www.jusletter.ch

Nadja Braun Binder / Thomas Burri / Melinda Florina Lohmann / Monika Simmler / Florent Thouvenin / Kerstin Noëlle Vokinger

Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht

Beeindruckende Fortschritte im Bereich der Künstlichen Intelligenz (KI) haben in den letzten Jahren für Aufsehen gesorgt, aber auch Ängste geweckt. Für Gesetzgeber stellt sich damit die Frage, ob und wie sie den Herausforderungen von KI begegnen wollen. Am 21. April 2021 hat die EU-Kommission einen Vorschlag für eine Verordnung zur Regulierung von KI präsentiert. Dieser Beitrag zeigt den Handlungsbedarf auf, der im Schweizer Recht besteht, und versteht sich als Anstoss für eine vertiefte Diskussion und als Aufruf an den Schweizer Gesetzgeber, die Erarbeitung eines Rechtsrahmens zur Erfassung der Herausforderungen von KI zeitnah anzugehen.

Beitragsart: Wissenschaftliche Beiträge
Rechtsgebiete: Informatik und Recht

Zitiervorschlag: Nadja Braun Binder / Thomas Burri / Melinda Florina Lohmann / Monika Simmler / Florent Thouvenin / Kerstin Noëlle Vokinger, Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht, in: Jusletter 28. Juni 2021

Inhaltsübersicht

1. Einleitung

[1] Innovation ist für ein Land wie die Schweiz von zentraler Bedeutung. Das gilt nicht nur für den privaten Sektor, sondern auch für die öffentliche Hand. Nicht wenige hoffen, dass wir die grossen Herausforderungen unserer Zeit – von der Pandemie über den Terrorismus bis zum Klimawandel – mithilfe von (technischen) Innovationen bewältigen können. Im Bereich der Informationstechnologie, die seit Jahren der wichtigste Treiber von Innovation in einer Vielzahl von Sektoren ist, haben in jüngerer Zeit vor allem die Distributed Ledger Technology (DLT, insb. Blockchain) und die Fortschritte im Bereich der Künstlichen Intelligenz (KI) für Aufsehen gesorgt.

[2] Innovationen eröffnen nicht nur Chancen, sondern sind regelmässig auch mit Risiken verbunden. Neuen Risiken kann teilweise durch geeignete technische Massnahmen begegnet werden, sie erfordern aber oftmals auch eine rechtliche Einhegung, sei es durch Erlass neuer Normen oder eine neue Auslegung und Anwendung des geltenden Rechts. Auf die Chancen und Risiken der DLT hat der Schweizer Gesetzgeber erstaunlich rasch reagiert.[1] Dies hat wesentlich dazu beigetragen, dass sich die Schweiz im internationalen Wettbewerb als Standort für DLT-Anwendungen («krypto-valley») etablieren und behaupten konnte. Im Bereich der KI hat sich die Schweiz dagegen weitgehend passiv verhalten. Für die interessierte Öffentlichkeit ist bisher nur ersichtlich, dass eine interdepartementale Arbeitsgruppe des Bundes im Dezember 2019 einen Bericht zu den Herausforderungen von KI vorgelegt[2] und der Bundesrat im November 2020 Leitlinien für den Umgang mit KI in der Bundesverwaltung verabschiedet hat.[3] Auch einzelne Kantone haben sich bereits mit dem Einsatz von KI in der Verwaltung befasst.[4] Immerhin beteiligt sich die Schweiz an einer Arbeitsgruppe des Europarates, dem «Ad hoc Committee on Artificial Intelligence (CAHAI)», welches die Machbarkeit und die möglichen Inhalte eines Rechtsrahmens für

1 Siehe dazu: Beschluss des Parlaments über das Bundesgesetz zur Anpassung des Bundesrechts an Entwicklungen der Technik verteilter elektronischer Register vom 25. September 2020, BBl 2020 7801–7820, https://www.admin.ch/opc/de/federal-gazette/2020/7801.pdf (7. Juni 2021); Botschaft zum Bundesgesetz zur Anpassung des Bundesrechts an Entwicklungen der Technik verteilter elektronischer Register vom 27. November 2019, BBl 2020 233–328.

2 Interdepartementale Arbeitsgruppe Künstliche Intelligenz, Herausforderungen der künstlichen Intelligenz, Bericht an den Bundesrat, 13. Dezember 2019.

3 Leitlinien «Künstliche Intelligenz» für den Bund, Orientierungsrahmen für den Umgang mit künstlicher Intelligenz in der Bundesverwaltung, 25. November 2020.

4 Namentlich hat der Kanton Zürich einen entsprechenden Bericht erstellen lassen; siehe Nadja Braun Binder/Matthias Spielkamp/Catherine Egli/Laurent Freiburghaus/Eliane Kunz/Nina Laukenmann/Michele Loi/Anna Mätzener/Liliane Obrecht/Jessica Wulf, Einsatz künstlicher Intelligenz in der Verwaltung: rechtliche und ethische Fragen, Schlussbericht vom 28. Februar 2021, https://www.zh.ch/de/politik-staat/kanton/kantonale-verwaltung/digitale-verwaltung/digitalisierungsprojekte.html#-718112641 (5. Juni 2021).

Entwicklung, Design und Anwendung von KI untersucht.[5] Nachdem der Bericht der interdepartementalen Arbeitsgruppe vom Dezember 2019 im Wesentlichen zum Schluss gekommen ist, in der Schweiz bestehe kein regulatorischer Handlungsbedarf,[6] ist wenig erstaunlich, dass der Gesetzgeber im Bereich der KI bisher nicht aktiv geworden ist. Die einzige Ausnahme betrifft – soweit ersichtlich – die Arbeiten an einem Rechtsrahmen für den Einsatz von KI in der automatisierten Mobilität.[7]

[3] Das Vorgehen der Schweiz unterscheidet sich damit grundlegend vom Ansatz der EU, in der seit mehreren Jahren intensiv an einer Regulierung von KI gearbeitet wird. Auf Grundlage umfassender Vorarbeiten hat die EU-Kommission am 21. April 2021 einen Vorschlag für eine Verordnung zur Regulierung von KI[8] präsentiert («EU-KI-Verordnungsvorschlag»). Dieser Schritt war mit Spannung erwartet worden, nachdem das vor etwas mehr als einem Jahr vorgelegte Weissbuch der Kommission zu KI[9] ein grosses Echo ausgelöst hatte. Der Verordnungsvorschlag orientiert sich am Risiko, das von verschiedenen Anwendungen von KI ausgeht, und reguliert primär die risikoreichsten Anwendungen.[10] Bemerkenswert ist, dass mit der EU-KI-Verordnung eine konkrete Technologie reguliert werden soll, entsprechend wird der Begriff der KI im Vorschlag auch definiert. Nach der weit gefassten Definition, die sich am etablierten Begriffsverständnis der Informatik orientiert,[11] umfasst der Begriff der KI im Sinn des EU-KI-Verordnungsvorschlags Ansätze des maschinellen Lernens («machine learning approaches»), logik- und wissensbasierte Ansätze («logic- and knowledge-based approaches») sowie statistische Ansätze, bayessche Schätzungen sowie Such- und Optimierungsverfahren.[12]

[4] Für die Schweiz, die an dem in der EU nun in Gang gesetzten Rechtsetzungsprozess formell nicht teilhaben wird, stellt sich die Frage, wie sie sich in Bezug auf diese Rechtsentwicklung positionieren soll. Soll die Schweiz eine Art «Gegenentwurf» zum sich abzeichnenden EU-Recht entwickeln und ebenfalls ein Gesetz vorlegen, das KI als Technologie reguliert? Soll sie die Herausforderungen bei KI-Anwendungen in den Blick nehmen und diese durch adäquate Regeln zu erfassen versuchen? Oder soll sie weiter abwarten und darauf vertrauen, dass sich die erkennbaren Risiken von KI nicht in konkreten Schäden manifestieren? Falls sich die Reaktion der Schweiz zu stark vom EU-Recht unterscheiden sollte, könnte unter Umständen der Zugang von Dienstleistungen und Produkten von Schweizer Unternehmen zum Binnenmarkt gefährdet wer-

5 https://www.coe.int/en/web/artificial-intelligence/cahai (14. Juni 2021).

6 Interdepartementale Arbeitsgruppe Künstliche Intelligenz (Fn. 2), 10 ff.

7 Siehe dazu Melinda F. Lohmann, Mobilität von morgen – Die Zulässigkeit automatisierter Fahrzeuge im Ländervergleich, AJP 2021, 617–627, 620 ff.

8 Europäische Kommission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, COM(2021) 206 final, 21. April 2021.

9 Europäische Kommission, White Paper on Artificial Intelligence – A European approach to excellence and trust, COM(2020) 65 final, 19. Februar 2020.

10 Für weitere Details zum Verordnungsvorschlag der Kommission siehe Thomas Burri/Fredrik von Bothmer, The New EU Legislation on Artificial Intelligence: A Primer, Version vom 21. April 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3831424 (20. Mai 2021).

11 Siehe dazu statt vieler Stuart J. Russell/Peter Norvig, Artificial Intelligence: A Modern Approach, 4. Aufl., Hoboken 2020.

12 In der Originalformulierung von Annex I zum EU-KI-Verordnungsvorschlag: «(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning, and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods.»

den. Falls der EU-KI-Verordnungsvorschlag auf neue Zugangshindernisse hinausliefe, könnte die Schweiz ihn allenfalls auch aktiv infrage stellen, bspw. als Handelshemmnis, das gegen das Recht der Welthandelsorganisation verstossen könnte. Wählt die Schweiz einen eigenständigen Weg, kann sie versuchen, ihren Rechtsraum bewusst von einer Regulierung von KI freizuhalten, um Innovation zu fördern. Ähnlich wie im Datenschutzrecht werden bestimmte Wirkungen des auf eine extraterritoriale Anwendung angelegten EU-Rechts in der Schweiz zwar kaum vermeidbar sein,[13] doch könnte man den Schweizer Unternehmen die Wahl lassen, ob und gegebenenfalls ab wann sie sich nach dem EU-Recht richten, um ihre Binnenmarktfähigkeit zu sichern. Durch einen solchen Ansatz könnte sich die Schweiz als innovationsoffener Experimentierraum für KI positionieren.

[5] Welchen Weg auch immer die Schweiz wählen wird, um den Risiken von KI-Anwendungen zu begegnen und sich zum Verordnungsvorschlag der EU zu positionieren, an einer Auseinandersetzung mit dem Bedarf nach einer rechtlichen Erfassung von KI und den möglichen regulatorischen Ansätzen führt kein Weg vorbei.

[6] Im Folgenden soll deshalb der Handlungsbedarf identifiziert werden, der sich durch die Herausforderungen von KI im Schweizer Recht ergibt. Der Verordnungsvorschlag der EU-Kommission wird dabei zwar nicht im Zentrum stehen, aber dennoch als Bezugspunkt dienen, um den Handlungsbedarf und die Möglichkeiten für eigenständige Ansätze im Schweizer Recht zu verdeutlichen. Der Beitrag identifiziert fünf zentrale Felder. Diese werden kurz erläutert und mit anschaulichen Beispielen illustriert, um anschliessend den Handlungsbedarf aufzuzeigen und mögliche Lösungsansätze zu skizzieren. In Abschnitt 2.1 werden Transparenz, Erklärbarkeit und Interpretierbarkeit von KI am Beispiel der medizinischen Diagnostik diskutiert. In Abschnitt 2.2 dient die Gesichtserkennung der Veranschaulichung der Herausforderungen beim Schutz von Privatsphäre, beim Datenschutz und bei der Problematik staatlicher Überwachung. Abschnitt 2.3 widmet sich der Nicht-Diskriminierung, der Fairness und der Verhinderung von Bias und veranschaulicht die Problematik anhand von Nutzungsbedingungen von Plattformen sowie Ergebnissen von Suchalgorithmen. In Abschnitt 2.4 wird die Manipulation mithilfe von KI diskutiert, wofür Empfehlungen abgebende Algorithmen als Anschauungsmaterial dienen. Abschnitt 2.5 widmet sich sodann der Haftung und der Verantwortlichkeit und illustriert die Probleme mit Beispielen aus dem Bereich des automatisierten Fahrens. Der Artikel schliesst mit einer Schlussfolgerung (Abschnitt 3).

[7] Angesichts der äusserst vielschichtigen Anwendungen von KI, der raschen technischen Fortschritte und der regulatorischen Entwicklungen in der EU erhebt dieser Beitrag keinen Anspruch auf Vollständigkeit. Er versteht sich vielmehr als Anstoss für eine vertiefte Diskussion und als Aufruf an den Schweizer Gesetzgeber, sich zeitnah ganz konkret und vertieft mit der Frage zu befassen, ob und gegebenenfalls wie das Schweizer Recht angepasst werden muss, um die Herausforderungen von KI angemessen erfassen zu können.

13 Siehe dazu ausführlich Anu Bradford, The Brussels Effect: How the European Union Rules the World, Oxford 2020.

2. Herausforderungen und Handlungsbedarf

2.1. Transparenz, Erklärbarkeit und Interpretierbarkeit

[8] In der Medizin ist die Früherkennung von akuten, schweren Erkrankungen (bspw. Blutvergiftung oder akute Nierenerkrankung) essenziell, damit die betroffenen Patientinnen und Patienten möglichst rasch therapiert werden können. Häufig werden hierfür klinische Parameter wie Blutdruck, Körpertemperatur oder Herzfrequenz manuell zusammengetragen, kalkuliert und für sog. Früherkennungssysteme («early warning scores») verwendet. Diese ermöglichen eine frühzeitige und adäquate Diagnosestellung. In den letzten Jahren wurden – insbesondere in den USA – zunehmend Instrumente entwickelt, die basierend auf der Auswertung klinischer Parameter mithilfe von KI aus elektronischen Patientendossiers schwere Erkrankungen prognostizieren bzw. diagnostizieren können. Häufig handelt es sich bei diesen Systemen um «black boxes», d. h., es ist nicht klar, wie bzw. basierend auf welchen Parametern solche Systeme zu ihren Voraussagen und Diagnosen gelangen.[14]

[9] Der Einsatz von KI in medizinischen Instrumenten ist nach dem Verordnungsvorschlag der Kommission mit einem hohen Risiko behaftet.[15] Daher finden die nach dem Vorschlag von hoch riskanter KI zu erfüllenden Bedingungen Anwendung, insbesondere jene, die die Transparenz sicherstellen sollen.[16] Transparenz soll in diesem Fall nicht nur gewährleisten, dass sich Menschen bewusst sind, mit einer KI-Anwendung zu interagieren. Die Nutzerinnen und Nutzer sollen vielmehr auch die Möglichkeit haben, zu verstehen, mit welchen Daten die KI trainiert wurde und wie sie technisch funktioniert und wirkt. Mit anderen Worten muss KI erklärbar und interpretierbar sein, damit sie entsprechend eingesetzt werden kann. Solche Aspekte werden in der Literatur deshalb auch als Erklärbarkeit bzw. Interpretierbarkeit bezeichnet.[17]

[10] Bei den Anforderungen an die Erklärbarkeit und Interpretierbarkeit sind verschiedene Personengruppen zu unterscheiden, die mit einer KI interagieren. Aus verschiedenen – insbesondere technischen sowie gesellschaftlichen – Perspektiven mag es zwar wünschenswert erscheinen, das Funktionieren von KI im Einzelnen nachvollziehen zu können, insbesondere dann, wenn sie auf neuronalen Netzen beruht. Diesem Bedürfnis stehen aber nicht nur technische Hürden, sondern auch die Geheimhaltungsinteressen derjenigen Unternehmen entgegen, welche die infrage stehende KI entwickelt haben. Aus rechtlicher Sicht geht es deshalb meist nicht darum, eine (mehr oder minder umfassende) technische Erklärbarkeit und Interpretierbarkeit anzustreben. Vielmehr sollen die mit einer KI interagierenden Personengruppen in jeweils unterschiedlichem Masse in die Lage versetzt werden, das Funktionieren der KI in den für sie relevanten Grundzügen zu verstehen. So muss bspw. eine Ärztin oder ein Arzt die Zuverlässigkeit einer auf KI beruhenden radiologischen Interpretation einschätzen können, um auf dieser Grundlage eine

14 Siehe Simon Meyer Lauritsen et al., Explainable artificial intelligence model to predict acute critical illness from electronic health records, Nature Communications 2020, https://www.nature.com/articles/s41467-020-17431-x (26. Mai 2021); siehe dazu auch Marco Tulio Ribeiro/Sameer Singh/Carlos Guestrin, «Why Should I Trust You?»: Explaining the Predictions of Any Classifier, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 2016, https://arxiv.org/abs/1602.04938 (22. Juni 2021).

