich bedeutsame Sachverhaltsabweichungen vorliegen.

Systeme, die auf einer Mustererkennung basieren, stehen daher insbesondere vor der Herausforderung, wie man zum Zeitpunkt des Trainings dieser Systeme noch unbekannte Muster erkennen kann, die sich zukünftig für die rechtliche Bewertung als entscheidend erweisen.²⁸ Zudem sind etliche Sachverhaltsabweichungen rechtlich irrelevant, sollten sich also nicht auf die Prognoseentscheidung auswirken. Letztlich erfordert all dies ein umfassendes Verständnis sowohl des Rechts als auch der Realität – und damit Fähigkeiten, die selbst modernsten neuronalen Netzen aktuell noch fehlen.

²⁵ *Sorge/Krüger* (Fn. 7), BRJ Sonderausgabe 1/2021, 13, 15.

²⁶ *Rühl/Reiß* (Fn. 9), BRJ Sonderausgabe 1/2021, 18, 19; *Sorge/Krüger* (Fn. 7), BRJ Sonderausgabe 1/2021, 13, 15.

²⁷ Siehe oben II.2.a. unter Verweis auf *Katz/Bommarito/Blackman* (Fn. 11), 12 PLOS ONE e0174698 (2017).

²⁸ Hierzu aus informationstechnischer Sicht *Pearl*, The Seven Tools of Causal Inference, with Reflections on Machine Learning, 62 Communications of the ACM 54 (2019).

Zusätzlich ist hiermit verbunden, dass Prognosesysteme inhärent ungeeignet sind, das Recht „voranzudenken“.²⁹ Auf Mustererkennung basierende Systeme vergleichen neue Eingabedaten lediglich mit der Vergangenheit und versuchen – überspitzt formuliert – die heutigen Probleme mit den Antworten von gestern zu lösen. Anpassungsbedarfe des Rechts und seiner Auslegung können solche Systeme nicht erkennen. Das erzeugt Spannungen etwa zum europäischen Verständnis der Grund- und Menschenrechte, dass diese dynamisch fortzuentwickeln sind, sodass z. B. die EMRK anerkanntermaßen „as a living instrument“³⁰ auszulegen ist.³¹

d) Korrelation statt Kausalität

Die bisherigen Prognosesysteme vermögen schließlich ausschließlich Korrelationen nachzuweisen. Sie zeigen auf, welche Muster in den Eingabedaten mit welchen Ergebnissen in den Ausgabedaten zusammenhängen. Mit Regressionsanalysen ist man z. B. zusätzlich in der Lage, die Stärke dieser Korrelationen zu bestimmen.

Es ist jedoch in aller Deutlichkeit darauf hinzuweisen, dass statistische Korrelationen nicht gleichzusetzen sind mit Kausalitäten; Korrelationen können lediglich auf vorhandene Kausalbeziehungen hindeuten. Am Beispiel: Dass die Uhrzeit der justiziellen Entscheidung (vor oder nach dem Mittagessen) in einer israelischen Studie mit der Strafaussetzung auf Bewährung korrelierte,³² kann mit dem Hunger der Richter*innen zusammenhängen, aber auch damit, dass leichtere Fälle vormittags, schwerwiegendere Straftaten nachmittags terminiert wurden.

2 Gefahren

Prognosesysteme basieren zunehmend auf Technologien des maschinellen Lernens, insbesondere unter Verwendung neuronaler Netzwerke. Daher sind die bei solchen Systemen allgemein bekannten Risiken zu berücksichtigen, die von einer antrainierten Diskriminierung (*bias*) über intransparente Datenverarbeitung und einer Enthumanisierung des Rechts hin zu einer großen Abhängigkeit staatlicher Akteure

²⁹ Hierzu Brodowski/Morsch (Hrsg.), Das Recht vorandenken? Reflexionen über die Rolle der (Straf-)Rechtswissenschaft. Kolloquium zum 80. Geburtstag von Heike Jung, 2023.

