

COLLANA DELLA RIVISTA DI DIRITTO ROMANO
SAGGI

AUTOMATISIERUNG
VON JURISTISCHEN
ENTSCHEIDUNGEN

DIE KÜNSTLICHE INTELLIGENZ
AM BEISPIEL
DER RÖMISCHRECHTLICHEN KASUISTIK

Herausgegeben von Iole Fargnoli

— Edizioni Universitarie di Lettere Economia Diritto —

ISSN 2499-6491
ISBN 978-88-5513-183-4
<https://doi.org/10.7359/1834-2024-aut-iur-ent>

Copyright 2024

LED Edizioni Universitarie di Lettere Economia Diritto
Via Cervignano 4 - 20137 Milano
Catalogo: www.lededizioni.com - www.ledonline.it

I diritti di riproduzione, memorizzazione e archiviazione elettronica, pubblicazione con qualsiasi mezzo analogico o digitale (comprese le copie fotostatiche, i supporti digitali e l'inserimento in banche dati) e i diritti di traduzione e di adattamento totale o parziale sono riservati per tutti i paesi.

Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume o fascicolo di periodico dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633.

Le fotocopie effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da CLEARedi, Centro Licenze e Autorizzazione per le Riproduzioni Editoriali, Corso di Porta Romana 108, 20122 Milano
e-mail autorizzazioni@clearedi.org - sito web www.clearedi.org

Umschlagbild:
Digitale Abbildung 'Der Traum des Erwachenden, der sich als Schmetterling fühlt'
© Renato Perani.

Stampa: Litogi - Milano

Inhaltsverzeichnis

Vorwort der Herausgeberin <i>Iole Fargnoli</i>	7
Autorenverzeichnis	15
Einleitung in das Kolloquium an der Universität Bern (20.-21. März 2024): «Automatisierung von juristischen Entscheidungen. Die künstliche Intelligenz am Beispiel der römischrechtlichen Kasuistik» <i>Fabio Addis</i>	17
Siebzig Jahre Rechtsanwendung durch Computer <i>Thomas Rüfner</i>	29
Herausforderungen beim Einsatz von künstlicher Intelligenz in der Justiz <i>Rolf H. Weber</i>	43
Chancen und Schwierigkeiten einer automatisierten Rechtsanwendung. Lassen sich Verzerrung und Verrauschung (Streuung) juristischer Entscheidungen durch den Einsatz künstlicher Intelligenz verringern? <i>Daniel Effer-Ube</i>	63
Revisiting the Digest through scripts and algorithms <i>Renato Perani</i>	81
<i>Damnum iniuria datum</i> , kasuistische Methode und ‚Künstliche Intelligenz‘ <i>Mario Varvaro</i>	91
Neuroheuristics, a flexible, problem-solving paradigm in Neuroscience <i>Alessandro E.P. Villa</i>	97

La persona e la sua identità: tra intelligenza artificiale e metaverso <i>Edoardo C. Raffiotta</i>	111
Literaturverzeichnis	125

Rolf H. Weber

Universität Zürich

Herausforderungen beim Einsatz von künstlicher Intelligenz in der Justiz

1. Einleitung – 2. KI-Programme als Arbeitsinstrument – 2.1. Einsatzmöglichkeiten – 2.2. Herausforderungen – 3. Transparenz und Datensicherheit als wichtige Rahmenbedingungen – 3.1. Transparenz bzw. Nachvollziehbarkeit – 3.2. Datenschutz und Datensicherheit – 3.3. Koordination der Schutzziele – 4. Einsatz von KI-Systemen in der Justiz – 4.1. Digitalisierung und IT-basierte KI-Systeme in der Rechtsordnung – 4.2. Einsatzfelder in einzelnen Verfahrensstadien der Justiz – 4.2.1. Feststellung des faktischen Sachverhalts – 4.2.2. Auffinden von Rechtssätzen – 4.2.3. Anwendung von Rechtsnormen – 4.3. Vorteile der Automatisierung – 4.4. Kontrolle von Risiken bei der Automatisierung – 4.4.1. Maschinen statt Menschen? – 4.4.2. Einzelne Risiken – 4.5. Gegenseitige Abwägung von Vorteilen und Risiken – 5. Ausblick.

1. EINLEITUNG

Der weite Begriff «Künstliche Intelligenz» (KI), der z.B. Erscheinungen wie automatisierte Entscheidungen, algorithmische Programmierungen und auch juristische Sprachmodelle bzw. allgemein Large Language Models (LLM) umfasst, hat in den letzten Monaten erheblich an Bedeutung gewonnen. Algorithmen und LLM sind in der Lage, die natürliche Sprache durch KI zu verarbeiten und gestützt darauf sachbezogene Texte zu entwickeln¹.

Die technischen Innovationen bzw. gar «Umwälzungen» führen aber auch zu neuen rechtlichen Herausforderungen, die rund um den Globus die Gesetzgeber haben aktiv werden lassen. Vorreiterin ist die Europäische Union mit dem Ar-

¹ Vgl. R. WEBER, *Juristische Sprachmodelle zwischen Transparenz und Datensicherheit*, in *Sprachmodelle: Juristische Papageien oder mehr?* (Hrsg. E. SCHWEIGHOFER et al.), Bern, 2024, S. 19.

tificial Intelligence Act (AIA) gewesen²; der Europarat hat anfangs Mai 2024 eine Artificial Intelligence Framework Convention verabschiedet³; der Schweiz will bis Ende 2024 die Grundlagen für die normative «Bewältigung» der KI bereitstellen⁴. Wichtige Themen im KI-Bereich sind der vorliegend nicht zu vertiefende Urheberrechtsschutz⁵, die gegenüber den Betroffenen geschuldete Gewährleistung der Transparenz bzw. Nachvollziehbarkeit von KI-Vorgängen, die Vermeidung des Eintritts von Risikofaktoren und (angesichts der grossen Datensammlungen) die Aufrechterhaltung der Datensicherheit⁶.

Mit Blick auf die Formulierung sachgerechter Rahmenbedingungen für den Einsatz von KI-Software in der Justiz spielen aber noch eine Reihe weiterer Faktoren eine Rolle, z.B. die Einhaltung von prozeduralen Verfahrensgarantien, die Beachtung staatspolitischer Grundsätze (z.B. Gewaltenteilung, demokratische Legitimation) und die Gewährleistung einer zumindest gewissen «Richtigkeitsgarantie» in der Entscheidungsfindung⁷.

2. KI-PROGRAMME ALS ARBEITSINSTRUMENT

2.1. Einsatzmöglichkeiten

KI-Programme und juristische Sprachmodelle zeichnen sich dadurch aus, dass sie die natürliche Sprache in schriftlicher Form auf der Basis von maschinellem Lesen verarbeiten und auch als Text präsentieren. Juristisches Arbeiten ist meist sprachorientiert; KI-Programme können deshalb ein gutes Unterstützungsinstrument bilden. Einsetzen lassen sich solche «Werkzeuge» in allen Bereichen, in denen eine (teil-)automatisierte Textverarbeitung und/oder Textproduktion stattfindet.

² Artificial Intelligence Act 2024/1689, 13. Juni 2024, OJ 2024 L vom 12. Juli 2024.

³ Committee on Artificial Intelligence (CAI), Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, Explanatory Report, 17 May 2024, CM(2024)52-addfinal.

⁴ Medienmitteilung vom 22. November 2023, Bundesrat prüft Regulierungsansätze für Künstliche Intelligenz, <https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen/bundesrat.msg-id-98791.html>.

⁵ W. STRAUB, *Immaterialgüterrechtlicher Schutz mit KI geschaffener Werke und Erfindungen*, in *Weblaw Jusletter*, 7. August 2023, Rz 2 ff. In den USA sind verschiedene Verfahren wegen Urheberrechtsverletzung hängig (z.B. New York, Times v. OpenAI); am 23. April 2024 hat der Internet Court Beijing die Verletzung der Persönlichkeitsrechte durch KI-generierte Stimmen als gegeben erachtet (https://english.bjinternet court.gov.cn /2024-04/24/c_706.htm).

⁶ WEBER, zit. (Fn. 1), S. 21 ff.

⁷ Im Einzelnen dazu hinten Ziff. 4.