15 Art. 5 in Verbindung mit Anhang 2A EU-KI-Verordnungsvorschlag.

16 Art. 13 EU-KI-Verordnungsvorschlag.

17 John Zerilli et al., Transparency in Algorithmic and Human Decision-Making: Is there a Double Standard?, Philosophy and Technology, 32(4), 5. September 2018, 661–683, 663; Mario Martini, Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, Berlin 2019, 41–47 & 189–207; Andrew D. Selbst/Solon Barocas, The Intuitive Appeal of Explainable Machines, Fordham Law Review 2018, 1085–1139, passim.

hinreichend sichere Diagnose zu stellen. Dies erfordert zwar ein vertieftes Verständnis der Technologie und ihrer Leistungsgrenzen, jedoch muss die Erklärbarkeit die Ärztin oder den Arzt nicht in die Lage versetzen, selbst ein entsprechendes System zu entwickeln. Für die Patientin oder den Patienten dürfte es hingegen in der Regel meist genügen, zu erfahren, dass die Diagnose auch auf dem Einsatz von KI beruht.[18]

[11] Im europäischen Recht wurden die hier aufgezeigten Herausforderungen – zumindest mit Blick auf automatisierte Entscheidungen – schon lange vor dem EU-KI-Verordnungsvorschlag identifiziert. Die Datenschutzgrundverordnung (DSGVO) enthält deshalb für automatisierte Entscheidungen weitreichende Regelungen, welche die Transparenz durch Informationspflichten (Art. 13 und 14 DSGVO), ein Auskunftsrecht (Art. 15 DSGVO) und spezifische Rechte beim Einsatz von automatisierten Entscheidungen sicherstellen, sofern diese rechtliche Wirkungen entfalten oder eine betroffene Person in anderer Weise erheblich beeinträchtigen (Art. 22 DSGVO). Dabei handelt es sich zwar nicht um Regelungen, die spezifisch auf KI abzielen; erfasst werden vielmehr alle Entscheidungen, die ganz ohne menschliche Intervention getroffen werden. Allerdings dürften solche Entscheidungen bereits heute weitgehend und in Zukunft noch vermehrt auf dem Einsatz von KI beruhen.

[12] Ähnliche Vorgaben wie in der DSGVO wurden auch in der Schweiz ins totalrevidierte Datenschutzgesetz (revDSG) aufgenommen. Dieser wird namentlich eine Regelung zu Informationspflichten bei automatisierten Entscheidungen (Art. 21 revDSG) und ein entsprechendes Auskunftsrecht (Art. 25 Abs. 2 lit. f revDSG) enthalten.[19] Diese Bestimmungen haben in der rechtswissenschaftlichen Literatur eine Diskussion über einen Mangel an eindeutigen rechtlichen Handlungsanweisungen entfacht.[20] Auch wenn verschiedene Massnahmen vorgeschlagen wurden, um die Erklärbarkeit sicherzustellen,[21] bleibt weitgehend unklar, wie den rechtlichen Anforderungen Genüge getan werden kann.[22] Für das europäische Recht präzisiert der EU-KI-Verordnungsvor-

---

18 Art. 13 EU-KI-Verordnungsvorschlag enthält den hier gemachten Ausführungen entsprechende Vorgaben, die auch die Bedürfnisse von Ärztinnen und Ärzten abdecken. Ob allerdings Patientinnen und Patienten auf der Grundlage des EU-KI-Verordnungsvorschlags entsprechende Informationsansprüche erheben könnten, bleibt zweifelhaft (siehe Art. 3 Ziff. 4, die den persönlichen oder nicht beruflichen Gebrauch ausnimmt). Zudem ist unklar, ob sie im hier diskutierten Beispiel überhaupt mit KI interagieren, wie es Art. 53 EU-KI-Verordnungsvorschlag verlangen würde.

19 Näheres zum Schweizer Recht: Florent Thouvenin/Alfred Früh, Automatisierte Entscheidungen: Grundfragen aus der Perspektive des Privatrechts, SZW 2020, 3–17; Nadja Braun Binder, Automatisierte Entscheidungen: Perspektive Datenschutzrecht und öffentliche Verwaltung, SZW 2020, 27–34; Dies., Künstliche Intelligenz und automatisierte Entscheidungen in der öffentlichen Verwaltung, SJZ 2019, 467–476; Florent Thouvenin/Alfred Früh/Damian George, Datenschutz und automatisierte Entscheidungen, in: Jusletter 26. November 2018. Zur allgemeinen Diskussion der Datenschutzgrundverordnung: Lilian Edwards, Data Protection and ePrivacy: From Spam and Cookies to Big Data, Machine Learning and Profiling, in: Lilian Edwards (Hrsg.), Law, Policy and the Internet, Oxford 2018, 119–164; Ronald Leenes/Silvia de Conca, Artificial intelligence and privacy: AI enters the house through the cloud, in: Woodrow Barfield/Ugo Pagallo (Hrsg.), Research handbook on the law of artificial intelligence, Cheltenham 2018, 280–306.

20 Lee A. Bygrave, Minding the Machine v2.0: The EU General Data Protection Regulation and Automated Decision-Making, in: Karen Yeung/Martin Lodge (Hrsg.), Algorithmic Regulation, Oxford 2019, 246–260; Thouvenin/Früh (Fn. 19); Rolf H. Weber/Simon Henseler, Regulierung von Algorithmen in der EU und in der Schweiz: Überlegungen zu ausgewählten Regulierungsthemen, Zeitschrift für Europarecht 2020, 28–42.

21 Sandra Wachter/Brent Mittelstadt/Chris Russell, Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR, Harvard Journal of Law & Technology 2018, 841–887, 861–879; Martini (Fn. 17), 189–207 & 340–347.

22 Bryce Goodman/Seth Flaxman, European Union regulations on algorithmic decision-making and a «Right to Explanation», AI Magazine 2017, 50–57, 55 f.; Lilian Edwards/Michael Veale, Slave to the Algorithm? Why a «Right to an Explanation» Is Probably Not the Remedy You Are Looking For, Duke Law & Technology Review 2017, 18–84, 54–65 & 81; Sandra Wachter/Brent Mittelstadt/Luciano Floridi, Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation, International Data Privacy Law 2017, 76–99, 98 f.

schlag im Rahmen der Anforderungen an die Transparenz von KI mit hohem Risiko,[23] was in der DSGVO hinsichtlich der Offenlegung der Logik bei automatisierten Entscheidungen verlangt wird. Eine solche Präzisierung fehlt aber für das revidierte Schweizer Datenschutzgesetz.

[13] Die Anwendungsbereiche der DSGVO und des Schweizer Datenschutzgesetzes sind – im Gegensatz zu jenem des EU-KI-Verordnungsvorschlags – auf die Verarbeitung personenbezogener Daten beschränkt. Da KI zwar häufig, aber nicht immer auf der Bearbeitung personenbezogener Daten beruht, vermag das Datenschutzrecht die Transparenz bei KI nicht lückenlos sicherzustellen. In medizinischen Anwendungen dürfte dies zwar regelmässig der Fall sein, bei anderen Erscheinungsformen, wie etwa beim raum- bzw. zeitbezogenen «Predictive Policing», aber kaum.[24] Als Predictive Policing oder vorausschauende Polizeiarbeit werden neue Formen der Polizeiarbeit bezeichnet, die anhand statistischer Prognosen wahrscheinliche Vorfälle identifizieren und der Polizei damit das Ergreifen präventiver Massnahmen ermöglichen.[25] Während beim personenbezogenen Predictive Policing die Gefährlichkeit bzw. die Gefährdung von Personen eruiert wird, beziehen sich raum- bzw. zeitbezogene Prognosen auf die Frage, wo bzw. in welchem Zeitraum eine bestimmte Gefahr auftreten könnte.[26]

[14] Ein weiteres Problem liegt darin, dass das Datenschutzrecht *de lege lata* kaum dazu beiträgt, dass die Daten, mit denen eine KI trainiert wird, repräsentativ und qualitativ hochstehend sind. Vielmehr besteht sogar die Gefahr einer umgekehrten Wirkung, weil der datenschutzrechtliche Grundsatz der Datenminimierung[27] mit dem Ansatz kollidiert, die Leistung von KI durch das Training mit möglichst vielen Daten zu verbessern.

[15] Für die Nutzer von KI-Anwendungen – im obigen Beispiel Ärztinnen und Ärzte bzw. unter besonderen Umständen auch deren Patientinnen und Patienten – ist Transparenz über das Niveau der DSGVO bzw. des revidierten Datenschutzgesetzes hinaus essenziell, um zu verstehen, mit welchen Daten die KI trainiert wurde und wie sie aufgebaut ist. Die Einhaltung der entsprechenden Vorgaben im EU-KI-Verordnungsvorschlag sollte den Nutzern ermöglichen, im Einzelfall informierte Entscheidungen über den Einsatz eines auf KI beruhenden Instruments zu treffen. In der Schweiz fehlen diesbezügliche gesetzgeberische Konkretisierungen.

---

23 Art. 13 Abs. 1 Satz 1 EU-KI-Verordnungsvorschlag lautet: «High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately».

24 Florent Thouvenin et al., Towards Principled Regulation of Automated Decision-Making (ADM) – A Workshop Report, Zürich 2019, https://www.itsl.uzh.ch/dam/jcr:edba006c-8452-4ffc-bbb0-fd8fde62d114/Workshop%20Report%20Lavin.pdf (26. Mai 2021), 1.

25 Jennifer Pullen/Patricia Schefer, Predictive Policing – Grundlagen, Funktionsweise und Wirkung, in: Monika Simmler (Hrsg.), Smart Criminal Justice, Basel 2021, 103–143, 105.

26 Tobias Knobloch, Vor die Lage kommen: Predictive Policing in Deutschland, Berlin 2018, https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/predictive.policing.pdf (11. Juni 2021), 13 & 17 f.; siehe dazu auch George Mohler/Martin B. Short/Sean Malinowski/Mark Johnson/George E. Tita/Andrea L. Bertozzi/P. Jeffrey Brantingham, Randomized Controlled Field Trials of Predictive Policing, Journal of the American Statistical Association 2015, 1399–1411; Jessica Saunders/Priscillia Hunt/John S. Hollywood, Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot, Journal of Experimental Criminology 2016, 347–371.

27 Art. 5(1)(c) DSGVO. In der Schweiz ergibt sich der Grundsatz der Datenminimierung aus dem in Art. 4 Abs. 2 DSG normierten Grundsatz der Verhältnismässigkeit; siehe dazu Florent Thouvenin, Forschung im Spannungsfeld von Big Data und Datenschutzrecht: eine Problemskizze, in: Volker Boehme-Nessler/Manfred Rehbinder (Hrsg.), Big Data: Ende des Datenschutzes?, Bern 2017, 27–53, 34; Bruno Baeriswyl, in: Bruno Baeriswyl/Kurt Pärli (Hrsg.), Datenschutzgesetz (DSG), SHK Stämpflis Handkommentar, Bern 2015, DSG 4 N 23; Philippe Meier, Protection des données, Bern 2010, Rz. 633 & 673.

2.2. Schutz der Privatsphäre, Datenschutz und Überwachung

[16] Unlängst befanden die Gerichte von Wales in Grossbritannien über die Zulässigkeit des polizeilichen Einsatzes von Gesichtserkennungstechnologie. Die walisische Polizei hatte eine neue, mit Videoüberwachung gekoppelte Software eingesetzt, um die Sicherheit bei bestimmten Grossereignissen wie z. B. Fussballspielen zu garantieren. Aufgrund einer vorab von der Polizei erstellten Liste half die Software, potenziell kriminelle Personen im Publikum zu identifizieren. Da man sich im Publikum der Überwachung nicht entziehen konnte, wandten sich einzelne Personen an die Gerichte und machten geltend, die Technologie greife in rechtswidriger Weise in ihre Privatsphäre ein, insbesondere, weil sie selbst nicht auf der Liste der Polizei mit den zu überwachenden Personen aufgeführt waren.[28]

[17] Der in diesem Fall geltend gemachte Schutzbedarf hat verschiedene Dimensionen: Der Schutz der Privatsphäre soll Menschen ermöglichen, selbst darüber zu entscheiden, welche Informationen über sie Dritten zugänglich sind; damit soll auch die Möglichkeit geschaffen werden, in bestimmten Bereichen «in Ruhe gelassen» zu werden.[29] Der Datenschutz erlaubt den betroffenen Personen darüber hinaus, auch nach dem Zugänglichmachen der sie betreffenden Daten eine gewisse Kontrolle über deren Bearbeitung auszuüben, etwa mithilfe des Auskunftsrechts. Das Datenschutzrecht beruht dabei im Wesentlichen auf der Idee der informationellen Selbstbestimmung, die meist als Grundrecht verstanden wird.[30] Der zunächst überzeugend erscheinende Gedanke der Ausübung einer umfassenden Kontrolle über die eigenen Daten lässt sich in der

---

28 Für weitere Details siehe die Entscheidung R (Bridges) v CCSWP and SSHD, CO/4085/2018, [2019] EWHC 2341 (Admin), und im Berufungsverfahren R (Bridges) v CCSWP and SSHD, C1/2019/2670, [2020] EWCA Civ 1058. Weitere Entscheidungen zur Gesichtserkennung wurden gefällt in Frankreich (La Quadrature du Net, No. 1901249: Gesichtserkennungssoftware angewendet beim Zutritt zur Schule) und Schweden (Supervision pursuant to the General Data Protection Regulation (EU) 2016/679 – facial recognition used to monitor attendance of students, DI-2019-2221: Gesichtserkennung angewendet zwecks Präsenzkontrolle in der Schule). Eine öffentlich zugängliche Datenbank, die Zwischenfälle mit künstlicher Intelligenz aufführt, enthält über 600 Einträge, von denen eine Vielzahl auf Gesichtserkennung zurückgehen, siehe das sog. AIAAIC Repository, https://docs.google.com/spreadsheets/d/1Bn55B4xz21-_Rgdr8BBb2lt0n_4rzLGxFADMlVW0PYI/edit#gid=888071280 (20. Mai 2021).