³⁰ EGMR, Urt. v. 25.4.1978 – Nr. 5856/72 (Tyrer ./ Vereinigtes Königreich).

³¹ Statt aller Grabenwarter/Pabel, Europäische Menschenrechtskonvention, 7. Aufl. 2021, § 5 Rn. 14 ff.

³² Danziger/Levav/Avnaim-Pesso, Extraneous factors in judicial decisions, 108 PNAS 6889 (2011).

von Privatunternehmen, die Eigeninteressen verfolgen, reichen.³³ Einige Gefahren müssen jedoch bei Prognosesystemen besonders hervorgehoben werden:

a) Begrenzter Aussagegehalt

Man darf nicht müde werden, den begrenzten Aussagegehalt von Entscheidungsprognosen zu betonen: Prognosesysteme sind nicht fähig, den Sachverhalt oder die Rechtslage zu ermitteln und zu subsumieren.³⁴ Sie können lediglich mit den zur Verfügung gestellten Eingabedaten arbeiten und Korrelationen zwischen diesen Eingabedaten und Präjudizien aufzeigen.³⁵ Infolgedessen sind sie nicht in der Lage, juristische Begründungen zu liefern. Sie vermögen lediglich statistische Zusammenhänge aufzuzeigen, die in den zugrunde gelegten Trainings- und Eingabedaten – gleich ob hoher oder niedriger Qualität – bereits enthalten sind. In aller Klarheit: Prognosesysteme versprechen zwar eine „Gleichheit“ und „Objektivität“; sie können diesem Anspruch jedoch nur im denjenigen Rahmen gerecht werden, der durch die Trainings- und weiteren Eingabedaten abgesteckt wird.

b) Schleichende Automatisierung

Besonders hervorzuheben ist zudem die Gefahr eines *automation bias*.³⁶ Selbst wenn ein Prognosesystem von Menschen lediglich als Hilfsmittel genutzt werden soll, führen Inertia-Effekte dazu, dass Menschen sich bei der nachfolgenden – autonom zu treffenden – Entscheidung an dieser Prognose orientieren. Ob solche Effekte zusätzlich durch ein Vertrauen in die vermeintlich objektive Technik verstärkt werden, muss die psychologische Forschung noch zeigen.

³³ Allgemein zu Gefahren des Einsatzes „Künstlicher Intelligenz“ in der (Straf-)Rechtspflege, statt vieler, *Bock/Höffler*, Künstliche Intelligenz und Kriminalität, *KriPoZ* 2022, 257; *Lelieur/Blount/Cherqaoui/Bampasika* (Fn. 1), 94(2) *RIDP* 11 (2023); *Malek*, Criminal courts' artificial intelligence: the way it reinforces bias and discrimination, 2 *AI Ethics* 233 (2022); *Martini*, Algorithmen als Herausforderung für die Rechtsordnung, *JZ* 2017, 1017; *Sprenger/Brodowski*, 'Predictive policing', 'Predictive Justice', and the use of 'Artificial Intelligence' in the Administration of Criminal Justice in Germany, *eRIDP A-02/2023*, 37 m.w.N.; *Valerius*, „Legal Tech“ im Strafverfahren?, *ZStW* 133 (2021), 152; sowie die Nachweise oben in Fn. 2.

³⁴ Statt vieler *Effer-Uhe* (Fn. 2), *JZ* 2023, 833, 834; *Hoch* (Fn. 23), *MMR* 2020, 295, 297 f.; *Timmermann*, Legal Tech-Anwendungen (Fn. 9), S. 273 f.

³⁵ Siehe bereits oben II.1.c.

³⁶ Hierzu, statt vieler, *Lelieur/Blount/Cherqaoui/Bampasika* (Fn. 1), 94(2) *RIDP* 43 ff. (2023); *Martini/Nink*, Strafjustiz ex machina?, in: Bertelsmann Stiftung (Hrsg.), *Automatisch erlaubt?*, 2020, S. 44, 50 f.