Für textorientierte KI-Programme stehen die folgenden konkreten Beispiele bzw. Vorgänge im Vordergrund⁸:

- *Textgenerierung*: Verfassen von Entwürfen für Dokumente, Verfassen von Texten in einem bestimmten Schreibstil, Werkzeuge zur Textfortführung oder -vervollständigung;
- *Textbearbeitung*: Rechtschreib- und Grammatikprüfung;
- *Textverarbeitung*: Wort- und Textklassifikation, Markierung von Begriffen im Text, Textzusammenfassung, Frage-Antwort-System, Übersetzung.

Die konkrete Ausgestaltung der KI-Programme und der juristischen Sprachmodelle hängt von den beabsichtigten Einsatzformen und den angestrebten Innovationen ab. Der Einsatz in der Justiz erscheint insbesondere im Rahmen der Erledigung von standardisierten Aufgaben als sinnvoll, d.h. in Bereichen, in denen eine Voll- oder eine Teilautomatisierung offensichtliche Effizienzvorteile herbeizuführen vermag⁹.

2.2. Herausforderungen

Der Einsatz von KI-Programmen in der Justiz hat jedoch unter Beachtung verschiedener Herausforderungen zu erfolgen, um den nicht zu unterschätzenden Risiken zu begegnen. Selbst wenn ein sprachlich fehlerfreier und oft auch inhaltlich überzeugender Text generiert wird, erweist sich nämlich ein unbegrenztes Vertrauen in die Aussagen von KI-Programmen als problematisch, weil bei den entsprechenden automatisierten Abläufen nicht ein menschenähnliches Leistungsvermögen vorausgesetzt werden darf¹⁰. In Betracht fallen folgende *Herausforderungen*¹¹:

- *Fehlende Vertraulichkeit der eingegebenen Daten*: Bei der Nutzung von externen Schnittstellen fließen alle Eingaben zunächst an den Betreiber des Modells; sensible und vertrauliche Informationen, die – über das allgemeine Datenschutzrecht¹² hinaus – nicht zuletzt im Rahmen von Berufs- und Amtsgeheimnissen in den Verfahren der Justiz¹³ oft auftreten, bedürfen deshalb der

⁸ Bundesamt für Sicherheit in der Informationstechnik, Grosse KI-Sprachmodelle, Chancen und Risiken für Industrie und Behörden, Bonn, 2021, S. 7.

⁹ Vgl. hinten Ziff. 4.3 und S. GLESS, *Künstliche Intelligenz in der Gerichtsbarkeit*, in ZSR, 1, 2023, S. 429, 445 f.

¹⁰ Vgl. auch R. WEBER, *Dürfen Menschen über Maschinen entscheiden?*, in *Schweizer Monat*, Februar 2019, S. 72, 73.

¹¹ Eingehender dazu Bundesamt für Sicherheit, zit. (Fn. 8), S. 10 f.

¹² Dazu hinten Ziff. 3.2.

¹³ In Justizverfahren spielen doppelte Vertraulichkeitsregeln eine Rolle, nämlich das (private) Anwaltsgeheimnis und das (öffentliche) Amtsgeheimnis.

Verknüpfung mit besonderen Schutzvorkehrungen.

- *Mangelnde Faktizität und Reproduzierbarkeit*: Die Generierung von Text beruht auf stochastischen Korrelationen, d.h. das «Wissen» wird aus (bereits gesehenen) Texten abgeleitet. Bezüge zur realen Welt existieren nicht, weshalb es bei für Menschen verständlichen Sachverhalten dennoch zu nicht korrekten Aussagen kommen kann; dieselbe Eingabe vermag auf Grund des wahrscheinlichkeitsbasierten Ansatzes zu unterschiedlichen Aussagen zu führen.
- *Ungenügende Aktualität*: Bietet das KI-Programm keinen Zugriff auf Live-Internetdaten (d.h. auf Informationen über aktuelle Ereignisse bzw. kürzlich ergangene Entscheide in der Justiz, national und grenzüberschreitend), werden ggf. aufgrund der in der Vergangenheit geprägten Datensammlungen nicht aktuelle Ergebnisse verarbeitet bzw. produziert.
- *Fehlerhafte Reaktion auf spezifische Eingaben*: Fehler vermögen aufzutreten, wenn Eingaben zu verarbeiten sind, die so stark von den Texten in den Trainingsdaten abweichen, dass sie sich nicht mehr korrekt in einem neuen Dokument verarbeiten lassen (im Schrifttum oft «Halluzination»¹⁴ als Ausdruck der Sinnestäuschung genannt); solche Abweichungen können zufällig auftreten oder auch absichtlich durch Interventionen Dritter als bewusster Täuschungsakt entstehen.
- *Anfälligkeit für «versteckte» Dateninfiltrationen mit manipulativer Absicht*: Falls ungeprüfte Dokumente Dritter im KI-System vorhanden sind oder Situationen eintreten, in denen Angreifende für Nutzende unbemerkt gewisse Eingaben einzubringen in der Lage sind, leben Risiken auf (z.B. Angriff auf ein Chat-Tool).

Eine grosse Herausforderung stellt die oft – gestützt auf vertragliche Absprachen bei einem Dritten verbleibende – fehlende Datenhoheit beim Einsatz von KI-Programmen dar; verstärkt wird die Problematik, wenn die fehlende Datenhoheit mit einer grossen Abhängigkeit vom Hersteller und/oder Betreiber des Modells verknüpft ist¹⁵.

¹⁴ Zur «Halluzination» aus technologischer Sicht insbesondere Y. ZHANG et al., *Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models*, September 2023, <https://arxiv.org/pdf/2309.01219v1.pdf>; vgl. auch GLESS, zit. (Fn. 9), S. 456.

¹⁵ Wettbewerbsrechtlich ergibt sich insbesondere das Problem des missbräuchlichen Verhaltens durch ein (relativ) marktmächtiges Unternehmen (Art. 7 Abs. 1 KG).

3. TRANSPARENZ UND DATENSICHERHEIT ALS WICHTIGE RAHMENBEDINGUNGEN

Abgesehen von den Besonderheiten, die beim Einsatz von KI in der Justiz eine Rolle spielen¹⁶, sind allgemeine Rahmenbedingungen zu beachten, die bei jeder KI-Verwendung bzw. Automatisierung von Bearbeitungsvorgängen eine Rolle spielen, nämlich die Transparenz bzw. die Nachvollziehbarkeit der KI-Abläufe sowie die Einhaltung der Datenschutz- und Datensicherheitskriterien.

3.1. Transparenz bzw. Nachvollziehbarkeit

Ein grundlegendes Problem beim Einsatz von KI-Programmen, das auch die Justiz betrifft, besteht in der mangelnden Transparenz mit Bezug auf deren Funktionsweise (z.B. Einsatz von Algorithmen, verwendete Daten, angewendete Methoden); akzentuiert wird das Risiko der Intransparenz bei den kaum steuerbaren selbstlernenden Algorithmen¹⁷.

Selbst soweit es gesetzliche Transparenzanforderungen gibt, wie z.B. mit Blick auf die automatisierten Entscheidungen gemäss Art. 21 DSGVO bzw. Art. 22 DSGVO, ist deren Wirkung nicht zu überschätzen¹⁸. Empirische Untersuchungen zu den vorhandenen daten- und Verbraucherschutzrechtlichen Informationspflichten lassen nämlich daran zweifeln, ob die gesetzlich geforderte Transparenz stets einen Mehrwert für die betroffenen Personen generiert und ihren angedachten Zweck tatsächlich erfüllt¹⁹. Die zur Verfügung gestellten Informationen sind oft nicht ausreichend verständlich und überfordern angesichts der Informationsintensität (Länge, Detaillierungsgrad) die Empfänger/innen²⁰. Diese allgemeine Problematik wird dadurch verschärft, dass selbstlernende System- und Sprachmodelle oft einen hohen Komplexitätsgrad aufweisen; überdies ist eine sachdienliche Erklärung bei Machine Learning-Verfahren bisweilen technisch nur schwer zu bewerkstelligen²¹.

Weiter müssen KI-Programme und Sprachmodelle grundsätzlich erklärbar und interpretierbar sein, um den betroffenen Personen die Nachvollziehbarkeit

¹⁶ Vgl. hinten Ziff. 4.

¹⁷ WEBER, zit. (Fn. 1), S. 21.

¹⁸ Vgl. dazu hinten Ziff. 3.2.