29 Zu den verschiedenen Konzepten des Schutzes der Privatsphäre bzw. der (teilweise weiter gefassten) «Privacy» siehe etwa: Daniel J. Solove, Understanding Privacy, 2008, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1127888 (26. Mai 2021), 13–38; Helen Nissenbaum, Privacy in Context: Technology, Policy, and the Integrity of Social Life, Stanford 2009, 67 ff.; Alan F. Westin, Privacy and Freedom, New York 1967, 77; Charles Fried, Privacy, The Yale Law Journal 1968, 475–493; Ruth Gavinson, Privacy and the Limits of Law, The Yale Law Journal 1980, 421–471; Randall P. Bezanson, The Right to Privacy Revisited: Privacy, News and Social Change 1890–1990, California Law Review 1992, 1133–1175; Adam D. Moore, Defining Privacy, Journal of Social Philosophy 2008, 411–428; Bert-Jaap Koops et al., A Typology of Privacy, University of Pennsylvania Journal of International Law 2017, 483–575.

30 Wegweisend der Entscheid des deutschen Bundesverfassungsgerichts vom 15. Dezember 1983, Az. 1 BvR 209/83, 1 BvR 484/83, 1 BvR 420/83, 1 BvR 362/83, 1 BvR 269/83, 1 BvR 440/83 – Volkszählung. In Frankreich findet sich gar eine ausdrückliche Normierung in Art. 1 al. 2 de la loi n. 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés: «Les droits des personnes de décider et de contrôler les usages qui sont faits des données à caractère personnel les concernant et les obligations incombant aux personnes qui traitent ces données s'exercent dans le cadre du règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016, de la directive (UE) 2016/680 du Parlement européen et du Conseil du 27 avril 2016 et de la présente loi». In der Schweiz wird das Grundrecht auf informationelle Selbstbestimmung aus Art. 13 Abs. 2 BV abgeleitet, wonach jede Person Anspruch auf Schutz vor Missbrauch ihrer persönlichen Daten hat; siehe dazu: BGE 146 I 11, E. 3; BGE 145 IV 42, E. 4.2; BGE 143 I 253, E. 4.8; BGE 142 II 340, E. 4.2; BGE 140 I 2, E. 9, alle m. w. H.; statt vieler: Giovanni Biaggini, BV Kommentar, Bundesverfassung der Schweizerischen Eidgenossenschaft, 2. Aufl., Zürich 2017, BV 13 N 11; Stephan Breitenmoser/Rainer J. Schweizer, in: Bernhard Ehrenzeller/Benjamin Schindler/Rainer J. Schweizer/Klaus A. Vallender (Hrsg.), Die schweizerische Bundesverfassung, St.Galler Kommentar, 3. Aufl., Zürich/St.Gallen 2014 (zit. Autor, St. Galler Kommentar), BV 13 N 64.

Realität allerdings kaum verwirklichen und wird deshalb zunehmend infrage gestellt.[31] Beide Dimensionen – der Schutz der Privatsphäre und der Datenschutz – werden im oben beschriebenen Fall berührt. Neben der individuellen Ebene wirken diese beiden Dimensionen auch auf einer gesellschaftlichen Ebene, indem sie das Sammeln und Bearbeiten von Daten durch Staat und Unternehmen Regeln unterwerfen und im Ergebnis auch beschränken.

[18] Das Beispiel der Gesichtserkennung macht deutlich, dass der Einsatz von KI die mit dem Sammeln und Bearbeiten von Personendaten ohnehin verbundenen Gefahren noch akzentuieren kann. Während die umfassende Videoüberwachung eines Fussballstadions mit Blick auf den Schutz der Privatsphäre und den Datenschutz für sich allein schon heikel erscheint, wird sie dies erst recht, wenn die Bilddaten mit weiteren Daten zusammengeführt werden, die bspw. von Videokameras in öffentlichen Verkehrsmitteln stammen, weil die Verbindung der Bilddaten ein Verfolgen der Bewegungen und Handlungen einzelner Personen oder ganzer Personengruppen im öffentlichen Raum erlaubt. Hinzu kommt, dass die automatisierte Gesichtserkennung ermöglicht, eine Vielzahl von Personen zu identifizieren, während sich eine menschliche Analyse von Bilddaten auf einige wenige polizeilich bekannte Personen beschränken müsste.

[19] Der Komplexität der Sachlage entsprechend würden sich weitere Fragen stellen, wenn in der Schweiz eine Kantonspolizei Gesichtserkennungstechnologie in ähnlicher Weise wie die walisische Polizei einsetzen würde.[32] Aus dem Blickwinkel des Polizei- und Strafprozessrechts müsste man insbesondere fragen, welche gesetzliche Grundlage für den Einsatz herangezogen werden könnte oder ob die polizeiliche Generalklausel genügen würde. Hätte die Polizei zudem eine hinreichende Kontrolle über die Funktion der Software? Wie zuverlässig und damit mit welcher Sensibilität und Spezifität müsste die Software arbeiten und würde ihr Output im Zweifel als Beweismittel ausreichen? Die Schweizer Rechtsordnung bleibt Antworten auf diese Fragen zurzeit schuldig.[33]

---

31 Möglicherweise auch hier wegweisend der Entscheid des deutschen Bundesverfassungsgerichts vom 6. November 2019, Az. 1 BvR 16/13 – Recht auf Vergessen I. Das Gericht hat festgehalten, dass das Recht auf informationelle Selbstbestimmung kein «allgemeines oder gar umfassendes Selbstbestimmungsrecht über die Nutzung der eigenen Daten» enthält. Es gewährleistet «den Einzelnen aber die Möglichkeit, in differenzierter Weise darauf Einfluss zu nehmen, in welchem Kontext und auf welche Weise die eigenen Daten anderen zugänglich und von ihnen genutzt werden. Es enthält damit die Gewährleistung, über der eigenen Person geltende Zuschreibungen selbst substantiell mitzuentscheiden». Für eine kritische Auseinandersetzung mit dem Grundrecht auf informationelle Selbstbestimmung siehe auch: Karl-Heinz Ladeur, Das Recht auf informationelle Selbstbestimmung: Eine juristische Fehlkonstruktion?, Die Öffentliche Verwaltung DÖV 2009, 45–55, 45; Marion Albers, Umgang mit personenbezogenen Informationen und Daten, in: Wolfgang Hoffmann-Riem et al. (Hrsg.), Grundlagen des Verwaltungsrechts, Band II, 2. Aufl., München 2012, 107–234; Gabriel Britz, Informationelle Selbstbestimmung zwischen rechtswissenschaftlicher Grundsatzkritik und Beharren des Bundesverfassungsgerichts, in: Wolfgang Hoffmann-Riem (Hrsg.), Offene Rechtswissenschaft, Tübingen 2010, 561–596; Paul De Hert/Serge Gutwirth, Privacy, Data Protection and Law Enforcement, Opacity of the Individual and Transparency of the Power, in: Erik Claes et al. (Hrsg.), Privacy and the Criminal Law, Antwerp/Oxford 2006, 61–104; Gloria Gonzáles Fuster, The Emergence of Personal Data Protection as a Fundamental Right of the EU, Cham 2014; Nikolaus Marsch, Das europäische Datenschutzgrundrecht, Tübingen 2018, 98 ff.; Ders., Artificial Intelligence and the Fundamental Right to Data Protection, in: Thomas Wischmeyer/Timo Rademacher (Hrsg.), Regulating Artificial Intelligence, Cham 2020, 33–52. Für das Verhältnis unter Privaten siehe Florent Thouvenin, Informational Self-Determination: A Convincing Rationale for Data Protection Law?, JIPITEC 2021 (erscheint demnächst).

32 Für das britische Recht siehe Biometrics and Forensics Ethics Group, Interim Report of BFEG Facial Recognition Working Group, Februar 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/781745/Facial_Recognition_Briefing_BFEG_February_2019.pdf (14. Juni 2021). Im Anhang dieses Berichts sind neun Grundprinzipien festgehalten, an denen sich der Gebrauch von Gesichtserkennung in der Strafverfolgung zu orientieren hat. Für den gegenwärtigen Gebrauch von Gesichtserkennungssoftware in den USA siehe z. B. Kashmir Hill, The facial-recognition app Clearview sees as spike in use after Capitol attack, New York Times, 9. Januar 2021.

33 Für den Fall der automatischen Fahrzeugschilderkennung siehe Elias Krummenacher/Nicole Ebneter, Bundesgericht, Urteil 6B_908/2018 vom 7. Oktober 2019 (zur Publikation vorgesehen), A. gegen Generalstaatsanwaltschaft

[20] Das Beispiel der Gesichtserkennung zeigt, dass das Schweizer Recht auf viele Fragen zum Einsatz von KI zu Zwecken der Überwachung bislang keine Antworten geben kann. Der Handlungsbedarf ist umso deutlicher, als im EU-KI-Verordnungsvorschlag vorgesehen ist, die laufende («real time») biometrische Erkennung von Personen in der Öffentlichkeit zum Zweck der Strafverfolgung zu untersagen. Allerdings ist eine potenziell weitreichende Ausnahme vorgesehen, wenn die biometrische Erkennung (welche die Gesichtserkennung umfasst) der öffentlichen Sicherheit dient.[34] Zudem ist davon auszugehen, dass die Erkennung biometrischer Merkmale durch Privatpersonen und zu anderen Zwecken als der Strafverfolgung erlaubt bleibt; immerhin wird sie als hoch riskant einzustufen sein und damit die entsprechenden Erfordernisse zu erfüllen haben.[35] Der Verordnungsvorschlag der EU macht jedenfalls deutlich, dass auch die Schweiz sich der Frage stellen muss, ob und unter welchen Voraussetzungen Behörden und Private Gesichtserkennungstechnologien einsetzen dürfen. Angesichts der grossen Risiken für den Schutz der Privatsphäre dürfte es angezeigt sein, dies spezifisch zu regeln. Denkbar wäre eine Berücksichtigung in den datenschutzrechtlichen Bestimmungen des Strafprozessrechts und der kantonalen Polizeigesetze; alternativ könnte auch ein Ansatz gewählt werden, wie er heute für die Überwachung des Fernmeldeverkehrs im Bundesgesetz in Bezug auf die Überwachung des Post- und Fernmeldeverkehrs (BÜPF)[36] besteht.

2.3. Diskriminierung, Fairness und Bias

[21] Die zentrale Aufgabe der meisten Anwendungen von KI besteht im weitesten Sinne darin, Entscheidungen zu treffen bzw. vorzubereiten, etwa indem eine E-Mail als Spam qualifiziert wird. Im Kern werden dabei gewisse Konstellationen nach bestimmten, allerdings nicht immer vollständig nachvollziehbaren Kriterien von anderen unterschieden. Zuweilen werden aufgrund der Unterscheidung auch unmittelbar Handlungen vorgenommen, so wird etwa eine E-Mail in einen Spam-Ordner statt in den Posteingang bewegt. Aufgrund der zentralen Funktion von KI, als verschieden qualifizierte Konstellationen unterschiedlich zu behandeln, gibt es ebenso zahlreiche wie unterschiedliche Beispiele für Diskriminierungen, die auf den Einsatz von KI zurückzuführen sind.

[22] Illustrativ ist bspw., dass vor einigen Jahren bei einer Google-Bilder-Suche nach dem Begriff «three black teenagers» überwiegend Polizeifotos von dunkelhäutigen Jugendlichen, beim Begriff «three white teenagers» hingegen Fotos von lachenden weissen Jugendlichen angezeigt wurden. Google reagierte umgehend auf die negativen Schlagzeilen und stellte sicher, dass nunmehr bei beiden Formulierungen Jugendliche in ähnlichen Lebenssituationen angezeigt werden.[37] Ein wei-

---

des Kantons Thurgau, Verwertbarkeit von Beweisen (mehrfaches Fahren ohne Berechtigung), Urteilsbesprechung, AJP 2020, 221–226, insb. 225 zur Übertragbarkeit auf die Gesichtserkennung. Das Bundesgericht befand in dem Entscheid den Eingriff in die informationelle Selbstbestimmung durch das im Kanton Thurgau angewandte System als schwer und verlangte eine formelle gesetzliche Grundlage.

34 Für weitere Voraussetzungen siehe Art. 5 Abs. 1 lit. d EU-KI-Verordnungsvorschlag. Vergleiche auch Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, Convention 108, Guidelines on Facial Recognition, TPD(2020)03rev4, 28. Januar 2021, https://rm.coe.int/guidelines-on-facial-recognition/1680a134f3 (26. Mai 2021).