Solche Inertia-Effekte können durch die Hintertüre zu einer schleichenden Automatisierung führen, wenn Entscheidungsvorschläge eines Prognosesystems nicht oder nur unzureichend hinterfragt und der eigenen Entscheidung zugrunde gelegt werden. So hat eine französische Studie gezeigt, dass Richterkandidat*innen, je nach der ihnen mitgeteilten Entscheidungsprognose, anders entscheiden und ihre Entscheidungen auch anders begründen.³⁷ Die Frage, ob und wie bei einer Automatisierung oder auch nur bei Entscheidungsvorschlägen *meaningful human oversight* gewährleistet werden kann, wird insoweit zur Gralsfrage der Digitalisierung.

c) Selbsterfüllende Prophezeiungen und die Petrifizierung des Rechts

Angesichts der Bedenken gegen eine Ersetzung menschlicher Richter*innen durch Maschinen bei verfahrensabschließenden Entscheidungen³⁸ wird für Prognose- und darauf basierende Entscheidungssysteme ein Potential vor allem bei Vor- und Zwischenentscheidungen gesehen. So drängt sich als ein Anwendungsfeld die Bewertung des Prozesskostenrisikos auf Seite der Kläger*innen auf. Von Justizvertreter*innen wird als potentiell automatisierungsfähig angesehen die Entscheidung über die Gewährung von Prozesskostenhilfe (PKH) und dabei insbesondere die Teilfrage, ob die Klage „hinreichende Aussicht auf Erfolg“ hat (§ 114 Abs. 1 Satz 1 ZPO).³⁹

Indessen droht hierbei eine selbsterfüllende Prophezeiung: Diejenigen, die infolge einer Negativprognose keinen gerichtlichen Rechtsschutz suchen oder, mangels Gewährung von PKH, suchen können, haben ihren Rechtsstreit schon verloren und damit die Prognose bestätigt. Zugleich wird den Gerichten jegliche Chance genommen, die der Prognose zugrunde gelegte Rechtsprechung zu korrigieren. Ein zusätzliches Problem folgt, wenn nur diejenigen Kläger einen Zivilrechtsstreit beschreiten – oder PKH gewährt bekommen –, bei denen ein solches Prognosesystem eine hohe Obsiege Wahrscheinlichkeit berechnet. Dann setzt dies für die Gerichte einen entsprechenden Anker, der – entsprechend dem in der

³⁷ Vial, *Prise en main d'un outil d'intelligence artificielle par des auditeurs de justice: l'office du juge sous 'influence des algorithmes*, Recueil Dalloz 2022, 1928.

³⁸ Siehe hierzu oben I. bei und mit Fn. 2.

³⁹ Vgl. das Grundlagenpapier zur 74. Jahrestagung der Präsidentinnen und Präsidenten der OLGs, des KG, des BayObLG und des BGH, Einsatz von KI und algorithmischen Systemen in der Justiz, 2022, S. 37 verfügbar unter https://oberlandesgericht-celle.niedersachsen.de/download/184478/Grundlagenpapier_der_Arbeitsgruppe_zum_Einsatz_von_KI_und_algorithmischen_Systemen_in_der_Justiz.pdf (17.12.2023).

(rechts-)psychologischen Literatur umfänglich beschriebenen Ankereffekt und dem sogenannten *confirmation bias* – zusätzlich dazu beizutragen droht, dass sich die Prognose bestätigen wird.

Neben der oben bereits beschriebenen Unfähigkeit von Prognosesystemen, das Recht voranzudenken,⁴⁰ besteht ein zusätzlicher Anlass zur Sorge, dass das Recht zunehmend versteinert und verkrustet. Den Gerichten werden Gelegenheiten genommen, ihre Rechtsprechung zu überdenken; die Signalwirkung von negativen Gerichtsentscheidungen – und damit ein gewichtiger Anlass für Gesetzgebung⁴¹ – droht abhanden zu kommen.