¹⁹ R. WEBER, S. HENSELER, *Regulierung von Algorithmen in der EU und in der Schweiz*, in *EuZ*, 2020, S. 28, 38.

²⁰ R. WEBER, *The Disclosure Dream - Towards a New Transparency Concept in EU Consumer Law*, in *EuCML*, 2023, S. 67 ff. m.w.H.

²¹ WEBER, HENSELER, zit. (Fn. 19), S. 39.

von algorithmischen Vorgängen zu ermöglichen²². Notwendig ist somit, abhängig vom Wissenshorizont des Betroffenen und der Bedeutung der Vorgänge, eine ausreichende Verständlichkeit, d.h. die zugrunde liegende Logik hat – gerade im Bereich der Justiz – auch begreiflich zu sein²³. Dieses Ziel ist in einer hoch technologisierten Umgebung nur schwer zu erreichen; deshalb ist davon auszugehen, dass selbst im Falle einer hohen Transparenz bei automatisierten Entscheidungen eine menschliche Nachkontrolle nicht vollständig unterbleiben kann²⁴.

3.2. Datenschutz und Datensicherheit

KI-Programme, insbesondere in Kombination mit juristischen Sprachmodellen, nutzen regelmässig auch Trainingsdaten mit Personenbezügen, die nur unter Einhaltung der datenschutzrechtlichen Vorgaben bearbeitet werden dürfen²⁵. Bei der Ausgestaltung der Automatisierungsvorgänge ist deshalb sicherzustellen, dass keine persönlichen Informationen über Datensubjekte, die in Gerichtsverfahren involviert sind, ggf. Drittpersonen zur Kenntnis gelangen. Im Rahmen von privatrechtlichen Verhältnissen lässt sich z.B. auf ein auf der OpenAI Website auffindbares Formular zurückgreifen, mit dem Widerspruch erhoben werden kann²⁶; materiell müssen also Auskunftsrechte, damit der Betroffene überhaupt Kenntnis von der Verwendung der Personendaten erhält, sowie Änderungs- und Löschungsrechte, welche den Betroffenen in die Lage versetzen, die Datenrichtigkeit herzustellen, vorhanden sein²⁷.

Im Falle der Bearbeitung von Personendaten durch Bundesorgane ist das Legalitätsprinzip zu beachten, d.h. der Einsatz von KI-Programmen setzt das Vorliegen einer gesetzlichen Grundlage (Art. 34 DSGVO) voraus²⁸. Diese Voraussetzung gilt in besonderem Masse für die Datenbearbeitung durch (staatliche) Gerichte, die regelmässig mit Personendaten zu tun haben; in der Praxis dürfte es derzeit oft

²² Aus technologischer Sicht vgl. M. DANILEVSKY et al., *A Survey of the State of Explainable AI for Natural Language Processing*, December 2020, <https://doi.org/10.48550/arXiv.2010.00711>.

²³ R. WEBER, *Künstliche Intelligenz: Regulatorische Überlegungen zum «Wie» und «Was»*, in *EuZ*, 2022, B1, B12.

²⁴ Bundesamt für Sicherheit, zit. (Fn. 8), S. 8; WEBER, zit. (Fn. 1), S. 22.

²⁵ Grundlegend D. ROSENTHAL, *Datenschutz beim Einsatz generativer künstlicher Intelligenz*, in *Weblaw Jusletter*, 6. November 2023; vgl. auch R. WEBER, *Künstliche Intelligenz und Datenschutz*, in *Weblaw IT-Jusletter*, 4. Juli 2024; F. THOUVENIN, ST. VOLZ, *White Paper: Datenschutz*, Juni 2024, <https://www.itsl.uzh.ch/de/Wissenstransfer/Publikationen.html#Positionspapier>.

²⁶ Vgl. https://share.hsforms.com/1UPy6xqxZSEq.TrGDh4ywo_g4sk30.

²⁷ Vgl. auch CH. LUTZ, B. DOMENIG, A. FLÜKIGER, *Datenschutzkonformer Einsatz von Chat-GPT an Schulen*, in *sic!*, 2024, S. 94, 96.

²⁸ Vgl. auch D. PINA, *Die Offenbarung von KI-Patenten*, Zürich, 2024, S. 30 f.

an der erforderlichen gesetzlichen Grundlage fehlen.

Im Kontext von automatisierten Entscheidungen kommt Art. 21 DSGVO bzw. Art. 22 DSGVO zur Anwendung; rechtsvergleichend ist der Unterschied zu beachten, dass im EU-Recht ein Verbotstatbestand vorgesehen ist, während das Schweizer Recht sich grundsätzlich mit Informations- und Korrekturpflichten begnügt, d.h. weniger restriktiv ist²⁹. Zur Einhaltung der Datenschutzvorgaben ist zudem an die Möglichkeit der Verschlüsselung von Personendaten zu denken; neuere Technologien bieten bessere Schutzmöglichkeiten an³⁰.

Ein wichtiger Aspekt des Datenschutzrechts ist die Datensicherheit (Art. 8 DSGVO bzw. Art. 32 DSGVO). Selbst wenn beim Stichwort «Datensicherheit» meist die Risiken im Vordergrund stehen, ist nicht zu übersehen, dass KI-Programme durchaus auch zur Verbesserung der IT-Sicherheit beitragen können; Beispiele sind (i) die Unterstützung bei der Detektion unerwünschter Inhalte, (ii) die Unterstützung bei der Berichterstattung zu Sicherheitsvorfällen, (iii) die Unterstützung bei der Erstellung und Analyse von Programmcodes (Untersuchung von Programmen auf Sicherheitslücken) sowie (iv) die Unterstützung bei der Analyse des Datenverkehrs³¹. Inwieweit die Verwendung von regelbasierten Systemen (z.B. Experten-Entscheidungsbäume) zu einer Erhöhung der IT-Sicherheit beizutragen vermag, weil sie strukturierte Programmabläufe ermöglichen, erscheint derzeit noch als ungeklärt³².

KI-Programme und automatisierte Entscheidungsverfahren sind aber auch verschiedenen Risiken und Missbrauchsgefahren ausgesetzt; abgesehen von den dargestellten allgemeinen Herausforderungen bestehen Risiken etwa mit Blick auf³³ (i) das Social Engineering (z.B. durch Preisgabe von Daten, Umgehung von Schutzmassnahmen, Installierung eines Schadcode), (ii) den Einsatz von Malware (z.B. Erzeugung eines Schadcode ohne viel Hintergrundwissen dank ausgereifterer Code-Generierungsfähigkeiten selbst bei Vorhandensein polymorpher Schadsoftware) und (iii) das Auftreten einer sog. «Hoax» (z.B. Falschmeldung wegen Desinformation oder Propaganda)³⁴.

Um eine ausreichende Datensicherheit zu gewährleisten, sind Massnahmen technischer und organisatorischer Art gegen mögliche Risiko- oder Missbrauchs-

²⁹ Vgl. WEBER, HENSELER, zit. (Fn. 19), S. 35; detailliert dazu F. SUTER, *Automatisierte Einzelentscheidungen im (Schweizer) Datenschutzrecht*, Diss. Zürich, 2024, S. 245 ff.

³⁰ Vgl. zur Verschlüsselung von Trainingsdaten auch PINA, zit. (Fn. 28), S. 46 ff.

³¹ Eingehender dazu Bundesamt für Sicherheit, zit. (Fn. 8), S. 9, und WEBER, zit. (Fn. 1), S. 22 m.w.H.

³² Zur Bedeutung von Entscheidungsbäumen vgl. auch GLESS, zit. (Fn. 9), S. 440 f.

³³ Bundesamt für Sicherheit, zit. (Fn. 8), S. 12 f.

³⁴ Aus technologischer Sicht vgl. L. WEIDINGER et al., *Taxonomy of Risks posed by Language Models*, June 2022, <https://doi.org/10.1145/35311463533088>.

fälle rechtzeitig einzurichten; Compliance-Vorkehrungen müssen die Authentizität von Textverfasser und Textinhalt sicherstellen bzw. den Nachweis bereit halten, welche Texte oder Informationen tatsächlich von einer bestimmten Person, Personengruppe oder Institution stammen³⁵.