35 Siehe Anhang 3 Ziff. 1 EU-KI-Verordnungsvorschlag.

36 Bundesgesetz betreffend die Überwachung des Post- und Fernmeldeverkehrs vom 18. März 2016 (BÜPF; SR 780.1).

37 Elle Hunt, «Three Black Teenagers»: anger as Google image search shows police mugshots, The Guardian 9. Juni 2016, https://www.theguardian.com/technology/2016/jun/09/three-black-teenagers-anger-as-google-image-

teres Beispiel betrifft eine US-amerikanische Sexarbeiterin, die für eine private Reise mit ihrem Partner über Airbnb eine Unterkunft in New York gebucht hatte. Wenig später wurde sie von der Nutzung der Plattform mit der Begründung ausgeschlossen, ihr Konto entspreche nicht den Nutzungsbedingungen und «community standards». Wie sich später herausstellte, verwendete Airbnb eine KI, die mithilfe von «öffentlichen digitalen Fussabdrücken», eine «Hintergrundprüfung» durchführte, um die Vertrauenswürdigkeit der Nutzerinnen und Nutzer zu bewerten. Menschen, bei denen die KI einen Zusammenhang mit unerwünschten Tätigkeiten oder Aussagen erkannte, wurden von der Nutzung von Airbnb auch dann automatisch ausgeschlossen, wenn keine Bezüge zur Nutzung der Plattform bestanden.[38]

[23] KI wird nicht nur von privaten Unternehmen, sondern auch von der öffentlichen Verwaltung verwendet. Für Aufmerksamkeit sorgte unlängst das Beispiel des österreichischen Arbeitsmarkt-Chancen-Assistenzsystems (AMAS). Dieses sollte die Arbeit des Arbeitsmarktservices (AMS) verbessern, indem auf Basis einer statistischen Analyse historischer Daten die zukünftigen Chancen von Arbeitssuchenden am Arbeitsmarkt berechnet wurden. In der Anwendung zeigte sich, dass eines der Regressionsmodelle insbesondere Frauen und Menschen mit Behinderungen schlechtere Chancen auf dem Arbeitsmarkt einräumte.[39] Eine Studie[40] aus dem Jahr 2020 kam zum Ergebnis, dass die vorliegende KI-Anwendung keine Mechanismen enthielt, um einem «Bias» vorzubeugen. Im August 2020 untersagte sodann die österreichische Datenschutzbehörde den Einsatz von AMAS aus datenschutzrechtlichen Gründen;[41] im Dezember 2020 gab das österreichische Bundesverwaltungsgericht aber einer Beschwerde des AMS statt und hob den Entscheid der Datenschutzbehörde auf.[42] Zwar ging es in der Rechtsstreitigkeit hauptsächlich um die Frage der ausreichenden gesetzlichen Grundlage für die Datenverarbeitung; das Bundesverwaltungsgericht sah es allerdings im Rahmen der Abgrenzung zwischen voll- und teilautomatisierten Verfahren auch als nicht erwiesen an, dass die Sachbearbeitenden die Prognose des Algorithmus übernommen hätten, ohne diese zu hinterfragen.

[24] Die Beispiele der Google-Bildersuche, des Ausschlusses von der Nutzung von Airbnb und des österreichischen Arbeitsmarktservices illustrieren, dass der Einsatz von KI zu problematischen Diskriminierungen führen kann. Wie der Einsatz von KI in Spam-Filtern zeigt, ist aber zu differenzieren: In der Informatik und in bestimmten Geistes- bzw. Sozialwissenschaften (etwa den Wirtschaftswissenschaften) bezeichnet der Begriff der «Diskriminierung» schlicht die unterschiedliche Behandlung verschiedener Konstellationen, etwa die Zuordnung von E-Mails in

search-shows-police-mugshots (14. Juni 2021); Ben Guarino, Google faulted for racial bias in image search results for black teenagers, The Washington Post, 10. Juni 2016, https://www.washingtonpost.com/news/morning-mix/wp/2016/06/10/google-faulted-for-racial-bias-in-image-search-results-for-black-teenagers/ (14. Juni 2021).

38 EJ Dickson, Who's allowed to use Airbnb, Rolling Stone, 8. Januar 2020, https://www.rollingstone.com/culture/culture-news/airbnb-sex-worker-discrimination-935048/ (26. Mai 2021).

39 Nicolas Kayser-Bril, Austria's employment agency rolls out discriminatory algorithm, sees no problem, 6. Oktober 2019, https://algorithmwatch.org/en/story/austrias-employment-agency-ams-rolls-out-discriminatory-algorithm/ (26. Mai 2021).

40 Doris Allhutter/Astrid Mager/Florian Cech/Fabian Fischer/Gabriel Grill, Der AMS-Algorithmus, eine Soziotechnische Analyse des Arbeitsmarktchancen-Assistenz-Systems (AMAS), Institut für Technikfolgen-Abschätzung der Österreichischen Akademie der Wissenschaften, Wien 2020, http://epub.oeaw.ac.at/0xc1aa5576_0x003bfdf3.pdf (26. Mai 2021).

41 Andreas Zavadil, Datenschutzrechtliche Zulässigkeit des «AMS-Algorithmus», Datenschutzbehörde Republik Österreich Newsletter 4/2020, 3–4.

42 Bundesverwaltungsgericht der Republik Österreich, Entscheidung vom 18. Dezember 2020, Az. W256 2235360-1/5E, https://www.ris.bka.gv.at/Dokumente/Bvwg/BVWGT_20201218_W256_2235360_1_00/BVWGT_20201218_W256_2235360_1_00.html (26. Mai 2021).

den Posteingang oder den Spam-Ordner. Von diesem wertneutralen Verständnis ist die qualifizierende Verwendung des Begriffs «Diskriminierung» im rechtlichen Sinn zu unterscheiden. Eine rechtlich relevante und unter Umständen verpönte Diskriminierung ist eine ungleiche Behandlung von Menschen, die aufgrund bestimmter besonders geschützter Merkmale – insbesondere aufgrund von Herkunft, Rasse, Geschlecht, Alter, Sprache, sozialer Stellung, Lebensform, religiöser, weltanschaulicher oder politischer Überzeugung sowie körperlicher, geistiger oder psychischer Behinderung – erfolgt und sachlich nicht gerechtfertigt werden kann (siehe Art. 8 Abs. 2 BV). Das Diskriminierungsverbot der Bundesverfassung ist dabei die grundrechtliche Antwort auf historische Erfahrungen der Ausgrenzung, Herabwürdigung und Stigmatisierung von Menschen aufgrund eines sensiblen und deshalb schützenswerten Merkmals der Persönlichkeit.[43] An den verfassungsrechtlich geprägten Begriff der Diskriminierung knüpfen verschiedene Rechtsnormen an, auf Bundesebene insbesondere das Bundesgesetz über die Gleichstellung von Frau und Mann (Gleichstellungsgesetz),[44] das Bundesgesetz über die Beseitigung von Benachteiligungen von Menschen mit Behinderungen (Behindertengleichstellungsgesetz)[45] und die Strafnorm zu Diskriminierung und Aufruf zu Hass (Art. 261bis StGB).[46] Zum Diskriminierungsverbot der Bundesverfassung gibt es zudem eine umfangreiche Rechtsprechung.[47]

[25] Im Gegensatz zur Diskriminierung sind die Begriffe «Fairness» und «Bias», denen im Zusammenhang mit KI-Anwendungen eine zentrale Bedeutung zukommt, rechtlich nicht klar umrissen. Zwar sind Überschneidungen mit dem Gebot der rechtsgleichen Behandlung (Art. 8 Abs. 1 BV), wonach «Gleiches nach Massgabe seiner Gleichheit gleich und Ungleiches nach Massgabe seiner Ungleichheit ungleich»[48] zu behandeln ist, auszumachen. Die Begriffe «Fairness» und «Bias» werden aber in verschiedenen Disziplinen weiter gefasst. Gerade der Begriff der Fairness ist ohnehin nur bedingt juristisch, dafür umso stärker moralphilosophisch besetzt.[49] In den Computerwissenschaften hat Fairness wiederum eine eigene Dimension; er wird dort meist mit «Bias» (bzw. Abwesenheit von «Bias») gleichgesetzt.[50]

---

43 Statt vieler Jörg Paul Müller/Markus Schefer, Grundrechte in der Schweiz, Im Rahmen der Bundesverfassung, der EMRK und der UNO-Pakte, 4. Aufl., Bern 2008, 651 & 684 ff.; René Rhinow/Markus Schefer/Peter Uebersax, Schweizerisches Verfassungsrecht, 3. Aufl., Basel 2016, Rn. 1889; Bernhard Waldmann, St.Galler Kommentar (Fn. 30), BV 8 N 45 & 65.

44 Bundesgesetz über die Gleichstellung von Frau und Mann (Gleichstellungsgesetz, GlG) vom 24. März 1994, SR 151.1.

45 Bundesgesetz über die Beseitigung von Benachteiligungen von Menschen mit Behinderungen (Behindertengleichstellungsgesetz, BehiG) vom 13. Dezember 2002, SR 151.3.

46 Schweizerisches Strafgesetzbuch (StGB) vom 21. Dezember 1937, SR 311.0.

47 In den letzten Jahren u. a. BGE 145 I 86, E. 5.1; BGE 145 I 146, E. 5.2; BGE 145 II 161, E. 4.3.6, 4.5.1; BGE 143 I 133, E. 2.3.1; BGE 143 I 368, E. 5.1; BGE 143 V 122, E. 5.3.2.1; BGE 142 V 323, E. 6.1.1, 6.1.2; BGE 141 I 12, E. 3.1; BGE 141 I 250, E. 4.3.2; BGE 141 I 251, E. 4.3.2. Siehe auch die Literatur zum verfassungsrechtlichen Diskriminierungsverbot in Art. 8 Abs. 2 BV, etwa Samantha Besson, L'égalité horizontale, l'égalité de traitement entre particuliers, des fondements théoriques au droit privé suisse, Fribourg 1999; Bernhard Pulver, L'interdiction de la discrimination, Basel 2003; Bernhard Waldmann, Das Diskriminierungsverbot von Art. 8 Abs. 2 BV als besonderer Gleichheitssatz, Bern 2003.

48 Siehe etwa BGE 145 I 259, E. 6.1, sowie Rainer Schweizer, St.Galler Kommentar (Fn. 30), BV 8 N 18 ff.

49 Für die moralphilosophische Dimension der Fairness siehe z. B. John Rawls, Justice as Fairness, Harvard 2001.

50 Vergleiche z. B. IBM, Everyday Ethics for Artificial Intelligence, 2018, https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf (11. Juni 2021), 32–39, wo unter dem Titel von «Fairness» Lösungen für das Problem des Bias bei künstlicher Intelligenz vorgeschlagen werden.

[26] Wenn somit von KI-Anwendungen erwartet wird, dass sie «fair» sind, keinen «Bias» haben und nicht diskriminieren,[51] ist einzig das Kriterium der Nicht-Diskriminierung rechtlich relevant und operabel. Dieses vermag allerdings viele Konstellationen zu erfassen, namentlich auch jene, in denen fehlerhafte oder unvollständige Daten für das Training einer KI verwendet wurden oder Probleme in der Architektur der Modelle bestehen. Voraussetzung für den Zugriff des Rechts ist zumindest in diesen Konstellationen regelmässig, dass der Output diskriminierende Wirkung entfaltet.[52]

[27] Mit dem Einsatz von KI hat sich die Gefahr einer indirekten bzw. verdeckten und unter Umständen einzig in der Auswirkung erkennbaren Diskriminierung erhöht. Von einer solchen indirekten Diskriminierung geht man aus, wenn die verwendeten Kriterien – bei KI insbesondere die Daten – neutral erscheinen, im Ergebnis aber Personen, die geschützte Merkmale aufweisen, benachteiligt werden.[53] Diskriminierungsverbote vermögen zwar indirekte Diskriminierung regelmässig zu erfassen; diese Form ist aber seit jeher mit besonderen Herausforderungen tatsächlicher Art verbunden (bspw. bei der Beweisführung), die bei KI-Anwendungen noch verstärkt werden. Ferner besteht bei KI eine erhöhte Gefahr der Diskriminierung aufgrund von Assoziierung. In diesem Fall erfolgt eine in den Auswirkungen unterschiedliche Behandlung allein aufgrund der Beziehung einer Person zu einer geschützte Merkmale aufweisenden Gruppe.[54]

[28] In öffentlich-rechtlichen Verhältnissen greift im Allgemeinen der Schutz der Bundesverfassung vor Diskriminierung. Setzt die öffentliche Verwaltung KI ein, sind daher in erster Linie die verfassungsrechtlichen Vorgaben umzusetzen. Die Behörden haben insbesondere die notwendigen Vorkehrungen zu treffen, damit potenziell diskriminierende Entscheidungsempfehlungen erkannt und von Sachbearbeitenden nicht übernommen werden. Die Verwaltung muss – u. a. aufgrund des Datenschutzrechts und des Untersuchungsgrundsatzes, wonach die Behörden von Amtes wegen für die Feststellung der rechtserheblichen Tatsachen zuständig sind – sicherstel-

---

51 Gesellschaft für Informatik, Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren, Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen, Berlin 2018, 12; Jon Kleinberg/Jens Ludwig/Sendhil Mullainathan/Ashesh Rambachan, Algorithmic Fairness, AEA Papers and Proceedings 2018, 22–27, https://doi.org/10.1257/pandp.20181018 (26. Mai 2021), 22; siehe dazu auch Anja Bechmann/Geoffrey C. Bowker, Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media, Big Data & Society 2019, 1–11, https://doi.org/10.1177/2053951718819569 (14. Juni 2021); Sam Corbett-Davies/Sharad Goel, The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning, 2018, https://arxiv.org/abs/1808.00023 (26. Mai 2021), 2; Danielle Ensign/Sorelle A. Friedler/Scott Neville/Carlos Scheidegger/Suresh Venkatasubramanian, Runaway Feedback Loops in Predictive Policing, Proceedings of Machine Learning Research 2018, 1–12.

52 Carsten Orwat, Diskriminierungsrisiken durch Verwendung von Algorithmen, Berlin 2019, https://www.antidiskriminierungsstelle.de/SharedDocs/Downloads/DE/publikationen/Expertisen/studie_diskriminierungsrisiken_durch_verwendung_von_algorithmen.pdf (8. Juni 2021), 104–106; Martini, (Fn. 17), 13; Wiebke Fröhlich/Indra Spiecker gen. Döhmann, Können Algorithmen diskriminieren?, Verfassungsblog, 26. Dezember 2018, www.verfassungsblog.de/koennen-algorithmen-diskriminieren (26. Mai 2021); siehe auch Le Chen/Ruijun Ma/Anikó Hannák/Christo Wilson, Investigating the Impact of Gender on Rank in Resume Search Engines, Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, 1–14.

53 Solon Barocas/Andrew D. Selbst, Big Data's Disparate Impact, California Law Review 2016, 671–732, 680–687; Frederik Zuiderveen Borgesius, Discrimination, artificial intelligence, and algorithmic decision-making, Strasbourg 2018, https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73 (26. Mai 2021), 10–13.