Einiges spricht dafür, dass dem durch eine begrenzte Randomisierung bzw. Stochastizität entgegengewirkt werden muss, sprich dass Prognosesysteme dahingehend programmiert werden müssen, dass sie gelegentlich Fehlprognosen liefern. Dies ist bereits aus informationstechnischer Sicht erforderlich, um die Zuverlässigkeit und Fairness von Prognosen zu verfeinern und nicht einseitig in eine Risikoaversität abdriften zu lassen.⁴² Integriert man Stochastizität in das Prognosesystem, kann dies für den Einzelnen bedeuten, dass sein Verfahren anders behandelt wird als das nach einer „reinen“ Prognose der Fall wäre. Damit können für diesen erhebliche Mühen eines langwierigen und prognostizierbar aussichtslosen Rechtsstreits verbunden sein. Allerdings können Gerichte nur dann, wenn sie auch vermeintlich wenig aussichtsreiche Fälle erreichen, einen Rechtsprechungswandel einläuten.⁴³ Damit kann, in begrenztem Umfang, eine Inpflichtnahme des Einzelnen zur Führung auch aussichtsloser Rechtsstreitigkeiten sogar notwendig sein, um das Recht insgesamt voranzubringen.

⁴⁰ Siehe oben III.1.b.

⁴¹ Brodowski, Die Evolution des Strafrechts, 2023, S. 356 ff.

⁴² Kilbertus/Rodriguez/Schölkopf/Muandet/Valera, Fair Decisions Despite Imperfect Predictions. Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, 108 PMLR 277 (2020); Rateike/Majumdar/Mineeva/Gummadi/Valera, Don't Throw it Away! The Utility of Unlabeled Data in Fair Decision Making, in: FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, 2022, S. 1421; Evans/Benn/Quintana/Robinson/Thiébaux, Stochastic Policies in Morally Constrained (C-)SSPs, in: AIES '22: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, 2022, S. 253.

⁴³ Ob die Zeit dafür reif ist und sich daher für den Einzelnen die Mühen lohnen, kann ein auf Präzedenzfällen beruhendes Prognosesystem jedoch gerade nicht voraussehen; siehe hierzu oben III.1.a. und III.1.b.

3 Resolution der AIDP zum Einsatz von „Predictive Justice“

Angesichts dieser Gefahren hat die Internationale Strafrechtsgesellschaft (AIDP), gestützt auf eine umfangreiche rechtsvergleichende Studie von *Lelieur u. a.*,⁴⁴ auf einem internationalen Kolloquium in Buenos Aires im März 2023 Resolutionen u. a. zum Einsatz von „Predictive Justice“-Systemen verabschiedet. Darin spricht sich die AIDP für ein striktes Verbot des Einsatzes solcher Systeme bei Entscheidungen über Schuld und Strafe aus, ebenso bei Entscheidungen über Zwangsmaßnahmen wie Untersuchungshaft. Allenfalls in Massenverfahren wegen geringfügiger Delikte sei ein unterstützender Einsatz nicht generell zu untersagen, aber von der Gewährleistung effektiven und menschlichen Rechtsschutzes abhängig zu machen.⁴⁵

4 Potentiale

Trotz dieser weit verbreiteten und im Ausgangspunkt zustimmungswürdigen Skepsis ist es lohnenswert, sich weiterhin mit datengestützten Systemen zur Prognose justizieller Entscheidungen auseinanderzusetzen und derartige Systeme weiterzuentwickeln. Diese Einschätzung beruht indes weniger auf dem Einsatzpotential bei Massenverfahren und mehr auf drei bislang unzureichend genutzten Potentialen solcher Systeme.