3.3. Koordination der Schutzziele

Zwischen der Transparenz bzw. der Nachvollziehbarkeit automatisierter Verfahren sowie dem Datenschutz bzw. der Datensicherheit bestehen aber Spannungsfelder, die es als notwendig erscheinen lassen, unterschiedliche Interessenlagen zu überbrücken.

- (i) Transparenz schafft grundsätzlich Klarheit und Nachvollziehbarkeit, die im Datenschutzbereich z.B. die Validierung von Personendaten (insbes. Datenrichtigkeit: Art. 6 Abs. 5 DSGVO bzw. Art. 5 Abs. 1 lit. d DSGVO) einschliesst und die auf die Rückverfolgung und Gewichtung der angewendeten Parameter ausreichenden Wert legt. Der Transparenz zu dienen vermögen auch neue Verfahren zur Authentisierung durch technische Prozesse³⁶. Gleichzeitig gilt der allgemeine (quantitativ zu verstehende) Grundsatz, dass je mehr Daten offengelegt werden, umso grösser auch das Risiko ist, dass Angriffe von Dritten erfolgreich sein können.
- (ii) Die Datensicherheit lässt sich z.B. erhöhen durch die Einschränkung der Nutzungsmöglichkeiten mit Bezug auf die zur Verfügung stehenden Datenbestände; dies liegt aber oft nicht im Interesse der auf KI-Systeme zurückgreifenden Betroffenen³⁷. Immerhin stehen heute automatisierte Detektionswerkzeuge mit Blick auf automatisch generierte Texte zur Verfügung; dadurch erhalten die Betroffenen, technische Kenntnisse vorausgesetzt, die Fähigkeit, einzelne Texte als maschinengeschrieben zu erkennen und damit ggf. ihre Authentizität und die Richtigkeit der enthaltenen Daten anzuzweifeln³⁸.

Mit Blick auf die Entwicklung transparenter und gleichzeitig sicherer KI-Programme sind verschiedene Risikovermeidungs- und Risikominderungsmaßnahmen in Betracht zu ziehen. Im Vordergrund steht die Sicherstellung der Datenqualität bei der Auswahl von Trainingsdaten. Die konkrete und transparente Bestimmung der Trainingsdaten ist wichtig, um ein hohes Qualitätsniveau mit Blick auf die algorithmischen Ergebnisse zu erreichen. Die verwendeten Trai-

³⁵ WEBER, zit. (Fn. 1), S. 23.

³⁶ Zu den neueren Entwicklungen bei den technologischen Authentifizierungsprozessen im Finanzmarktbereich vgl. M. BOŠKIĆ, S. HEPP, *zkKYC in Decentralized Finance (DeFi)*, in *GesKR*, 2023, S. 209, 214 ff.

³⁷ Eingeschränkte Nutzungsmöglichkeiten verringern den Wert des Einsatzes von Algorithmen bzw. der verwendeten KI-Systeme.

³⁸ WEBER, zit. (Fn. 1), S. 24.

ningsdaten sind zudem Grundlage für die späteren Textgenerierungs-Modelle³⁹. Zur Datenqualität gehört weiter die Vermeidung von (versteckten) Diskriminierungen, die sich durch die Verwendung der Algorithmen perpetuieren könnten⁴⁰.

Weiter sind Massnahmen zu implementieren, die gegen Angriffe Dritter schützen. Technische Experten haben verschiedene Arten von möglichen Angriffen identifiziert. Mit Blick auf Justizverfahren sind folgende Angriffe erwähnenswert⁴¹:

- *Privacy Attacks*: Möglichkeiten zur Verminderung der Anfälligkeit für solche Angriffe sind z.B. die manuelle Auswahl oder automatische Filterung bzw. Anonymisierung von Daten, die Entfernung von Dopplungen aus Trainingsdaten zwecks Herabsetzung der Wahrscheinlichkeit einer möglichen Rekonstruktion, die Anwendung von die Differential Privacy verbessernden Mechanismen, die Einschränkung der Zahl von Modellen bzw. des Zugriffs auf das Modell sowie die Vornahme von zusätzlichen Trainings bzw. eines besonderen Trainings für sensible Daten⁴².
- *Adversarial and Poisoning Attacks*: (i) Vorkehrungen gegen solche Angriffe zielen darauf ab, auch zwecks Erschwerung der Wahrnehmung einer Veränderung absichtlich nur leicht veränderte oder intransparent gestaltete Texte als Verfälschungen erkennbar zu machen⁴³. (ii) Möglichkeiten zur Herabsetzung der Anfälligkeit für solche Angriffe sind die Verwendung vertrauenswürdiger Quellen als Trainingsdaten, der Einsatz von geschultem und erfahrenem Personal, die intensivierte Analyse von Bewertungen und die Beschränkung der Auswirkungen des Einsatzes auf ein kontrollierbares Feld⁴⁴.

Gesamthaft betrachtet lässt sich deshalb die Annahme treffen, dass bei Vornahme sachgerechter Schutzvorkehrungen der Einsatz von KI-Programmen sowohl zur Sicherstellung einer ausreichenden Transparenz bzw. Nachvollziehbarkeit als auch zur Einhaltung ausreichender Datensicherheitsgarantien beizutragen in der Lage ist⁴⁵.

³⁹ Eingehender dazu ROSENTHAL, zit. (Fn. 25), Rz 26 ff.

⁴⁰ WEBER, HENSELER, zit. (Fn. 19), S. 31 f.

⁴¹ Eingehender dazu Bundesamt für Sicherheit, zit. (Fn. 8), S. 16 ff. m.w.H.

⁴² Aus informationstechnologischer Sicht vgl. R. STAAB et al., *Beyond Memorization: Violating Privacy via Interference with Large Language Models*, October 2023, <https://arxiv.org/pdf/2310.07298v1.pdf>.

⁴³ Bundesamt für Sicherheit, zit. (Fn. 8), S. 18.

⁴⁴ WEBER, zit. (Fn. 1), S. 25.

⁴⁵ Vgl. dazu auch ROSENTHAL, zit. (Fn. 25), Rz 55 ff.

4. EINSATZ VON KI-SYSTEMEN IN DER JUSTIZ

4.1. Digitalisierung und IT-basierte KI-Systeme in der Rechtsordnung

KI-Programme sind Teil der umfassenden Digitalisierung des Lebens und der beruflichen Tätigkeiten der letzten Jahre. Die Digitalisierung ist aber auch mit Bezug auf die Justizverfahren viel breiter. Das Projekt «Justitia 4.0» will den Aufbau und den Betrieb der Plattform «Justitia Swiss» realisieren⁴⁶. Rechtliche Grundlage soll das derzeit im Parlament zur Debatte stehende Bundesgesetz über die Plattformen für die elektronische Kommunikation in der Justiz (BEKJ) sein⁴⁷; der Gesetzesentwurf sieht ein Obligatorium für Justizbehörden und Anwaltschaft vor, diese Plattform in Justizverfahren zu benutzen. Die Grundversion der Plattform ist schon entwickelt worden; mit ihrem Einsatz ist aber frühestens 2025 zu rechnen⁴⁸. Diese Plattform stellt ein digitalisiertes Werkzeug zur Vereinfachung des Dokumentenaustausches dar, hat indessen wenig zu tun mit der Verwendung von KI-Programmen⁴⁹.

In den letzten Jahren ist die Diskussion aufgekommen, ob Software-Programmen und/oder Robotern eine Rechtspersönlichkeit zukommen kann oder nicht. So wie vor knapp 200 Jahren Carl von Savigny die fiktive Rechtspersönlichkeit von juristischen Personen anerkannt hat, wäre es denkbar, eine ähnliche Fiktion für IT-Systeme zu schaffen; entsprechende Diskussionsansätze sind v.a. in der anglo-amerikanischen Doktrin durchaus zu finden⁵⁰.

Bisher geht die überwiegende Lehre in der Schweiz und in Kontinentaleuropa indessen davon aus, dass IT-Systeme formal keine Rechtspersönlichkeit innehaben können⁵¹. Zwar wird erwogen, die Verleihung der Rechtspersönlichkeit an die von den Betroffenen zu übernehmende Verantwortung zu knüpfen, die von IT-Systemen generierten Rechtslösungen als Vergleichsvorschläge zu verstehen,

⁴⁶ Dazu M. GYSIN ROBERT, *Justitia 4.0: Plattform geht in die Pilotierung*, in *Anwalts Revue*, 2024, S. 64 f.; B. WAEBER, *Effizienzsteigerung durch Digitalisierung. Nutzen und Mehrwert von Justitia 4.0 für die Justiz*, in *Weblaw Jusletter*, 2. September 2024.