54 Sandra Wachter, Affinity Profiling and Discrimination by Association in Online Behavioural Advertising, Berkeley Technology Law Journal 2020, 367–430, 394–412.

len, dass die verwendeten Trainings- und Sachverhaltsdaten richtig sind und nur Daten genutzt werden, die für das entsprechende Verfahren geeignet sind.[55]

[29] Ein mit dem allgemeinen Diskriminierungsverbot im öffentlichen Recht vergleichbares Instrument fehlt im Verhältnis unter Privaten.[56] Faktische Unterscheidungen anhand von geschützten Merkmalen (Diskriminierungen), die Google, Airbnb und andere Dienste unter Einsatz von KI vornehmen, werden in der Schweiz deshalb rechtlich nur punktuell erfasst, namentlich im Anwendungsbereich des Gleichstellungs- und des Behindertengleichstellungsgesetzes, in der Strafnorm über Diskriminierung und Aufruf zu Hass (Art. 261bis Abs. 5 StGB) und in bestimmten Konstellationen der bilateralen Verträge mit der Europäischen Union.[57] In Bezug auf das Verhältnis zwischen Privaten könnte folglich Handlungsbedarf bestehen. Eine naheliegende und in der Literatur bereits vertretene Möglichkeit besteht darin, eine Diskriminierung von natürlichen Personen aufgrund bestimmter geschützter Merkmale als Persönlichkeitsverletzung im Sinn von Art. 28 ZGB zu verstehen.[58] Dadurch würde ein breiter Zugriff auf diskriminierendes Verhalten von Individuen und Unternehmen ermöglicht. Dieser Ansatz liesse sich auch ohne Weiteres im Rahmen des geltenden Rechts und insbesondere durch eine Weiterentwicklung der Rechtsprechung umsetzen. Die nach dem allgemeinen Persönlichkeitsrecht für eine rechtswidrige Verletzung erforderliche Intensität des Eingriffs[59] würde zudem bewirken, dass Bagatellfälle vom Zugriff des Rechts ausgenommen blieben. Durch die zur Verfügung stehenden Rechtfertigungsmöglichkeiten bei Persönlichkeitsverletzungen (Art. 28 Abs. 2 ZGB) wären solche Diskriminierungen ferner – ähnlich wie im öffentlichen Recht – einer Rechtfertigung wegen Vorliegens eines sachlichen Grundes für die Ungleichbehandlung zugänglich.

[30] Somit bestehen insgesamt durchaus Möglichkeiten, das bestehende Schweizer Recht organisch weiterzuentwickeln, um durch KI neu auftretende Diskriminierungen aller Art zu erfassen, ohne den Gesetzgeber notwendigerweise damit befassen zu müssen. Der Verordnungsvorschlag der EU-Kommission verlässt sich bezüglich möglicher Diskriminierungen durch KI primär auf die Bestimmung zu den Transparenzanforderungen (Art. 13 EU-KI-Verordnungsvorschlag) und

---

55 Zu diesem Themenkreis siehe Braun Binder (Fn. 19), 474 f.; Dies., Als Verfügungen gelten Anordnungen der Maschinen im Einzelfall: Dystopie oder künftiger Verwaltungalltag?, ZSR I 2020, 253–278, 275 f.; Braun Binder et al. (Fn. 4), 38 f. & 44–46.

56 Zur Diskriminierung im privatrechtlichen Kontext siehe Tarkan Göksu, Rassendiskriminierung beim Vertragsabschluss als Persönlichkeitsverletzung, Fribourg 2003; Ruth Arnet, Freiheit und Zwang beim Vertragsschluss, Bern 2007; Florent Thouvenin, Privatversicherungen: Datenschutzrecht als Grenze der Individualisierung?, in: Astrid Epiney/Déborah Sangsue (Hrsg.), Datenschutz und Gesundheitsrecht, Zürich 2019, 15–42. Zur persönlichen Integrität im Beschäftigungskontext siehe Ulrich Leicht-Deobald/Thorsten Busch/Christoph Schank/Antoinette Weibel/Simon Schafheitle/Isabelle Wildhaber/Gabriel Kasper, The Challenges of Algorithm-Based HR Decision-Making for Personal Integrity, Journal of Business Ethics 2019, 377–392, https://doi.org/10.1007/s10551-019-04204-w (22. Juni 2021); Gabriel Kasper/Isabelle Wildhaber, Big Data am Arbeitsplatz, Datenschutz- und arbeitsrechtliche Herausforderungen von People Analytics in Schweizer Unternehmen, in: Ueli Kieser/Kurt Pärli/Ursula Uttinger (Hrsg.), Datenschutztagung 2018 – Ein Blick auf aktuelle Rechtsentwicklungen, Zürich 2019, 189–232; Isabelle Wildhaber/Melinda F. Lohmann/Gabriel Kasper, Diskriminierung durch Algorithmen – Überlegungen zum schweizerischen Recht am Beispiel prädiktiver Analytik am Arbeitsplatz, ZSR 2019, 459–489.

57 Siehe dazu beispielhaft Art. 2 des Abkommens zwischen der Schweizerischen Eidgenossenschaft einerseits und der Europäischen Gemeinschaft und ihren Mitgliedstaaten andererseits über die Freizügigkeit, SR 0.142.112.681.

58 Arnet (Fn. 56), Rn. 356; Peter Gauch/Walter Schluep/Jörg Schmid, OR AT: Band 1, 11. Aufl., Zürich/Basel/Genf 2020, Rn. 1111; Thouvenin (Fn. 56), 26 m. w. H.; Göksu (Fn. 56), Rz. 214–267; Tarek Naguib, Diskriminierende Verweigerung des Vertragsabschlusses über Dienstleistungen Privater: Diskriminierungsschutz zwischen Normativität, Relativität und Idealität, AJP 2009, 993–1017, 1006 ff.

59 Andreas Meili, in: Thomas Geiser/Christiana Fountoulakis (Hrsg.), Zivilgesetzbuch I, Basler Kommentar, 6. Aufl., Basel 2018, ZGB 28 N 38; Bianka S. Dörr, in: Andrea Büchler/Dominique Jakob (Hrsg.), Schweizerisches Zivilgesetzbuch, Kurzkommentar, 2. Aufl., Basel 2018, ZGB 28 N 2; Heinz Hausheer/Regina E. Aebi-Müller, Das Personenrecht des Schweizerischen Zivilgesetzbuches, 5. Aufl., Bern 2020, Rz. 547.

statuiert keine separate Bestimmung, um Diskriminierungsschutz sicherzustellen. Allerdings ist diese Zurückhaltung im Lichte des im Vergleich zur Schweiz stärker ausgebauten Diskriminierungsrechts der EU zu sehen.[60]

2.4. Manipulation

[31] KI wird nicht selten eingesetzt, um menschliches Verhalten zu beeinflussen. Die Konstellationen sind ebenso zahlreich wie unterschiedlich. Dabei kann es sich um die verdeckte Beeinflussung eines einzelnen Individuums oder einer Gruppe von Individuen handeln, um deren Selbstkontrolle und Entscheidungskraft zu unterlaufen.[61] Entscheidungssituationen werden zu diesem Zweck so modelliert, dass sich eine relevante Anzahl von Nutzerinnen und Nutzern mit einer hinreichenden Wahrscheinlichkeit in der gewünschten Art und Weise verhält. Im Vordergrund stehen die Verwendung von sog. Empfehlungsalgorithmen und die Verbreitung von falschen oder unvollständigen Informationen sowie das Unterdrücken von zutreffenden Informationen (Desinformation).

[32] Bestens bekannt ist die Verwendung von Empfehlungsalgorithmen auf Plattformen der sozialen Medien wie Facebook, Twitter, LinkedIn oder YouTube. Solche Algorithmen werden in den sozialen Medien vor allem eingesetzt, um Nutzerinnen und Nutzern aufgrund ihres bisherigen Verhaltens auf der Plattform und durch Analyse anderer Daten laufend weitere Inhalte vorzuschlagen, die sie ebenfalls interessieren könnten, um sie möglichst lange auf der Plattform zu halten und weitere Werbung anzuzeigen und Daten zu gewinnen. Verbreitet ist auch die Anwendung von Empfehlungsalgorithmen im E-Commerce. Das prominenteste Beispiel sind die auf maschinellem Lernen beruhenden Empfehlungen von Produkten in Echtzeit auf der Plattform von Amazon. Amazon verwendet die entsprechenden Algorithmen nicht nur selbst, sondern bietet sie auch Dritten zur Nutzung an.[62] Ein weiteres Ziel der Verwendung von KI besteht folglich darin, Konsumentinnen und Konsumenten bei der Suche nach einem Produkt oder bei dessen Kauf weitere Produkte vorzuschlagen, die sie ebenfalls interessieren könnten.

[33] Das Denken und Handeln von Menschen kann nicht nur durch massgeschneiderte Empfehlungen, sondern auch durch Desinformation manipuliert werden. KI kann dazu beitragen, die Verbreitung von Desinformationen zu beschleunigen, aber auch bei deren Bekämpfung helfen.[63] Auch hier spielen die sozialen Medien insofern eine zentrale Rolle, als sie von Dritten zur Verbreitung von Desinformation genutzt werden können.

---

60 Siehe z. B. für das Verbot der Diskriminierung aufgrund des Alters, EuGH vom 22. November 2005, C144/04 – Mangold.

61 Siehe zur Manipulation Daniel Susser/Beate Roessler/Helen Nissenbaum, Technology, autonomy, and manipulation, Internet Policy Review 8/2 2019, 1–22; Tal Z. Zarsky, Privacy and Manipulation in the Digital Age, Theoretical Inquiries in Law 2019, 157–189; Frank Pasquale, The Black Box Society, The Secret Algorithms That Control Money and Information, Cambridge, MA/London 2015.

62 https://aws.amazon.com/personalize (14. Juni 2021).

63 Samuel Woolley, We're fighting fake news AI bots by using more AI, That's a mistake, MIT Technology Review, 8. Januar 2020, https://www.technologyreview.com/2020/01/08/130983/were-fighting-fake-news-ai-bots-by-using-more-ai-thats-a-mistake/ (26. Mai 2021), der allerdings darauf hinweist, dass künstliche Intelligenz allein Desinformation nicht zu verhindern vermag; John Villasenor, How to deal with AI-enabled disinformation, 23. November 2020, https://www.brookings.edu/research/how-to-deal-with-ai-enabled-disinformation/ (26. Mai 2021); Brian T. Horowitz, Can AI Stop People from Believing Fake News?, 15. März 2021, https://spectrum.ieee.org/tech-talk/robotics/artificial-intelligence/ai-misinformation-fake-news (26. Mai 2021).

[34] In welchen Konstellationen die Beeinflussung von Menschen als problematisch zu qualifizieren ist, hängt nicht in erster Linie vom Wahrheitsgehalt der Information oder vom verwendeten Empfehlungsalgorithmus bzw. von den Daten ab, mit denen eine KI trainiert wurde oder auf die sie zugreift. Entscheidend sind vielmehr der Kontext der Verwendung sowie das Ziel und das Mass der Beeinflussung. In vielen Fällen kann eine Beeinflussung wohl als unproblematisch hingenommen werden, etwa beim Anzeigen weiterer Inhalte auf sozialen Medien, um die Nutzungsdauer zu erhöhen. Problematisch wäre dann aber jedenfalls, wenn die Verwendung von KI zu einem suchtähnlichen Medienkonsum führen würde. In wiederum anderen Fällen ist eine Beeinflussung allerdings höchst problematisch, so z. B. bei der Manipulation von politisch relevanten Inhalten unmittelbar vor Wahlen oder Abstimmungen.

[35] Bei der rechtlichen Erfassung der aufgezeigten Konstellationen ist zu differenzieren. So stehen gegen das Verbreiten von Desinformation der strafrechtliche Ehrverletzungsschutz (Art. 173 ff. StGB) sowie das allgemeine Persönlichkeitsrecht (Art. 28 ZGB) zur Verfügung, wenn sich die Äusserungen auf eine bestimmte Person beziehen. Fehlt ein Personenbezug, sind dagegen kaum rechtliche Mittel vorhanden. Dieser offene Raum lässt sich im Wesentlichen darauf zurückführen, dass jede Regulierung in diesem Bereich eine Einschränkung der Meinungsfreiheit nach sich ziehen würde.[64] Wirken sich falsche oder irreführende Informationen auf den Wettbewerb aus, greifen die Bestimmungen des Bundesgesetzes gegen unlauteren Wettbewerb (UWG).[65] Diese stellen sicher, dass Anbieterinnen und Anbieter, Kundinnen und Kunden, Wirtschaftsverbände und Konsumentenschutzorganisationen gegen die Verbreitung marktrelevanter Desinformation vorgehen können (Art. 2 und insb. Art. 3 Abs. 1 lit. b, lit. d und lit. i UWG; Art. 9 f. UWG). Allerdings fehlt es in diesem Bereich oft an der Rechtsdurchsetzung.

[36] Im Vorfeld von Volksabstimmungen und -wahlen stehen besondere Mechanismen zur Verfügung, um – unabhängig von den genutzten Informationskanälen – gegen schwerwiegende Desinformation vorzugehen. Die entsprechenden rechtlichen Rahmenbedingungen umhegen sowohl Informationsflüsse seitens der Behörden wie auch unter Privaten.[66] Behörden müssen auch auf den sozialen Medien stets objektiv und sachlich kommunizieren; zudem muss die Herkunft der Information ersichtlich sein und die Behörden haben den Verhältnismässigkeitsgrundsatz zu beachten (Art. 10*a* Abs. 2 BPR für eidgenössische Volksabstimmungen). Für die Informationstätigkeit durch Private gilt grundsätzlich der Schutz der Meinungsäusserung durch die Kommunikationsgrundrechte. Eine schwerwiegende Irreführung über zentrale Abstimmungsinhalte durch Private – in den sozialen Medien oder über andere Kanäle – wird in erster Linie einen Anspruch auf Richtigstellung der betreffenden Informationen begründen.[67] Die Behörden können allerdings

64 Sylvain Métille, Évaluation de la régulation existante et des options de la régulation concernant les intermédiaires d'information en Suisse, Étude juridique réalisée à la demande de l'OFCOM, Lausanne 2020, 70, 77, 88; Denis Masmejan, Débat public en ligne et protection des libertés de communication, Étude réalisée sur mandat de l'Office fédéral de la communication, Genf 2020, 13 ff. Beide Studien sind abrufbar unter https://www.bakom.admin.ch/bakom/de/home/elektronische-medien/studien/einzelstudien.html (14. Juni 2021). Das Bundesamt für Kommunikation (BAKOM) ist daran, gemeinsam mit der Bundeskanzlei (BK) einen Governance-Bericht zuhanden des Bundesrates zu erarbeiten, in dem u. a. die Frage von weiterem Regulierungsbedarf behandelt werden soll.

65 Bundesgesetz gegen den unlauteren Wettbewerb vom 19. Dezember 1986, SR 241.0.

66 Siehe dazu Michel Besson, Behördliche Information vor Volksabstimmungen, Bern 2002; Andrea Töndury, Intervention oder Teilnahme? Möglichkeiten und Grenzen staatlicher Kommunikation im Vorfeld von Volksabstimmungen, Schweizerisches Zentralblatt für Staats- und Verwaltungsrecht 2011, 341–374; Benedikt Pirker, Behördliche Interventionen in Abstimmungskämpfe, AJP 2017, 1366–1381.