a) Wissenschaftlicher Erkenntnisgewinn

Prognosen müssen nicht zwingend die Zukunft betreffen; Prognosen können sich auch – wie es bei den genannten Studien zur Rechtsprechung des *U.S. Supreme Court* und des *EGMR* der Fall ist – auf historisches Fallmaterial beziehen. Wenn Prognosesysteme dazu genutzt werden, Rechtsprechung und Rechtsentwicklung in quantitativer Hinsicht zu analysieren, können sie beispielsweise Anhaltspunkte dafür liefern, welches Gewicht welche Faktoren für den Ausgang eines Verfahrens haben – etwa die Uhrzeit der Entscheidung (und damit der Hunger der Richter*innen),⁴⁶ die Besetzung des Gerichts⁴⁷ oder auch die Vorstrafen des Beschul-

⁴⁴ *Lelieur/Blount/Cherqaoui/Bampasika* (Fn. 1), 94(2) RIDP 11 (2023); Deutscher Landesbericht: *Sprenger/Brodowski* (Fn. 33), eRIDP A-02/2023.

⁴⁵ AIDP Resolutions, Section 3 – AI and the administration of criminal justice: „Predictive policing,“ „predictive justice,“ and evidence. Buenos Aires, 28–31 March 2023, 94(2) RIDP 383, 390 f. (2023) Nr. 16 bis 23.

⁴⁶ Siehe erneut *Danzinger/Levav/Avnaim-Pesso* (Fn. 32), 108 PNAS 6889 (2011).

digten. Sie können daher wichtige, empirisch unterlegte Evidenzen für das Funktionieren, aber auch für Fehler des Rechtssystems und damit Potentiale für dessen Optimierung aufzeigen.

Prognosesysteme sollten somit nicht allein aus dem Blickwinkel einer Anwendungsorientierung, sondern auch aus dem der juristischen Grundlagenforschung gesehen werden: Sich abseits jeglicher Anwendungsfälle für die Justizpraxis mit der Prognose von Gerichtsentscheidungen zu befassen, kann dazu beitragen, das Rechtssystem, die Anwendung des Rechts und die Praxis juristischer Entscheidungen – und damit das Recht – besser verstehen zu lernen.⁴⁸

b) Einsatz dort, wo menschliche Entscheidungen besonders problembehaftet sind

Einsatzzwecke von Prognosesystemen sollten nicht schlicht nach Effizienzgesichtspunkten ausgewählt werden, sondern danach, wo ihr Einsatz einen besonderen Mehrwert bringen könnte. Das rückt den bislang vorrangig diskutierten Einsatzzweck „Massenverfahren“ in den Hintergrund, dafür aber justizielle Entscheidungen in den Vordergrund, bei denen eine menschliche Entscheidung durch das Gericht besonders kritikwürdig oder – weil sie einen Anker setzen, einen *confirmation bias* begründen oder Inertia-Effekte auslösen können – für das weitere Verfahren nachteilig ist.⁴⁹ Hierzu drei Beispiele aus dem Strafprozessrecht, die in meinen Augen eher diskussionswürdig sind als eine Automatisierung von Strafzumessung oder Gefährdungsprognosen:

Ein Computersystem könnte beispielsweise anstelle des bzw. der Vorsitzenden (vgl. § 213 Abs. 1 StPO) die Hauptverhandlungstermine nach der Verfügbarkeit der Beteiligten automatisch bestimmen. Ein solches System ließe sich rein regelbasiert programmieren und käme somit ohne „Künstliche Intelligenz“ in Gestalt maschinellen Lernens aus. Der Vorteil eines solchen Systems: Eine frühzeitige Belastung des Verhandlungsklimas durch Terminkonflikte und durch Streit über die gerichtliche Terminhoheit und deren Begrenzungen ließe sich weitestgehend vermeiden.

⁴⁷ In Frankreich ist eine Auswertung der Tätigkeit einzelner Richter*innen unter Strafe gestellt worden, Art. L. 10 Code de la justice administrative und Art. L. 111–113 Code de l'organisation judiciaire.