⁴⁷ Vgl. *Botschaft* vom 15. Februar 2023, *BBl*, 2023, S. 679 ff. Der Nationalrat hat im Herbst 2023 zugestimmt, der Ständerat im Herbst 2024, doch ist die Differenzbereinigung noch ausstehend.

⁴⁸ GYSIN ROBERT, zit. (Fn. 46), S. 65.

⁴⁹ Im Zusammenhang mit dem Fall «Vincenz» wird in den Medien bereits diskutiert, ob das Projekt «Justitia 4.0» ein «Befreiungsschlag» für die «Aktenschieberei» sein könnte: vgl. A. STECK, in *NZZ am Sonntag*, 3. März 2024, S. 27.

⁵⁰ Zur anglo-amerikanischen Diskussion vgl. P. CHESTERMAN, *We, the Robots?*, Cambridge, 2021; zur Diskussion in Deutschland vgl. den eben erschienenen Beitrag von J. MAATZ, *Rechtssubjektivität im digitalen Zeitalter*, in *CR*, 2024, S. 323 ff.

⁵¹ Vgl. WEBER, zit. (Fn. 10), S. 73; GLESS, zit. (Fn. 9), S. 435.

welche von den Rechtsuchenden ggf. zu akzeptieren wären, was eine Zustimmung bedeuten würde⁵²; bisher hat dieser Gedankenansatz aber noch keine verbreitete Akzeptanz gefunden.

Die Rechtspersönlichkeit basiert zudem auch bei juristischen Personen letztlich auf personalen Entscheidungsmechanismen, d.h. natürliche Personen handeln für die «künstliche Einheit»; eine entsprechende Zuordnung müsste vergleichbar bei IT-Systemen bzw. Robotern möglich sein, um die Handlungsfähigkeit zu begründen⁵³. Überdies wäre ein Konzept zu entwickeln, das es ermöglichen würde, eine sachgerechte Haftung von IT-Systemen (mit einem entsprechenden Haftungssubstrat) vorzusehen⁵⁴.

4.2. Einsatzfelder in einzelnen Verfahrensstadien der Justiz

In der Justiz lassen sich verschiedene Verfahrensstadien unterscheiden, die mit Bezug auf die Automatisierung der Abläufe und insbesondere den Einsatz komplexer KI-Programme unterschiedliche Beurteilungen notwendig machen. Oft wird differenziert zwischen der Feststellung des faktischen Sachverhalts, dem Auffinden von Rechtssätzen und der Anwendung der relevanten Rechtsnormen auf den gegebenen Sachverhalt (zum Teil mit einer besonderen Phase für die Begründung des Entscheids⁵⁵); kürzlich ist auch eine Aufteilung in Rechtsmitteleingang, Spruchkörper, Instruktion, Entscheidungsfindung und Fallabschluss vorgeschlagen worden⁵⁶.

4.2.1. Feststellung des faktischen Sachverhalts

Der erste Schritt in einem Justizverfahren betrifft regelmässig die Feststellung des faktischen Sachverhaltes. Ein Gericht wird nur tätig, wenn ein entsprechender Antrag, der mit den relevanten Daten unterlegt ist, anhängig gemacht wird⁵⁷; dabei kann die Entscheidung über die Relevanz bzw. Irrelevanz der Daten, verbunden

⁵² GLESS, zit. (Fn. 9), S. 461.

⁵³ WEBER, zit. (Fn. 10), S. 75.

⁵⁴ Weil IT-Systeme selbst über kein Haftungssubstrat verfügen, müsste eine Zuordnung an dafür verantwortliche Einzelpersonen oder juristische Gemeinschaften erfolgen.

⁵⁵ Ein gesondertes Einsatzfeld «nur» für die Begründung wird von GLESS, zit. (Fn. 9), S. 443 f., befürwortet, was indessen nicht überzeugt, weil die Anwendung von Rechtssätzen die Begründung bereits impliziert.

⁵⁶ So D. SCHNEEBERGER, *Digitale Wende in der Rechtsprechung: Der Einfluss von KI auf die Justiz*, in *Justice - Justiz - Gustizia*, 1, 2024, Rz 6 ff.

⁵⁷ Vgl. GLESS, zit. (Fn. 9), S. 435 f.

mit dem Problem der Gewährung des rechtlichen Gehörs⁵⁸, durchaus bereits eine anfängliche Herausforderung sein. Grundsätzlich ist aber davon auszugehen, dass KI-Programme bei der Feststellung des Sachverhalts zu helfen in der Lage sind⁵⁹, wie Beispiele aus der Praxis zeigen:

- In Schadenersatzverfahren wegen Flugverspätungen greifen deutsche Gerichte schon heute auf IT-Systeme zu, die Bordkarten, Wetterkarten und vergangene Urteile auswerten, d.h. die Gerichte können sich relevante Informationen automatisiert zuliefern lassen⁶⁰.
- In einem Forschungsprojekt hat die ETH Zürich einen «Enforcement Bot» entwickelt, der Verletzungen des Datenschutz- und Verbraucherrechts automatisiert im Internet erkennen kann; technische Grundlage für die Compliance-Beurteilung sind sog. «Webcrawler»⁶¹, welche (automatisch) Informationen von Webseiten «abkratzen» («scraping») und auf Muster von mutmasslichen Gesetzesverstößen prüfen. Im Falle von Verletzungen erfolgt eine Abmahnung des Rechtsverletzers und ggf. eine Anzeige bei der zuständigen Behörde⁶².

Die bisherigen Erfahrungen zeigen dementsprechend, dass KI-Programme bei der Feststellung des Sachverhalts durchaus wertvolle Dienste anzubieten vermögen.

4.2.2. Auffinden von Rechtssätzen

Das Auffinden und Auswerten möglicherweise anwendbarer Rechtssätze mit Blick auf die Entscheidung in einem bestimmten Rechtsfall ist ein Vorgang, der sich für Automatisierungen grundsätzlich eignet. Bereits heute gibt es das «Swiss Federal Supreme Court Dataset» (SCD)⁶³; «Webcrawler» bieten aufgrund publizierter Datensätze von Bundesgerichtsurteilen sachdienliche Auswertungsergeb-

⁵⁸ Dazu hinten Ziff. 4.4.2.

⁵⁹ AILexA strebt die Aufteilung des Sachverhalts in seine Bestandteile nach der Logik der Vereinfachung und Optimierung an; das Programm soll mittels prädiktiver Analysen anhand vergleichender Algorithmen seine Antwort geben; vgl. dazu R. PERANI, *Intelligenza Artificiale e Digesta Iustiniani*, Milano, 2023.

⁶⁰ Vgl. dazu J. HORN, *Automatische Informationsbeschaffung durch Zivilgerichte*, in *RDi*, 2022, S. 469 f.

⁶¹ Der entwickelte Prototyp analysiert Verletzungen des Datenschutzes; vgl. K. KUBICEK et al., *Checking Websites' GDPR Consent Compliance for Marketing Emails*, in *PoPETs*, 2022, S. 282, 283 ff.

⁶² Vgl. auch GLESS, zit. (Fn. 9), S. 437.

⁶³ Vgl. dazu J. MERANE, F. GEERING, *Bundesgerichtsurteile nutzen: Das Swiss Federal Supreme Court Dataset (SDC)*, in *Weblaw Jusletter*, 18. September 2023.

nisse an⁶⁴. Ohne menschliche Intervention geht es aber regelmässig nicht, weil eine Feinjustierung im Verhältnis zwischen den anwendbaren Rechtsnormen und dem konkreten Sachverhalt meist notwendig ist⁶⁵.

In Deutschland (z.B. FRAUKE) und in den USA sind bereits entsprechende KI-Programme im Einsatz⁶⁶. AILexA ist wie folgt aufgebaut: Eine Person (actor) erkundigt sich nach dem möglichen Ergebnis (responsum), welches sich als Reaktion auf ein bestimmtes Verhalten (factum) von Dritten (agens) ergeben und ihre Interessensphäre in irgendeiner Weise (victima) stören kann⁶⁷. Bei statistischen Explorationen nimmt AILexA gleichzeitig eine Vorhersageanalyse auf verschiedenen Ebenen vor.