67 Zur Intervention von Behörden zur Richtigstellung von privater Information siehe z. B. BGE 140 I 338, E. 5.3; BGE 132 I 104, E. 4.1.

unter bestimmten Voraussetzungen zur Intervention und im Extremfall sogar zur Aufhebung des Resultats einer Volksabstimmung berechtigt bzw. verpflichtet sein.[68]

[37] Bei der Erfassung einer Manipulation, die nicht auf Desinformation beruht, steht das Recht noch ganz am Anfang. Eine erste durchaus komplexe Herausforderung besteht darin, Kriterien zu identifizieren, die es erlauben, eine von der Rechtsordnung hinzunehmende Beeinflussung von Nutzerinnen und Nutzern von einer rechtlich problematischen Manipulation zu unterscheiden. Auch diesbezüglich wird relevant sein, ob das Denken und Handeln von Menschen in einem politischen oder kommerziellen Kontext mittels eines Empfehlungsalgorithmus oder einer anderen KI-Anwendung beeinflusst wird. Unter welchen Umständen sogar eine «blosse» Modellierung einer Entscheidungssituation als rechtlich relevante Manipulation zu qualifizieren wäre, bleibt damit aber noch offen.

[38] Wie beim Verbreiten von Desinformation können sich Konsumentinnen und Konsumenten grundsätzlich auf das UWG stützen, wenn ihre Entscheidung über den Erwerb eines Produktes mithilfe von Empfehlungsalgorithmen oder durch andere KI-Anwendungen in rechtlich relevanter Weise manipuliert wird. Anders als zur in mehreren Bestimmungen des UWG geregelten Irreführung gibt es hierzu in der Schweiz aber kaum relevante Rechtsprechung. Folglich ist gänzlich unklar, in welchen Konstellationen Gerichte eine rechtlich relevante Manipulation annehmen würden. Bei personalisierter Werbung[69] dürfte dies wohl ebenso zu verneinen sein wie bei personalisierten Preisen.[70] Anderes dürfte gelten, wenn Entscheidungssituationen derart modelliert werden, dass Konsumentinnen und Konsumenten (vermeintlich) keine Wahl haben. Völlig unklar ist die Rechtslage bei nicht marktrelevanten Manipulationen. Denkbar wäre, eine relevante Einflussnahme auf Entscheidungen als Eingriff in die Autonomie der Individuen zu verstehen und als Persönlichkeitsverletzung im Sinn von Art. 28 ZGB zu qualifizieren. Ob hier aber überhaupt Bedarf an einem Zugriff des Rechts besteht und unter welchen Voraussetzungen ein persönlichkeitsrechtlich relevanter Eingriff in die Autonomie angenommen werden könnte, ist erst noch vertieft zu untersuchen. Das blosse Anzeigen weiterer Inhalte mit dem Ziel, die Nutzerinnen und Nutzer möglichst lange auf einer Plattform zu halten, dürfte jedenfalls nicht genügen.

[39] Es zeigt sich, dass das Schweizer Recht hinreichend flexibel ist, um Probleme zu erfassen, die durch Manipulationen mittels KI entstehen können. Damit besteht kein Anlass, die entsprechende Regelung im EU-KI-Verordnungsvorschlag zu übernehmen. Dieser sieht in Art. 5 Abs. 1 lit. a und b ein Verbot für gewisse Formen des manipulativen Gebrauchs von KI und für KI-Anwendungen vor, die auf die Beeinflussung gewisser, besonders verletzlicher Bevölkerungsgruppen zielen. Insbesondere der erste Tatbestand ist derart offen formuliert, dass auch harmlose Praktiken erfasst werden könnten – unter Umständen sogar das durch KI gesteuerte Ausspielen von Werbung. Beim zweiten Tatbestand werden zudem die Aspekte von Manipulation und Dis-

68 Siehe dazu etwa BGE 140 I 338, E. 5.3; BGE 135 I 292, E. 4.1; Urteil des Bundesgerichts 1C_472/2010 vom 20. Januar 2011, E. 4.

69 Florent Thouvenin, Datenschutz auf der Intensivstation, digma 2019, 206–213, 208. Siehe dazu auch Rolf H. Weber, Online-Marketing und Datenschutz, digma 2012, 110–115, 111. Zum deutschen Recht siehe Martin Ebers, Beeinflussung und Manipulation von Kunden durch Behavioral Microtargeting, Verhaltenssteuerung durch Algorithmen aus der Sicht des Zivilrechts, MMR 2018, 423–428, 424 f.

70 Näheres dazu bei Florent Thouvenin, Dynamische Preise, Eine Herausforderung für das Datenschutz-, Wettbewerbs- und Vertragsrecht, in: Jusletter IT 22. September 2016, Rz. 29–46; Roland Mathys/Helen Reinhart, Bestimmung von Vertragskonditionen im Rahmen automatisierter Entscheidungen, SZW 2020, 35–42, 40 ff.; Michael Isler, Meine Daten machen meinen Preis, digma 2015, 18–23, 23.

kriminierung in problematischer Weise vermischt. Angesichts dieser Schwächen erscheint fraglich, ob diese Vorschläge im angelaufenen Legislativprozess der EU Bestand haben werden.

2.5. Haftung und Verantwortlichkeit

[40] Eine zentrale Herausforderung beim Einsatz von KI sind die Klärung der zivilrechtlichen Haftung im Schadensfall und die Sicherstellung der strafrechtlichen Verantwortlichkeit.[71] Ein plastisches Beispiel ist das automatisierte Fahrzeug, das etwa aufgrund einer fehlerhaften Objekterkennung einen Unfall verursacht. Während bei einem manuell gesteuerten Fahrzeug der Mensch für die Beherrschung des Fahrzeugs (inklusive Objekterkennung und Kollisionsvermeidung) zuständig ist, übernimmt bei hochgradig automatisierten Fahrzeugen die Maschine die Steuerung. Beim automatisierten Fahrzeug kommt als zivilrechtliches Haftungssubjekt weiterhin die kausal haftende Halterin bzw. der Halter in Betracht (Art. 58 SVG).[72] Verstärkt in den Fokus rücken wird allerdings neu die Herstellerin, typischerweise ein Unternehmen, das bislang nur in den seltenen Fällen eines schadensursächlichen technischen Defekts (z. B. eines geplatzten Reifens) belangt werden konnte. Eine weitere Einflussgrösse ist beim vernetzten Fahrzeug ausserdem die (mangelnde) Qualität der verwendeten Daten, womit u. a. aufgrund externer Datenzulieferung Haftungssubjekte hinzutreten können.[73]

[41] Auf europäischer Ebene befasst man sich seit mehreren Jahren intensiv mit Haftungsfragen für Schäden, die durch KI verursacht sind.[74] Da insbesondere die Überarbeitung des harmonisierten Produkthaftungsrechts gesondert diskutiert wird, äussert sich der EU-KI-Verordnungsvorschlag nicht zu Haftungsfragen.[75] Allerdings handelt es sich beim Schweizer Produkthaftungsgesetz (PrHG) um eine Abbildung der europäischen Richtlinie 85/374/EWG über die Haftung für fehlerhafte Produkte, weswegen die laufende europäische Diskussion zur Haftung auch in die Schweiz hineinwirkt und aufmerksam verfolgt werden sollte. Anknüpfend an die europäischen Entwicklungen ist nachfolgend auf ausgewählte Aspekte der spezialgesetzlichen Produkthaftung nach PrHG einzugehen, das in der Praxis allerdings aufgrund verschiedener Einschränkungen oftmals durch die deliktische Produzentenhaftung nach Art. 55 OR verdrängt wird. Mit bestimmten Anpassungen könnte das PrHG im Lichte der technischen Entwicklungen an Bedeutung gewinnen.

---

71 European Commission Expert Group on Liability and New Technologies, Liability for Artificial Intelligence and Other Emerging Digital Technologies, 2019, 11 f.; European Parliamentary Research Service, A common EU approach to liability rules and insurance for connected and autonomous vehicles, 2008, http://www.europarl.europa.eu/RegData/etudes/STUD/2018/615635/EPRS_STU(2018)615635_EN.pdf (26. Mai 2021), 8; Peter M. Asaro, The Liability Problem for Autonomous Artificial Agents, AAAI Spring Symposia, 2016, 190–194, 191; Andrea Bertolini, Robots as Products: The Case for a Realistic Analysis of Robotic Applications and Liability Rules, Law Innovation and Technology 2013, 214–247, 235 f. Siehe auch Ben Wagner, Liable, But Not in Control? Ensuring Meaningful Human Agency in Automated DecisionMaking Systems, Policy & Internet 2019, 104–122, https://doi.org/10.1002/poi3.198, passim; Monika Simmler/Olivia Zingg, Rechtliche Aspekte sozialer Roboter, Gutachten im Auftrag der TA-SWISS, 2011, 27 ff.

72 Dazu Melinda F. Lohmann, Automatisierte Fahrzeuge im Lichte des Schweizer Zulassungs- und Haftungsrechts, Baden-Baden 2016, 211 ff.

73 Ausführlich Herbert Zech, Entscheidungen digitaler autonomer Systeme: Empfehlen sich Regelungen zu Verantwortung und Haftung?, Gutachten für den 73. Deutschen Juristentag, München 2020, A 52.

74 Siehe Überblick bei Melinda F. Lohmann, Ein zukunftsfähiger Haftungsrahmen für Künstliche Intelligenz, Warum die Schweiz ihr Produkthaftungsrecht aktualisieren muss, HAVE 2021, 111–120, 113 f.; Melinda F. Lohmann, Ein europäisches Roboterrecht – überfällig oder überflüssig?, ZRP 2017, 168–171.

75 Der EU-KI-Verordnungsvorschlag findet auch (fast) keine Anwendung auf Fahrzeuge, siehe dessen Art. 2 Ziff. 2.

[42] Das Produkthaftungsrecht ist auf herkömmliche Produkte und damit auf physische Gegenstände zugeschnitten, die nach der Herstellung einmalig in den Verkehr gebracht und danach nicht mehr beeinflusst werden.[76] Auf KI-Anwendungen passt diese Konzeption nur schlecht, wenn es um die Bestimmung der Produkteigenschaft von Software, die Fehlerhaftigkeit von Entscheidungen, die Entlastungsmöglichkeiten oder die Nachmarktpflichten der Herstellerin geht.[77] Zu klären ist ferner die Einordnung digitaler Dienste (sog. Digital Services), bspw. vernetzter Navigationssysteme, die das Produkthaftungsgesetz wegen ihrer Eigenschaften als Dienstleistungen grundsätzlich nicht erfasst.[78] Diskutiert wird sodann, ob der vom Produkthaftungsrecht erfasste Tatbestand des Sachschadens nicht auch Vermögensschäden, die etwa im Zuge von Datenverlust entstehen, einbeziehen und ob der Sachschaden auch bei gewerblicher Nutzung gedeckt sein sollte.[79] Auch die Position der Herstellerin verändert sich angesichts der Vielzahl von Personen, die das Design, die Funktionsweise und die Nutzung von KI-Anwendungen beeinflussen, was eine Überprüfung des Begriffs erforderlich macht.[80]

[43] Zeitnah geklärt werden muss primär die (umstrittene) Frage, ob die isolierte Steuerungssoftware als (Teil)Produkt qualifiziert werden soll. Diese Frage hat erhebliche Auswirkungen auf die Haftung. Für das Schweizer Recht sprechen sich zu Recht zahlreiche Lehrmeinungen für die Produkteigenschaft von Software aus.[81] Software in jeglicher Form ist als «typische Erscheinung der fortschreitenden Technisierung» unter den Anwendungsbereich des Produktehaftpflichtgesetzes zu subsumieren.[82] Die Auslegung des Begriffs hat nach produkthaftpflichtrechtlichen und nicht

76 Lohmann (Fn. 74, Künstliche Intelligenz), 119.

77 Lohmann (Fn. 74, Künstliche Intelligenz), 119.

78 Siehe dazu Astrid Seehafer/Joel Kohler, Künstliche Intelligenz: Updates für das Produkthaftungsrecht?, EuZW 2020, 213–218, 214; siehe auch European Commission Expert Group on Liability and New Technologies (Fn. 71), 28; Andrea Bertolini, Artificial Intelligence and Civil Liability, Legal Affairs, Policy Department for Citizens' Rights and Constitutional Affairs (Hrsg.), Study requested by the JURI committee, PE 621.926, Juli 2020, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf (17. Mai 2021), 57.

79 Bejahend Friedrich von Westphalen, Haftungsfragen beim Einsatz künstlicher Intelligenz in Ergänzung der Produkthaftungs-RL 85/374/EWG, ZIP 2019, 889–895, 894 f.; Bernhard A. Koch, Product Liability 2.0 – Mere Update or New Version?, in: Sebastian Lohsse/Reiner Schulze/Dirk Staudenmayer (Hrsg.), Liability for Artificial Intelligence and the Internet of Things, Baden-Baden 2019, 99–116, 103 f.; ebenso Susana Navas, Robot Machines and Civil Liability, in: Martin Ebers/Susana Navas (Hrsg.), Algorithms and Law, Cambridge 2020, 157–173, 172 f.; eine Ausdehnung des Schutzbereichs auf gewerbliche oder berufliche Produktbenutzer befürwortend Walter Fellmann, Haftpflichtrecht im Zeichen der Digitalisierung, HAVE 2021, 105–111, 109.

80 Siehe European Commission Expert Group on Liability and New Technologies (Fn. 71), 28.

81 Jüngst Fellmann (Fn. 79), 107; Lohmann (Fn. 74, Künstliche Intelligenz), 115; Walter Fellmann, in: Corinne Widmer Lüchinger/David Oser (Hrsg.), Obligationenrecht I, Basler Kommentar, 7. Aufl., Basel 2020, PrHG 3 N 10; Silvio Hänsenberger, Die Haftung für Produkte mit lernfähigen Algorithmen, in: Jusletter 26. November 2018, Rz. 14; Barbara Klett, Digitalisierte Gesundheit – Abgrenzungen und Regulierung, HAVE 2017, 104–113, 112; Heinz Rey/Isabelle Wildhaber, Ausservertragliches Haftpflichtrecht, 5. Aufl., Zürich 2018, § 8 Rz. 1427; Vito Roberto, Haftpflichtrecht, 2. Aufl., Bern 2018, § 9 Rz. 09.10; Ingeborg Schwenzer/Christiana Fountoulakis, Schweizerisches Obligationenrecht: Allgemeiner Teil, 8. Aufl., Bern 2020, § 53 Rz. 53.35; Joachim Hess, Produktehaftpflichtgesetz (PrHG), SHK Stämpflis Handkommentar, 3. Aufl., Bern 2016, PrHG 3 N 30–34. A. A. etwa Heinrich Honsell/Bernhard Isenring/Martin A. Kessler, Schweizerisches Haftpflichtrecht, 5. Aufl., Zürich 2013, § 21 Rz. 31; Claire Huguenin, Obligationenrecht – Allgemeiner und Besonderer Teil, 3. Aufl., Zürich 2019, § 24 Rz. 2110; siehe auch Corinne Widmer Lüchinger, Apps, Algorithmen und Roboter in der Medizin: Haftungsrechtliche Herausforderungen, HAVE 2019, 3–15, 7, allerdings mit Kritik zu dieser Leseart.