⁴⁸ In diese Richtung bereits *Zambrano*, *Précédents et prédictions jurisprudentielles à l'ère des Big Data*: papier sur le résultat (probable) d'un procès, 2015, S. 9.

⁴⁹ In diese Richtung auch *Busche*, *Legal Tech* und die verhaltenswissenschaftliche Analyse juristischer Entscheidungen, RW 2023, 69, 84 ff. sowie *Nink*, *Justiz und Algorithmen* (Fn. 2), S. 360 f.

Ein zweites Anwendungsfeld bezieht sich auf die Auswahl von Pflichtverteidiger*innen: Wenn der bzw. die Beschuldigte von seinem Bezeichnungsrecht keinen Gebrauch macht, ist dieser in der Regel vom Gericht zu bestellen (§ 142 Abs. 3 und 4 StPO) – nach Erhebung der Anklage konkret durch den bzw. die „Vorsitzende des Gerichts, bei dem das Verfahren anhängig ist“ (§ 142 Abs. 3 Nr. 3 StPO). Empirische Untersuchungen weisen darauf hin,⁵⁰ dass die Auswahl auch aus sachwidrigen Erwägungen auf solche Verteidiger*innen fallen kann, von denen wenig effektive Verteidigung zu erwarten ist. Eine Ersetzung durch ein starres Turnusverfahren hätte den Nachteil, dass möglicherweise besonders spezialisierte Pflichtverteidiger*innen übergangen würden; eine Verlagerung der Auswahlentscheidung auf die Rechtsanwaltskammern fand keine Unterstützung.⁵¹ Ließe sich die Auswahlentscheidung hingegen durch ein Prognosesystem automatisieren oder auch nur vorbereiten (z. B. fünf Vorschläge, aus denen der bzw. die Vorsitzende sodann auszuwählen hat), wäre sachwidrigen Erwägungen bei der Auswahlentscheidung Einhalt geboten. Noch lohnenswerter könnte ein Einsatz automatisierter Systeme beim Verteidigerwechsel nach § 143a Abs. 2 Nr. 3 StPO sein – hat die bisherige Rechtslage, solche Entscheidungen durch den bzw. die Vorsitzende(n) treffen zu lassen, doch den klaren Nachteil, dass vor diesem bzw. dieser Verteidigungsinterna auszubreiten sind, um zu begründen, dass das „Vertrauensverhältnis [...] endgültig zerstört ist“.

Drittens: Bei der Entscheidung über die Eröffnung des Hauptverfahrens (§§ 203, 204 StPO) ist der Maßstab anzulegen, ob „nach den Ergebnissen des vorbereitenden Verfahrens der Angeschuldigte einer Straftat hinreichend verdächtig erscheint“ (§ 203 StPO), also mit überwiegender Wahrscheinlichkeit wegen der angeklagten Tat zu verurteilen sein wird. Wäre es sinnvoll, ein – hypothetisch zu solch umfassender Prüfung fähiges – Prognosesystem anstelle des Hauptsachegerichts diese Entscheidung treffen zu lassen? Vorteilhaft wäre möglicherweise, dass diese Richter*innen nicht im selben Maße wie bisher durch die psychologischen Effekte, die mit einer *eigenen* Vor-Entscheidung einhergehen (Inertia-Effekt und *confirmation bias*), beeinträchtigt wären. Es ist aber zu befürchten, dass die Nachteile einer Automatisierung überwiegen – etwa dadurch, dass die vorgenannten Effekte dann lediglich durch einen *automation bias*⁵² ersetzt werden.

50 Jahn, Zur Rechtswirklichkeit der Pflichtverteidigerbestellung: Eine Untersuchung zur Praxis der Beordnung durch den Strafrichter nach § 140 Abs. 1 Nr. 4 der Strafprozessordnung in der Bundesrepublik Deutschland, 2014; Schöller, Die Praxis der Beordnung von Pflichtverteidigern, 2016.