4.2.3. Anwendung von Rechtsnormen

Die Anwendung der einschlägigen Rechtsnormen auf den gegebenen Sachverhalt und die Begründung des Entscheids bieten die grössten Herausforderungen für den Einsatz von KI-Programmen in der Justiz. Vorerst hängt der Erfolg des Einsatzes automatisierter Programme von der Komplexität der Rechtsfrage(n) ab; je schwieriger die Gemengelage der zu entscheidenden Rechtsprobleme ist, desto weniger eignen sich KI-Programme⁶⁸. Gewisse Vorhersagen, wie dies auch bei AILexA der Fall ist, sind indessen regelmässig möglich, aber es ist genau zu analysieren, wie hoch die mögliche Fehlerquote ist⁶⁹. Regelbasierte Systeme (Entscheidungsbäume, Expertensysteme) vermögen dabei eine wertvolle, auch automatisierte, Unterstützung zu bieten⁷⁰.

Gesamthaft betrachtet ist somit festzuhalten, dass sich in allen Ablaufschritten bzw. Verfahrensstadien (auch) KI-Programme einsetzen lassen, dass deren

⁶⁴ Vgl. MERANE, GEERING, zit. (Fn. 63), Rz 11 ff.; vgl. auch I. GLASER, P. MOSER, F. MATTHES, *Generation of Legal Norm Chains: Extracting the Most Relevant Norms from Court Rulings*, in *Legal Knowledge and Information Systems* (Hrsg. E. SCHWEIGHOFER), Amsterdam, 2021, S. 43, 45 ff.; aus dem anglo-amerikanischen Rechtskreis vgl. S. VANDEN BROUCKE, B. BAESENS, *Practical Web Scraping for Data Science: Best Practices and Examples with Python*, Berkeley, 2018, S. 25 ff.; N. DE VITO et al., *Tools That Ease Data Collection From the Web*, in *Nature*, 2020, S. 621 f.

⁶⁵ Vgl. auch GLESS, zit. (Fn. 9), S. 437.

⁶⁶ Vgl. dazu A. SESING-WAGENPFEIL et al., *Legal Tech im Richterzimmer? Streiflichter aus Wissenschaft und Praxis zum KI-Einsatz bei Flugverspätungen*, in *Weblaw Jusletter IT*, 30. März 2023, Rz 11 ff.

⁶⁷ Vgl. Fn. 59.

⁶⁸ Vgl. GLESS, zit. (Fn. 9), S. 438 f.

⁶⁹ In den USA sind bereits empirische Studien zur «Legal Judgment Prediction» vorhanden; vgl. die Nachweise bei GLESS, zit. (Fn. 9), S. 439 Fn. 52 und 54.

⁷⁰ Für weitere Einzelheiten vgl. GLESS, zit. (Fn. 9), S. 440.

Nutzen jedoch von der konkreten Situation und vom gegebenen Komplexitätsgrad der in Frage stehenden Justizangelegenheit abhängig ist. Zudem muss jedenfalls sichergestellt sein, dass die Rahmenbedingungen der Rechtsordnung eingehalten werden, insbesondere die datenschutzrechtlichen Vorgaben⁷¹, aber auch die Bestimmungen des Lauterkeitsrechts zum Schutz vor der unmittelbaren Übernahme und Verwendung eines marktreifen Arbeitsergebnisses (Art. 5 lit. c UWG)⁷².

Die eingesetzten KI-Programme können mit Bezug auf die Entscheidungsfindung auch zu einer grösseren Transparenz und Nachvollziehbarkeit⁷³ zugunsten der Verfahrensbeteiligten führen. Ein solche Entwicklung lässt sich durch einen sog. Legal-BERT konkretisieren⁷⁴; Ziel ist eine möglichst umfassende Wiedergabe aller möglichen Daten, die für eine Entscheidungsfindung von Bedeutung sind⁷⁵. Entsprechende Datensammlungen sind zudem nicht nur für die Justiz als Unterstützungsinstrumente wertvoll, sondern auch für private Unternehmen, die sich im Bereich der Rechtsdurchsetzung engagieren (z.B. die Plattform «Flightright», die Entschädigungen von Passagieren gegen Fluggesellschaften automatisiert durchsetzt)⁷⁶.

4.3. Vorteile der Automatisierung

Als Chancen des Einsatzes von KI-Programmen und von algorithmischen Verfahren stehen nach verbreiteten Meinungsäusserungen in der Doktrin folgende effizienzfördernden Ziele im Vordergrund:

- *Erwartung von kürzeren und kostengünstigeren Verfahren*: Die Effizienzgewinne durch Verfahrensstraffungen, sofern die verfahrensrechtlichen Grundprinzipien (z.B. rechtliches Gehör, Nichtdiskriminierung, usw.⁷⁷) eingehalten sind, dürften wohl ein unbestrittener Vorteil einer Automatisierung sein⁷⁸. Auch AILexA hat die Fähigkeit, in Sekundenbruchteilen alle Inhalte zu verarbeiten, indem unter allen vorhandenen Datensätzen jeweils die am besten geeignete Lösung gesucht wird. Der Effizienz-Aspekt ist übrigens kürzlich im Strafverfahren Vincenz und Konsorten sowohl von der Ankläger-Seite als auch von der

⁷¹ Vgl. vorne Ziff. 3.2.

⁷² Vgl. auch PINA, zit. (Fn. 28), S. 19 f.

⁷³ Vgl. vorne Ziff. 3.1.

⁷⁴ BERT steht für Bidirectional Encoder Representations from Transformers.

⁷⁵ Vgl. auch GLESS, zit. (Fn. 9), S. 441 und 443.

⁷⁶ Vgl. <https://www.flightright.de/>.

⁷⁷ Dazu hinten Ziff. 4.4.2.

⁷⁸ Vgl. vorne Ziff. 2.1 und GLESS, zit. (Fn. 9), S. 445 f.

Betroffenen-Seite angesprochen worden ⁷⁹.

- *Präzisere Rechtsanwendung*: Die automatisierte Subsumtion, basierend auf empirisch gesicherten Vorgaben sowie auf breiten, systematisch abgesuchten Informationssammlungen, ist ggf. der menschlichen Entscheidungsfähigkeit überlegen; eine nachträgliche Plausibilitätskontrolle erweist sich aber in der Regel dennoch als erforderlich ⁸⁰. Immerhin lässt sich nicht übersehen, dass es bereits, insbesondere in den USA, recht präzise Vorhersagemodelle für mögliche Prozessergebnisse gibt ⁸¹.
- *Größere Einheitlichkeit der Rechtsanwendung*: Ja mehr Daten von früheren und parallelen Verfahren automatisch gesammelt und in KI-Programmen verarbeitet werden, umso einheitlicher dürfte die Rechtsanwendung durch die Gerichte sein ⁸². Diese Einschätzung gilt schon für die nationale Rechtsprechung, noch viel mehr aber für die internationalen Gremien, die (verbindliche und unverbindliche) Entscheidungen fällen ⁸³.
- *Ausschluss emotionaler menschlicher Elemente im Entscheidungsprozess*: Wenn Menschen arbeiten, ist es unumgänglich, dass auch personenbezogene Einflussfaktoren eine Rolle spielen. Beteiligte eines Gerichtsverfahrens können mehr oder weniger sympathisch sein, geäußerte Wertvorstellungen entsprechen mehr oder weniger den eigenen Ideen ⁸⁴. Obwohl das Sachlichkeitsgebot von der Justiz fordern würde, dass solche Einschätzungen für die Entscheidungsfindung nicht relevant sein dürften, ist deren vollständige «Verbannung aus den Gerichtsräumen» eine Illusion ⁸⁵. Automatisierungen und von KI-Programmen «produzierte» Ergebnissen können deshalb menschlichen Risikofaktoren durchaus entgegenwirken.

Verschiedene Effizienzvorteile, die sich aus der Automatisierung ergeben, sind also dementsprechend beachtlich, d.h. algorithmische Verfahren und KI-Programme vermögen in der Justiz positive Wirkungen herbeizuführen, wenn sie sachgerecht eingesetzt werden; diese Vorteile sind indessen gegen die Nachteile der Automatisierung abzuwägen.

⁷⁹ Vgl. vorne Fn. 49.