82 Fellmann (Fn. 79), 107.

nach sachenrechtlichen Kriterien zu erfolgen und Software kann durchaus eine produkttypische Schädigungsgefahr schaffen.[83] Daher empfiehlt sich, diese Deutung im Gesetz zu verankern.[84]

[44] Klärungsbedürftig ist sodann die Übertragung der herkömmlichen Massstäbe zur Bestimmung der Fehlerhaftigkeit auf KI-Anwendungen: Worin liegt der Produktfehler bei einem System, das gerade aus seinen Fehlern lernen soll und bei dem je nach Konfiguration ein Teil des verhaltensbestimmenden Trainings im Rahmen der Nutzung und damit ausserhalb des Verantwortungsbereichs der Herstellerin stattfindet?[85] Die Fehlerhaftigkeit eines Produkts liegt darin, dass es nicht die Sicherheit bietet, die man unter Berücksichtigung aller Umstände erwarten darf (Art. 4 Abs. 1 PrHG).[86] Dass eine bestimmte Produkteigenschaft typisch und möglicherweise unvermeidbar ist, schliesst das Vorliegen eines Produktfehlers nicht aus.[87] Gerade bei sicherheitskritischen KI-Anwendungen werden die Sicherheitserwartungen berechtigterweise besonders hoch sein.[88]

[45] Die Fehlerhaftigkeit bestimmt sich heute nach dem Zeitpunkt des Inverkehrbringens (Art. 4 Abs. 1 lit. c PrHG); gehaftet wird nicht für Fehler, die erst nach Inverkehrbringen entstanden sind (Art. 5 Abs. 1 lit. b PrHG). Bei einer auf eine mangelnde Risikobegrenzung zurückgehenden Fehlentscheidung im Rahmen einer KI-Anwendung liegt allerdings bereits zum Zeitpunkt des Inverkehrbringens ein Konstruktionsfehler vor.[89] Bei adaptiven Systemen beruhen die Lernprozesse auf Algorithmen, die zum Zeitpunkt des Inverkehrbringens Bestandteil des Produktes waren, «aber erst nach diesem Zeitpunkt neue und eigenständige Lösungen entwickeln».[90] Indem die Herstellerin die Weichen für die Entwicklungsfähigkeit legt, schafft sie die Gefahr einer aus dieser Fähigkeit resultierenden Schädigung.[91] Damit wird jedoch nicht etwa die Lernfähigkeit mit einem Produktfehler gleichgesetzt.[92] Letzterer gründet vielmehr in der Standardabweichung,[93]

83 Lohmann (Fn. 72), 316; siehe auch Saskia Wittbrodt, Industrie 4.0 und die Haftung für Maschinensoftware, InTer 2020, 74–81, 76.

84 Fellmann (Fn. 79), 107.

85 Lohmann (Fn. 74, Künstliche Intelligenz), 115.

86 Dazu BGE 133 III 81, E. 3.1; Roberto (Fn. 81), § 9 Rz. 09.11; Honsell/Isenring/Kessler (Fn. 81), § 21 Rz. 32; Schwenzer/Fountoulakis (Fn. 81), § 53 Rz. 53.37.

87 Lohmann (Fn. 74, Künstliche Intelligenz), 116.

88 Susanne Wende, Haftungsfragen bei vernetzten und autonomen Systemen, in: Thomas Sassenberg/Tobias Faber (Hrsg.), Rechtshandbuch Industrie 4.0 und Internet of Things, 2. Aufl., München 2020, 93–121, 106; siehe auch von Westphalen (Fn. 79), 893; siehe dazu auch die Definition von «high-risk» KI-Anwendungen im EU-KI-Verordnungsvorschlag.

89 So auch Jan-Philipp Günther, Roboter und rechtliche Verantwortung, München 2016, 181 f.; Melinda F. Lohmann, Roboter als Wundertüten – eine zivilrechtliche Haftungsanalyse, AJP 2017, Sonderheft Roboterrecht, 152–162, 158; Zech (Fn. 73), A 70; siehe auch BSK-Fellmann (Fn. 81), PrHG 4 N 28e; ferner auch Simmler/Zingg (Fn. 71), 29.

90 Seehafer/Kohler (Fn. 78), 215.

91 Lohmann (Fn. 74, Künstliche Intelligenz), 116 f.; so auch David Rosenthal, Autonome Informatiksysteme: Wie steht es mit der Haftung, in: Albert Kündig/Danielle Bütschi (Hrsg.), Die Verselbständigung des Computers, TA-SWISS 51/2008, Zürich 2008, 131–144, 133; Zech (Fn. 73), A 35; BSK-Fellmann (Fn. 81), PrHG 4 N 28e: «Jeder Hersteller muss das Gefahrenpotenzial beherrschen.»

92 So aber Hänsenberger (Fn. 81), Rz. 25; Widmer Lüchinger (Fn. 81), 11.

93 Siehe Gerhard Wagner, Roboter als Haftungssubjekte? Konturen eines Haftungsrechts für autonome Systeme, in: Florian Faust/Hans-Bernd Schäfer (Hrsg.), Zivilrechtliche und rechtsökonomische Probleme des Internet und der künstlichen Intelligenz, Tübingen 2019, 1–40, 14.

also der aus einer ungenügenden Absicherung resultierenden Verletzung der Sicherheitserwartungen.[94]

[46] Im Zentrum der Haftungsfrage steht die Abgrenzung der Risikosphären aufgrund von Herstellung und Nutzung: Grundsätzlich erschafft die Herstellerin die KI und kann daher das Risiko der Gefährdung durch sorgfältige Programmierung und Instruktion beherrschen.[95] Bei lernfähigen KI-Anwendungen, die nach dem Inverkehrbringen aufgrund der Nutzung trainiert werden, verringert sich jedoch ihre Kontrolle.[96] Hingegen können Nutzerin und Nutzer durch die Auswahl des Lernverfahrens, der Trainingsdaten und der Dauer des Lernprozesses auf die KI einwirken.[97] Führt die Einflussnahme der Nutzerin bzw. des Nutzers zu einer schadensursächlichen Fehlentscheidung, scheint es unter Risikogesichtspunkten angemessen, sie bzw. ihn haften zu lassen.[98]

[47] Entsprechend der Abgrenzung der Risikosphären kann sich die Herstellerin, soweit sie für die sichere (Weiter)Entwicklung ihres Produkts haftet, bei unsachgemässer Einflussnahme durch andere Beteiligte entlasten.[99] Sie haftet nicht für Fehler, die durch «unsachgemässe Änderung» ihres Produkts nach Inverkehrbringen entstehen (Art. 5 Abs. 1 lit. b PrHG),[100] weil das Produkt im Zeitpunkt des Inverkehrbringens den Machtbereich der Herstellerin verlässt.[101] Steht die Fehlerhaftigkeit einer KI-Anwendung fest, könnte sie sich durchaus auf eine nachträgliche Veränderung durch unsachgemässes Training seitens der Nutzerin oder des Nutzers berufen.[102] Auch diesbezüglich empfiehlt sich allerdings eine gesetzgeberische Klarstellung.

[48] Eine weitere Entlastungsmöglichkeit besteht sodann für sog. Entwicklungsrisiken (Art. 5 Abs. 1 lit. e PrHG), d. h. für unvorhersehbare Risiken, die im Zeitpunkt des Inverkehrbringens des Produkts nach dem damaligen Stand der Wissenschaft sowie der Technik nicht erkennbar waren.[103] Ein solches Entwicklungsrisiko liegt nur vor, wenn die Gefährlichkeit des Produkts im Zeitpunkt, als eine Schadensabwendung in Betracht kam, weder bekannt noch nach dem Stand von Wissenschaft und Technik erkennbar war (bspw. bei den Asbestfällen, bei HIV-infizierten Blutkonserven und bei den Contergan-Fällen[104]).[105]

---

94 Siehe dazu Melinda F. Lohmann/Markus Müller-Chen, Selbstlernende Fahrzeuge – eine Haftungsanalyse, SZW 2017, 48–58, 55.

95 Lohmann (Fn. 74, Künstliche Intelligenz), 120.

96 Zech (Fn. 73), A 89; Lohmann (Fn. 89), 158.

97 Zech (Fn. 73), A 35 ff.; Lohmann (Fn. 89), 158.

98 Siehe auch die Pflichten der Nutzerinnen und Nutzer nach Art. 29 EU-KI-Verordnungsvorschlag, die allerdings angesichts der gesonderten Abhandlung des Haftungsrechts nicht auf die Haftung durchschlagen. In bestimmten Sektoren greift beim Einsatz von künstlicher Intelligenz bereits eine verschuldensunabhängige Nutzerhaftung, z. B. bei den eingangs erwähnten automatisierten Fahrzeugen (Art. 58 SVG), dazu Lohmann (Fn. 72), 211 ff. Fehlt eine solche, wird den Geschädigten der Nachweis einer unerlaubten, insbesondere fahrlässigen und schadensursächlichen Handlung des Nutzers gestützt auf Art. 41 OR regelmässig schwerfallen (Melinda F. Lohmann (Fn. 89), 158 f.; siehe auch Simmler/Zingg (Fn. 71), 43 f.).

99 Lohmann (Fn. 74, Künstliche Intelligenz), 118.

100 BSK-Fellmann (Fn. 81), PrHG 5 N 6; Huguenin (Fn. 81), § 24 Rz. 2119; Rey/Wildhaber (Fn. 81), § 8 Rz. 1464.

101 Huguenin (Fn. 81), § 24 Rz. 2119; Rey/Wildhaber (Fn. 81), § 24 Rz. 2119.

102 A. A. Fellmann (Fn. 79), 108; BSK-Fellmann (Fn. 81), PrHG 4 N 28e; offengelassen bei Clara-Ann Gordon/Tanja Lutz, Haftung für automatisierte Entscheidungen – Herausforderungen in der Praxis, SZW 2020, 53–61, 58 f.

103 BGE 137 III 226, E. 4.1; BSK-Fellmann (Fn. 81), PrHG 5 N 15; Huguenin (Fn. 81), § 24 Rz. 2122.

104 Anne Marie Meermann, Entwicklungsrisiko und State-of-the-Art, Diss. Hamburg, Baden-Baden 2013, 35 ff.

105 Hans-Bernd Schäfer/Claus Ott, Lehrbuch der ökonomischen Analyse des Zivilrechts, 6. Aufl., Berlin 2020, 421; Schwenzer/Fountoulakis (Fn. 81), § 53 Rz. 53.36. In BGE 137 III 226 bejahte das Bundesgericht die Haftungs-

[49] Die einzelnen Aktionen einer KI mögen unvorhersehbar sein, dies gilt jedoch nicht für ihre Gefährlichkeit. Diese ist in der Wissenschaft abstrakt bekannt, weshalb hier nicht von einem Entwicklungsrisiko, sondern von einer Entwicklungslücke auszugehen ist.[106] Eine Entwicklungslücke stellt nicht auf die Grenzen menschlicher Erkenntnisfähigkeit, sondern auf die Grenzen technologischer Gefahrvermeidungsmöglichkeiten ab.[107] Die Herstellerin kann sich bei Vorliegen einer solchen Lücke nicht unter Berufung auf ein Entwicklungsrisiko für Fehlentscheidungen einer KI entlasten.[108]

[50] Zur Schliessung möglicher zivilrechtlicher Haftungslücken werden verschiedene (ausservertragliche) Lösungsansätze diskutiert, die allerdings nicht alle gleichermassen zu überzeugen vermögen:[109] die Übertragung bestehender Haftungsnormen per Analogie,[110] die Einführung weiterer sektorspezifischer Gefährdungshaftungen aufgrund der Nutzung[111] oder Herstellung[112] von KI bzw. die Einführung einer allgemeinen Gefährdungshaftung kombiniert mit einer Versicherungspflicht[113] sowie die Schaffung einer Eigenhaftung von KI (auch als ePersonhood bezeichnet).[114] Eine für sich allein stehende Lösungsmöglichkeit wäre sodann das vollständige Ersetzen einer Haftung durch einen Versicherungsschutz für Unfälle, an denen KI beteiligt ist.[115]

[51] Nicht nur im Haftpflichtrecht, d. h. bei der Bestimmung der Schadensersatzpflicht, stellen «Unfälle» mit Beteiligung von KI eine Herausforderung dar. Auch die strafrechtliche Verantwortungszuschreibung kann bei entsprechenden Sachverhalten an ihre Grenzen stossen. Eine Strafbarkeit für die fahrlässige Herbeiführung eines Schadens aufgrund der Anwendung einer KI

---

entlastung der Herstellerin einer vorzeitig abgenutzten Hüftprothese unter Hinweis darauf, dass die vorzeitige Abnutzung vor der Schädigung nicht Gegenstand wissenschaftlicher Veröffentlichungen gewesen sei (E. 4.2); dazu Barbara Klett/Dominique Müller, Rechtsentwicklung zum PrHG und PrSG, HAVE 2018, 438–442, 440.

106 Lohmann (Fn. 74, Künstliche Intelligenz), 119; Herbert Zech, Künstliche Intelligenz und Haftungsfragen, ZfPW 2019, 198–219, 213; siehe auch Malte Grützmacher, Die deliktische Haftung für autonome Systeme – Industrie 4.0 als Herausforderung für das bestehende Recht?, CR 2016, 695–698, 696; von Westphalen (Fn. 79), 893; ein Entwicklungsrisiko nur ausnahmsweise in Betracht ziehend BSK-Fellmann (Fn. 81), PrHG 4 N 28e; a. A. Alexia Sidiropoulos, Haftung für Gerätefehler bei der medizinischen Diagnostik und Behandlung, Sicherheit & Recht 2020, 49–56, 51; Hänsenberger (Fn. 81), Rz. 21; Seehafer/Kohler (Fn. 78), 215 f.

107 Meermann (Fn. 104), 33; SHK-Hess (Fn. 81), PrHG 4 N 70; siehe auch Olaf Sosnitza, Das Internet der Dinge – Herausforderung oder gewohntes Terrain für das Zivilrecht?, CR 2016, 764–772, 769 f.