51 Ein entsprechender Regelungsvorschlag – Schlothauer/Neuhaus/Matt/Brodowski, Vorschlag für ein Gesetz zur Umsetzung der Richtlinie (EU) 2016/1919 betreffend Prozesskostenhilfe für Verdächtige und Beschuldigte in Strafverfahren, HRRS 2018, 55, 57, 66 f. – wurde nicht aufgegriffen.

52 Siehe hierzu oben III.2.b.

c) Verbesserung menschlicher Entscheidungsfindung

Jedenfalls drängt sich aber auf, nach Systemen zu suchen, welche die menschliche Entscheidungsfindung *verbessern*, anstatt diese zu *ersetzen*.⁵³ Am Beispiel: Richter*innen des *EGMR* könnten das von *Aletras* u.a. entwickelte Prognosesystem⁵⁴ dazu nutzen, ihre Entscheidungsentwürfe daraufhin zu überprüfen, ob die darin enthaltenen gerichtlichen Feststellungen den Entscheidungstenor mit hoher Wahrscheinlichkeit tragen. Ist dies nicht der Fall, könnte dies ihnen Veranlassung dazu geben, ihre Begründung nachzuschärfen. Ein *automation bias* oder eine schleichende Automatisierung wäre nicht zu befürchten – aber möglicherweise ließe sich so ein Qualitätsgewinn in der Formulierung der Entscheidungsgründe erzielen.

IV Schluss

Der Einsatz datengestützter Systeme zur Prognose justizieller Entscheidungen ist inhärent mit einer Orientierung an Präjudizien verbunden und verträgt sich nur begrenzt mit einem Verständnis des Rechts, das dieses kontinuierlich zu optimieren, zu verbessern und „voranzudenken“ sucht. Solche Systeme können zu einer Petrifizierung des Rechts und zu einer risikoreichen schleichenden Automatisierung der Rechtsprechung führen. Zurecht setzten sich daher die Europäische Kommission und vor allem das Europäische Parlament in den Verhandlungen zum sogenannten KI-Gesetz dahingehend durch, dass es sich bei Prognosesystemen, die zur Automatisierung justizieller Entscheidungen oder auch nur zur Entscheidungsunterstützung herangezogen werden, um Hochrisikosysteme handelt (Anhang III Nr. 8 lit. a KI-Gesetz-E).⁵⁵

⁵³ In diese Richtung auch *Nink*, Justiz und Algorithmen (Fn. 2), S. 447f. mit dem Vorschlag eines Kontrollsystems *ex post*.

⁵⁴ Siehe oben II.2.b. unter Verweis auf *Aletras/Tsarapatsanis/Preoțiuc-Pietro/Lampos* (Fn. 15), 2 PeerJ Computer Science e93 (2016).

⁵⁵ Europäische Kommission, Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, COM(2021) 206 final i. d. F. Ratsdok. 5662/24. Das Europäische Parlament (P9_TA(2023)0236) schlug mit Abänderung 738 eine Verschärfung vor, während der Rat in seiner Allgemeinen Ausrichtung (Ratsdok. 15698/22) als Hochrisikosysteme nur solche „KI-Systeme“ einordnen wollte, „die bestimmungsgemäß von Justizbehörden oder in deren Namen zur Ermittlung und Auslegung von Sachverhalten und Rechtsvorschriften und zur Anwendung des Rechts auf konkrete Sachverhalte verwendet werden sollen“.

Dennoch haben Prognosesysteme, auch und gerade in einem menschenzentrierten Grundrechtsstaat, Potential und ihre Berechtigung: Entscheidend ist es, sie sinnvoll zu konzipieren und dergestalt einzusetzen, dass sie einen Mehrwert liefern – nicht singulär bezogen auf Einsparungen im Justizhaushalt, sondern umfassend: für das Verständnis des Rechts und seiner Anwendung, für die Qualität des Rechtssystems als Ganzes, und für die einzelnen Akteure im Rechtssystem.