⁸⁰ Vgl. GLESS, zit. (Fn. 9), S. 446.

⁸¹ Vgl. vorne Fn. 69.

⁸² Vgl. GLESS, zit. (Fn. 9), S. 447.

⁸³ Der EMGR wendet schon Programme an, die den Charakter von Voraussagesystemen haben.

⁸⁴ Vgl. J. RISSE, *Der Homo iuridicus - ein gefährliches Trugbild*, in *NJW*, 2018, S. 2848, 2850.

⁸⁵ Eingehender dazu GLESS, zit. (Fn. 9), S. 448 f.

4.4. Kontrolle von Risiken bei der Automatisierung

4.4.1. Maschinen statt Menschen?

Der Einsatz von KI-Programmen bzw. algorithmischen Modellen in der Justiz verursacht auch spezifische Herausforderungen, die über die allgemeinen, durch automatisierte Verfahren verursachten Risiken hinausgehen; deren Erörterung ist in den bisherigen Diskussionen im Vordergrund gestanden, was nicht überrascht, weil in einer vereinfachten Betrachtungsweise die Menschen durch Maschinen ersetzt werden sollen. Die Themen sind zumindest teilweise bereits einlässlich analysiert worden, so etwa vom Autor vor knapp fünf Jahren unter dem Titel «Dürfen Maschinen über Menschen entscheiden?»⁸⁶ und kürzlich von Sabine Gless mit der Frage, ob es weiter Menschen in der Justiz brauche⁸⁷.

Die Problematik der bisherigen Untersuchungen und Regelungsbemühungen liegt aber darin, dass sie, was für rechtliche Abhandlungen nicht untypisch ist, abstrakt bleiben und empirische Analysen weitgehend fehlen⁸⁸. Insoweit besteht somit ein nicht zu unterschätzender wissenschaftlicher Handlungsbedarf.

4.4.2. Einzelne Risiken

Als Risiken im Falle einer Automatisierung in der Justiz sind insbesondere die folgenden Aspekte identifiziert worden⁸⁹:

- (i) Gerichtsverfahren zeichnen sich durch eine grosse Zahl von *formell-rechtlichen Vorgaben* aus; deren Nichteinhaltung führt dazu, dass ein gefällttes Urteil anfechtbar ist. Beispiele sind das rechtliche Gehör und die Begründung des Urteils⁹⁰. Gewisse KI-Hilfsmittel sind einsetzbar, z.B. «Chatbots» oder «Explainers»⁹¹, doch ist deren Präzision bisher empirisch noch nicht ausreichend nachgewiesen. Immerhin lässt sich nicht übersehen, dass die Entscheidungs-

⁸⁶ WEBER, zit. (Fn. 10), so der Titel.

⁸⁷ GLESS, zit. (Fn. 9), so im Untertitel.

⁸⁸ Im anglo-amerikanischen Bereich sind mehr empirische Analysen vorhanden (vgl. vorne Fn. 64).

⁸⁹ Eingehend zu den Risiken GLESS, zit. (Fn. 9), S. 449 ff. m.w.H.

⁹⁰ Art. 29 Abs. 2 BV; statt Vieler vgl. BGE 134 I 83 E. 4; zum Begründungsaspekt vgl. N. BRAUN BINDER, L. OBRECHT, *Die Begründung von Verfügungen beim Einsatz algorithmischer Systeme*, in *SJZ*, 120, 2024, S. 707 ff.; allgemein zur KI-Verwendung in der Verwaltung vgl. M. MONTAVON, M. FOROWICZ, *Les décisions automatisées et semi-automatisées en droit administratif suisse*, in *Weblaw Jusletter*, 23. September 2024, Rz 46 ff.

⁹¹ Vgl. immerhin vorne Ziff. 4.2.1 zu den Webcrawlern.

gründe menschlicher Richter oft auch nicht vollständig nachvollziehbar sind⁹².

- (ii) In der Diskussion wird oft auf die *Gefahr von Fehlurteilen* hingewiesen⁹³. Ein solches Risiko besteht insbesondere bei quantitativ oder qualitativ ungenügenden Datensammlungen oder bei mangelhaften Formulierungen von sog. Prompts (je nach Wortlaut ist Antwort unterschiedlich): Der Einsatz von KI-Programmen kann jedenfalls nur zu befriedigenden Ergebnissen führen, wenn alle relevanten Gerichte ihre Urteile digitalisiert in einer zentralen Datenbank abspeichern. Ob diese Voraussetzung heute bereits erfüllt ist, erscheint zumindest als ungewiss.

Zudem fehlt es an grenzüberschreitenden Vorgaben, wie solche Datensammlungen aufzubauen sind; als Ausnahme liegen seit 2022 immerhin Prinzipien für ein Folgeabschätzungsverfahren und ein Genehmigungsverfahren mit Bezug auf die Nutzung von KI-Systemen in der öffentlichen Verwaltung, erarbeitet durch das European Law Institute, vor⁹⁴. Weil KI-Programme zudem mit statistischen Methoden arbeiten müssen, welche für die Lösung von Regelfällen geeignet sind, erscheint als nicht gesichert, dass auch Sonderfälle mit ungewöhnlichen Konstellationen einer richtigen Lösung zugeführt werden⁹⁵. Schliesslich vermag das erwähnte Problem der «Halluzination»⁹⁶ zu Sinnestäuschungen und damit zu falschen Schlüssen zu führen; im Vordergrund steht die Fehleinschätzung mit Bezug auf die eigenen Fähigkeiten («Overconfidence»)⁹⁷, verbunden mit dem Verzicht auf eine kritische Reflexion und Überprüfung des automatisch generierten Ergebnisses⁹⁸. Mit zunehmender Leistungsfähigkeit von KI-Programmen wird es für Menschen auch immer schwieriger, ein falsches Ergebnis von einem richtigen Ergebnis zu unterscheiden⁹⁹.

- (iii) Den automatisierten Entscheidungen in der Justiz stehen auch *verfassungsrechtliche Bedenken* entgegen, insbesondere die Argumente der Gewaltenteilung.

⁹² Vgl. GLESS, zit. (Fn. 9), S. 458.

⁹³ Vgl. etwa H. FRY, *Hello World, Being Human in the Age of Algorithms*, New York, 2018, S. 76 f.

⁹⁴ ELI Model Rules on Impact Assessment of Algorithmic Decision-Making Systems used by Public Administration, <https://www.europeanlawinstitute.eu/projects-publications/completed-projects/ai-and-public-administration/>.

⁹⁵ Vgl. GLESS, zit. (Fn. 9), S. 455 f.

⁹⁶ Vgl. vorne Ziff. 2.2.

⁹⁷ Vgl. vorne Ziff. 3.1.

⁹⁸ Vgl. auch GLESS, zit. (Fn. 9), S. 457.

⁹⁹ Vgl. auch G. RÜHL, *KI in der gerichtlichen Streitbeilegung*, in *Rechtshandbuch Artificial Intelligence and Machine Learning* (Hrsg. M. KAULARTZ, T. BRAEGELMANN), München, 2020, S. 617, Rz 24.

lung und der demokratischen Legitimation, z.B. mit Blick auf die Unabhängigkeit und die Unparteilichkeit der Richter/innen¹⁰⁰. Zweifellos würde der Einsatz umfassender KI-Programme in der Justiz aus verfassungsrechtlicher Sicht zu neuen Herausforderungen führen; dabei kann indessen nicht eine Anpassung der heutigen Rahmenbedingungen im Vordergrund stehen, sondern die zentralen Begrifflichkeiten sind der IT-Charakteristik entsprechend zu umschreiben.

Die mit Bezug auf Menschen verlangte innere Unabhängigkeit, die zwar in der Realität auch nicht immer verwirklicht ist, muss neu gedacht werden, und zwar im Sinne der interessenfreien Programmierung von KI-Systemen, verbunden mit der Möglichkeit, die Parameter regelmässig zu überprüfen (Nachvollziehbarkeit des Entscheidungsvorganges) und ggf. anzupassen¹⁰¹. Vorzusehen ist auch die Möglichkeit des Menschen, Ergebnisse von IT-Systemen, die nicht als plausibel erscheinen, abzulehnen und durch eigene Erkenntnisse zu ersetzen¹⁰². Um die Gewähr für die Unabhängigkeit und Unparteilichkeit von KI-Programmen in der Justiz zu erhöhen, wäre auch daran zu denken, deren Zulassung zur automatisierten Rechtsanwendung an eine öffentlich-rechtliche Zertifizierung zu knüpfen, welche die nach Auffassung der Öffentlichkeit bzw. der politischen Entscheidungsträger relevanten Anforderungen gewährleistet¹⁰³. Zudem sollte kein Zwang bestehen, die von KI-Programmen vorgesehenen Ergebnisse in jedem Fall übernehmen zu müssen.