108 Siehe Zech (Fn. 106), 213; Grützmacher (Fn. 106), 696.

109 Lohmann (Fn. 74, Künstliche Intelligenz), 120.

110 Etwa die Haftung des Tierhalters; zum Ganzen Lohmann (Fn. 89), 159 ff.

111 Günther (Fn. 89), 237 ff.; Zech (Fn. 73), A 98 ff., möchte die Nutzerhaftung auf professionelle Nutzer beschränken, also auf «Betreiber, deren hauptsächlicher Unternehmensgegenstand der Betrieb digitaler Systeme ist [...].» Das EU-Parlament unterscheidet zwischen «Frontend-» und «Backend-Betreibern»: Entschliessung des Europäischen Parlaments vom 20. Oktober 2020 mit Empfehlungen an die Kommission für eine Regelung der zivilrechtlichen Haftung beim Einsatz künstlicher Intelligenz (2020/2014(INL)), E. 12, Art. 3(d)–(f).

112 De lege lata haftet die Herstellerin nach PrHG nur für *fehlerhafte* Produkte; zur (umstrittenen) Qualifikation der spezialgesetzlichen Produkthaftung als gewöhnliche Kausalhaftung Rey/Wildhaber (Fn. 81), § 8 Rz. 1410; zu den unterschiedlichen Lehrmeinungen Lohmann (Fn. 72), 390. De lege ferenda wäre die blosse Verursachung des Schadens durch ein Produkt haftungsauslösend, womit die Schwierigkeiten bei der Bestimmung der Fehlerhaftigkeit umgangen würden; siehe etwa Zech (Fn. 106), 214; ausführlich zu den Vor- und Nachteilen einer strikten Herstellerinnenhaftung Wagner (Fn. 93), 18 ff.

113 Dazu etwa Lohmann (Fn. 89), 161; Jochen Hanisch, Haftung für Automation, Diss. Erlangen-Nürnberg, Göttingen 2010, 205 f.; Gunther Teubner, Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten, AcP 2018, 155–205, 191 ff., fordert stattdessen die Einführung einer «digitalen Assistenzhaftung», d. h., einer strikten deliktischen Haftung des Prinzipals für Fehlverhalten der KI.

114 Dazu Ruth Janal, Extra-Contractual Liability for Wrongs Committed by Autonomous Systems, in: Martin Ebers/Susana Navas (Hrsg.), Algorithms and Law, Cambridge 2020, 174–206, 175 f.; Thomas Wischmeyer, Regulierung intelligenter Systeme, AöR 2018, 1–66, 37 ff.

115 Dazu Zech (Fn. 73), A 105 ff.

hängt einerseits davon ab, ob deren Einsatz als Sorgfaltswidrigkeit oder als sog. erlaubtes Risiko qualifiziert wird.[116] Wo genau die Grenze verläuft, ist jedoch noch weitgehend unklar. Welche Sorgfaltspflichten haben im Hinblick auf die Herstellung, den Vertrieb und die Nutzung von KI zu gelten?[117] Sorgfaltspflichten müssen sicherlich *ex ante* erkennbar sein. Eine faktische Kausalhaftung im Schadensfall wäre mit dem im Strafrecht hochgehaltenen Schuldprinzip nicht vereinbar.[118] Dieses besagt, dass nur ein persönlich als schuldhaft vorwerfbares Verhalten zu einer Strafe führen darf.

[52] Andererseits ist mit Blick auf KI, deren Funktionsweise mit der weiteren Entwicklung unvorhersehbarer werden mag, fraglich, ob *de lege lata* überhaupt noch ein Rechtsadressat in die Pflicht genommen werden könnte. Die Verselbständigung des technischen Systems, das selbst kein Strafrechtssubjekt darstellt,[119] kann Kausal- und damit Zurechnungsketten «durchbrechen». Die Verantwortungszuschreibung wird dadurch erschwert, wobei in Anwendung der bestehenden Zurechnungslehre auch ungewollte Verantwortlichkeitslücken entstehen könnten.[120] Ebenso wahrscheinlich ist eine Vorverlagerung des Vorwurfs: Jene, die KI entwickeln, herstellen oder implementieren, würden in solchen Fällen bereits für den Einsatz von KI zur Rechenschaft gezogen.[121] Auch hier ergäben sich jedoch Unklarheiten, bspw. mit Blick auf die Grenzen der Unternehmensstrafbarkeit oder der fahrlässigen Mittäterschaft.[122]

[53] Die Grundlagen der Haftung und Verantwortlichkeit beim Einsatz von KI bedürfen, wie sich gezeigt hat, der Klärung. Zumindest im Produktehaftpflichtrecht sollte sich die Schweiz (anders als in der Vergangenheit) nicht ausschliesslich darauf beschränken, den Weg nachzuvollziehen, den das EU-Recht nun vorzuzeichnen beginnt. Sollte eine Klärung im Schweizer Recht nicht gelingen, drohen willkürliche und für die Normadressaten unberechenbare Verantwortungszuweisungen, die Innovation behindern können. Ähnliche Schwierigkeiten zeigen sich auch im Strafrecht, das seit jeher stark durch nationale Eigenheiten geprägt ist. Die Herausforderungen bei der Regelung von Haftung und Verantwortlichkeit im Zivil- und Strafrecht weisen auf die Notwendigkeit eines baldigen rechtspolitischen Diskurses hin.

---

116 Zur Diskussion des erlaubten Risikos bei KI im Strassenverkehr Nadine Zurkinden, Strafrecht und selbstfahrende Autos – ein Beitrag zum erlaubten Risiko, recht 2016, 144–156.

117 Vergleichbar auch die Fragen, die sich in Bezug auf (autonome) soziale Roboter stellen, siehe dazu Simmler/Zingg (Fn. 71), 19 ff. und 48 ff.

118 Monika Simmler, Strafrechtliche Verantwortung im Zeitalter autonomer Technik: Vom Individual- zum Unternehmensstrafrecht?, in: Daniel Fink/Jörg Arnold/Joëlle Vuille/Niklaus Oberholzer (Hrsg.), Strafjustiz zwischen künstlicher Intelligenz und prädiktiven Algorithmen, Basel 2021 (erscheint demnächst).

119 Zur Möglichkeit einer originären strafrechtlichen Verantwortlichkeit von Technik siehe aber m. w. N. Monika Simmler/Nora Markwalder, Roboter in der Verantwortung? – Zur Neuauflage der Debatte um den funktionalen Schuldbegriff, ZStW 2017, 20–47, 22; Nora Markwalder/Monika Simmler, Roboterstrafrecht, Zur strafrechtlichen Verantwortlichkeit von Robotern und künstlicher Intelligenz, AJP 2017, 171–182, 172.

120 In der Literatur wird deshalb auch das Risiko einer «Responsibility Gap» oder «Lawlessness» identifiziert; siehe u. a. Susanne Beck, Google Cars, Software Agents, Autonomous Weapons Systems – New Challenges for Criminal Law?, in: Eric Hilgendorf/Uwe Seidel (Hrsg.), Robotics, Autonomics, and the Law, Baden-Baden 2017, 227–251, 232; Andreas Matthias, The responsibility gap: Ascribing responsibility for the actions of learning automata, Ethics and Information Technology 2004, 175–183, passim; Robert Sparrow, Killer Robots, Journal of Applied Philosophy 2007, 62–77, passim; Mark A. Chinen, The Co-Evolution of Autonomous Machines and Legal Responsiblity, Virginia Journal of Law & Technology 2016, 338–393, passim; Jack M. Balkin, The Path of Robotics Law, California Law Review Circuit 2015, 45–60, passim.

121 Siehe zu dieser Verlagerungstendenz ausführlich und auf empirischer Grundlage Simmler (Fn. 118).

122 Zu den entsprechenden Herausforderungen bereits Simmler/Zingg (Fn. 71), 55 ff.; Simmler (Fn. 118).

3. Schlussfolgerung

[54] Die Skizzierung der Herausforderungen von KI und der Blick auf den Verordnungsvorschlag der EU-Kommission haben gezeigt, dass (auch) im Schweizer Recht Handlungsbedarf besteht. Ziel einer rechtlichen Erfassung von KI muss sein, die mit KI-Anwendungen in bestimmten Sektoren (bspw. Medizinalprodukte und automatisierte Fahrzeuge) verbundenen Risiken zu minimieren, eine hinreichende Transparenz zu gewährleisten und die erforderlichen Mittel bereitzustellen, damit betroffene Personen gegen konkrete Nachteile (insb. Eingriffe in die Privatsphäre, Diskriminierung und Manipulation) vorgehen und den Ersatz allfälliger Schäden einfordern können.

[55] Im Gegensatz zum EU-Recht und in Übereinstimmung mit dem bewährten Ansatz des Schweizer Rechts sollten diese Herausforderungen allerdings nicht mit einem «KI-Gesetz» angegangen werden. Da KI-Anwendungen zahlreiche unterschiedliche Fragen in einer Vielzahl von Rechtsbereichen aufwerfen, wäre es wenig sinnvoll, alle diese Rechtsfragen in einem einzigen Erlass erfassen zu wollen. Stattdessen sollten in den betroffenen Rechtsbereichen punktuelle Anpassungen der bestehenden Normen vorgenommen werden, soweit diese erforderlich sind. Dabei werden einerseits gewisse sektorspezifische Regelungen zu überarbeiten sein (bspw. bei der Zulassung von Medizinalprodukten und automatisierten Fahrzeugen). Andererseits wird es bestimmte Anpassungen bei allgemein anwendbaren Normen durch eine Revision des Gesetzeswortlautes oder durch eine neue Auslegung und Anwendung brauchen. Möglicherweise bildet KI auch den Anlass, ein allgemeines Diskriminierungsverbot zu schaffen, sofern sich erweisen sollte, dass sich Diskriminierungen durch Private nicht hinreichend durch eine entsprechende Auslegung des allgemeinen Persönlichkeitsrechts (Art. 28 ZGB) erfassen lassen. In Anlehnung an den Verordnungsvorschlag der EU[123] könnte es zudem sinnvoll sein, eine allgemeine Pflicht einzuführen, betroffene Personen zu informieren, wenn sie mit einer KI interagieren.

[56] Angesichts der raschen technischen Entwicklung von KI sollten die Anpassungen im Schweizer Recht grundsätzlich technologieneutral formuliert werden. Auch wenn man sich nicht der Illusion hingeben darf, dass die technologieneutrale Formulierung von Rechtsnormen diese gegenüber Veränderungen ihrer Wirkungen aufgrund von technischen Entwicklungen abschirmen kann,[124] vermag doch nur dieser Ansatz zu verhindern, dass Rechtsnormen wegen technischer Entwicklungen überhaupt nicht mehr greifen, mithin schlicht ins Leere gehen. Strikt durchzuhalten ist der technologieneutrale Ansatz jedenfalls bei der Formulierung allgemein anwendbarer Normen, etwa bei einem allfälligen Diskriminierungsverbot. Sektorspezifische Regelungen (bspw. Zulassungsvorschriften bei Medizinalprodukten oder automatisierten Fahrzeugen) werden hingegen meist derart spezifisch ausfallen müssen, dass sie kaum noch technologieneutral formuliert werden können. Immerhin dürfte es hier möglich sein, die meisten Spezifika auf Verordnungsstufe zu regeln, wodurch die erforderlichen Anpassungen bei technischen Weiterentwicklungen vergleichsweise einfach vorgenommen werden können.

[57] Insgesamt scheinen eine umfassende Analyse der Herausforderungen in den jeweiligen Rechtsbereichen und die Erarbeitung eines differenzierten Sets an neuen Normen ebenso un-

123 Art. 52 EU-KI-Verordnungsvorschlag.

124 Siehe dazu Herbert Burkert/Peter Hettich/Florent Thouvenin, Eine kritische Geschichte des Informationsrechts: Erlebte, bevorstehende und versäumte Paradigmenwechsel, in: Lukas Gschwend/Peter Hettich/Markus Müller-Chen/Benjamin Schindler/Isabelle Wildhaber (Hrsg.), Recht im digitalen Zeitalter, FS zum Schweizerischen Juristentag 2015 in St.Gallen, Zürich/St.Gallen 2015, 49–71, 57 f.

umgänglich wie die Entwicklung von Vorgaben für die Auslegung und Anwendung bestehender Normen. Entscheidend wird dabei sein, dass die Arbeiten in den verschiedenen Rechtsbereichen koordiniert werden, weil sich nur so Widersprüche und Doppelspurigkeiten vermeiden lassen. Diese Arbeit kann nur eine fachlich breit aufgestellte Expertenkommission leisten. Mit der Einsetzung einer solchen Kommission sollte die Schweiz nicht länger zuwarten, wenn sie nicht Gefahr laufen will, in einen Zustand der Rechtsunsicherheit zu geraten, der weder der Forschung, Entwicklung und Anwendung von KI durch Bund, Kantone und Unternehmen noch der Bevölkerung nutzt.

[58] Die Arbeiten an einem Schweizer Rechtsrahmen für KI dürfen selbstredend nicht isoliert von den Rechtsetzungsbemühungen auf europäischer und internationaler Ebene erfolgen. Auch wenn das künftige EU-Recht nicht ins Schweizer Recht überführt werden sollte, ist dennoch auf die Anbindung an den europäischen Binnenmarkt zu achten. Zugleich sollte die Schweiz den Spielraum nutzen, um einen eigenständigen Ansatz zu entwickeln, der erlaubt, unter Wahrung der grundrechtlichen Vorgaben Innovation im Bereich der KI zu fördern – auf der Ebene der Forschung ebenso wie bei der Entwicklung und Anwendung.

Nadja Braun Binder, Prof. Dr., MBA, Professorin für Öffentliches Recht an der juristischen Fakultät der Universität Basel, Mitglied des Center for Information Technology, Society, and Law (ITSL) und der Digital Society Initiative (DSI).

Thomas Burri, Prof. Dr., LL.M., Rechtsanwalt, Professor für internationales Recht und Europäisches Recht an der Universität St. Gallen.

Melinda Florina Lohmann, Prof. Dr., Rechtsanwältin, Assistenzprofessorin für Wirtschaftsrecht, Schwerpunkt Informationsrecht und Direktorin der Forschungsstelle für Informationsrecht (FIR) an der Universität St. Gallen, Mitglied des Center for Information Technology, Society, and Law (ITSL) und der Digital Society Initiative (DSI).

Monika Simmler, Prof. Dr., Assistenzprofessorin für Strafrecht, Strafprozessrecht und Kriminologie und Co-Direktorin des Kompetenzzentrums für Strafrecht und Kriminologie (SK-HSG) an der Universität St. Gallen.

Florent Thouvenin, Prof. Dr., Rechtsanwalt, Inhaber des Lehrstuhls für Informations- und Kommunikationsrecht, Vorsitzender des Leitungsausschusses des Center for Information Technology, Society, and Law (ITSL) und Direktor der Digital Society Initiative (DSI) der Universität Zürich.

Kerstin Noëlle Vokinger, Prof. Dr. iur. et Dr. med., LL.M., Rechtsanwältin, Assistenzprofessorin und DSI-Professorin an der Universität Zürich, Faculty Associate am Berkman Klein Center for Internet and Society, Harvard University.