- (iv) Als *weitere Risiken* werden in der Diskussion die «Versteinerung» der Rechtsentwicklung¹⁰⁴ und der Wegfall von menschlicher Kompetenz vorgebracht¹⁰⁵. Zwar dürften KI-Programme, selbst wenn die Daten regelmässig und sachgerecht aktualisiert werden¹⁰⁶, zum Teil weniger innovativ sein als Menschen, doch ist die individuelle Bereitschaft, neue rechtliche Lösungen zu entwickeln, auch nicht zu überschätzen. Die menschliche Kompetenz ist selbstredend wichtig und muss auch beim Einsatz von KI-Programmen z.B. im Sinne einer Plausibilitätskontrolle ausgeübt werden; ihre teilweise Zurückdrängung hat

¹⁰⁰ Vgl. A. BOOS, N. POMPE, *Die demokratische Legitimität der Anwendung Künstlicher Intelligenz in der Rechtsprechung*, in *Recht und Innovation, Innovation durch Recht, im Recht und als Herausforderung für das Recht* (Hrsg. J. MEIER et al.), Zürich, 2020, S. 243, 250 ff.

¹⁰¹ Zur Nachvollziehbarkeit vgl. vorne Ziff. 3.1.

¹⁰² GLESS, zit. (Fn. 9), S. 450 f. m.w.H.

¹⁰³ Dazu S. GLESS, W. WOHLERS, *Subsumtionsautomat 2.0 - Künstliche Intelligenz statt menschlicher Richter?*, in *Festschrift für Urs Kindhäuser* (Hrsg. M. BÖSE et al.), Baden-Baden, 2019, S. 147, 163 f.

¹⁰⁴ Dazu GLESS, zit. (Fn. 9), S. 459.

¹⁰⁵ Dazu GLESS, zit. (Fn. 9), S. 460.

¹⁰⁶ Vgl. PINA, zit. (Fn. 28), S. 52 f.

aber auch den Vorteil, dass der erwähnte Ausschluss menschlicher Emotionen¹⁰⁷ eher möglich ist.

- (v) Gesamthaft betrachtet lassen sich Entscheidungsverfahren in der Justiz somit nicht einfach vollumfänglich automatisieren. Als Unterstützungsinstrument erweisen sich KI-Programme aber schon heute als sinnvolle Hilfsmittel. Zudem ist der erweiterte Einsatz von IT-Systemen angesichts ihrer unbestreitbaren Vorteile nur eine Frage der Zeit; zukunftsgerichtet sind deshalb die Rahmenbedingungen zu entwickeln, die den Einsatz von KI-Systemen in der Justiz unter Beachtung der verfassungs- und prozessrechtlichen Vorgaben ermöglichen.

4.5. Gegenseitige Abwägung von Vorteilen und Risiken

Die Automatisierung der Justizverfahren, neuerdings gestützt auf KI-Programme, ist – wie eingangs erwähnt – zu Recht auch zu einem Diskussionsthema in Kontinentaleuropa geworden. Die Entwicklung verläuft schrittweise, ausgehend von der erwähnten Plattform «Justitia 4.0»¹⁰⁸ über partielle KI-Einsätze bis zur Implementierung ganzer KI-Systeme. Grundsätzlich sind KI-Systeme aber eher als Instrumente oder Werkzeuge, welche die Justizverfahren positiv unterstützen, denn als finale Problemlöser von Entscheidungsverfahren zu sehen.

Risiken wegen ungenügender Transparenz und Datensicherheit sind ernst zu nehmen. Zudem ist menschliches Wissen bei der Feinjustierung zwischen Sachverhalt und anwendbaren Rechtsregeln oft unerlässlich. Immerhin dürfen weit entwickelte KI-Programme schon bald in der Lage sein, relevante Informationen zu extrahieren, Zusammenhänge zu erkennen und Muster zu identifizieren, um die einschlägigen Rechtsgrundlagen zur Anwendung zu bringen, jedenfalls wenn ausreichend grosse und qualitativ hochwertige Datengrundlagen zur Verfügung stehen¹⁰⁹.

Der Einsatz von KI-Programmen drängt sich vor allem in stark standardisierten Bereichen auf, also mit Bezug auf Fälle, in denen die Rechtsanwendung auf wenige Vorgaben reduziert ist; die menschliche Interventionsmöglichkeit wird in dessen unumgänglich bleiben. Für die autonome Rechtsanwendung sind insbesondere die generative KI und die LLM von Bedeutung; deren Einsatz dürfte sich in der Justiz in den nächsten Jahren vermehrt durchsetzen. Gerade diese Einschätzung bedeutet, dass wissenschaftlich eine verstärkte Bearbeitung von KI-bezogenen Themen als sehr erwünscht erscheint; neben rechtlichen Analysen sind aber

¹⁰⁷ Vgl. vorne Ziff. 4.3 mit Fn. 84/85.

¹⁰⁸ Vgl. vorne Ziff. 4.1.

¹⁰⁹ Vgl. auch WEBER, zit. (Fn. 1), S. 25.

auch empirische Daten-Untersuchungen unabdingbar, um sachgerechte Entscheidungen zum KI-Einsatz treffen zu können.

5. AUSBLICK

Die Technologie wird sich unaufhaltsam weiterentwickeln. Ein grundsätzlicher Widerstand gegen eine bzw. jegliche Automatisierung von Justizverfahren ist deshalb nicht zukunftsweisend. Offenheit gegenüber neuen methodischen Ansätzen und Verknüpfungen mit komplexen KI-Modellen ist somit geboten. In einem ersten Schritt lassen sich dadurch etwa «Urteilkonfigurationen», wie sie in Schadenersatzverfahren in Frankfurt eingesetzt werden, entwickeln (z.B. System «FRAUKE») ¹¹⁰.

Justizorgane sind auch eingeladen, zur Qualitätssteigerung an der sachgerechten Aufarbeitung von Trainingsdaten mitzuwirken (in Form eines sog. «LegalBERT» als spezialisiertes Sprachmodell, das Akten nach bestimmten Kriterien sowie nach Namen und Begriffen durchsucht) ¹¹¹. Dänemark kennt das Modell «Explainable AI and the Law», das sich zur Vorprüfung von Fällen einsetzen lässt. Weiterentwicklungen und Anpassungen an die Schweizer Rechtslandschaft sind durchaus denkbar.

Wie bereits erläutert, ist die Nachvollziehbarkeit der IT-Vorgänge und der KI-Systeme von grosser Bedeutung, d.h. eine «Explainable AI» muss transparent machen, wie ein bestimmter Rechtsanwendungsvorschlag zustande gekommen ist ¹¹². Die semi-autonome, KI-unterstützte Anonymisierung von Personendaten in Gerichtsurteilen ist ein weiteres gutes Beispiel ¹¹³.

Der richtige Weg sollte also dahin gehen, im Einzelnen zu prüfen, in welchen Bereichen der Einsatz von KI-Systemen eine Entlastung für die Justiz im Sinne einer Ressourcenoptimierung bringen kann (z.B. Automatisierung schematisierter Rechtsbereiche) und wie das Design einer Automatisierung auszugestalten ist, um die notwendige Qualität der Rechtsprechung sicherzustellen ¹¹⁴.

¹¹⁰ Vgl. vorne Fn. 66.

¹¹¹ Vgl. vorne Fn. 74.

¹¹² Vgl. dazu L. MALANDRI et al., *The Good, the Bad, and the Explainer: A Tool for Contrastive Explanations of Text Classifiers*, in *IJCAI*, 2022, S. 5936 ff.

¹¹³ Vgl. auch GLESS, zit. (Fn. 9), S. 461.

¹¹⁴ Vgl. auch N. BRAUN BINDER, *Künstliche Intelligenz und automatisierte Entscheidungen in der öffentlichen Verwaltung*, in *SJZ*, 2019, S. 467, 472 ff.