

CENTER FOR INFORMATION TECHNOLOGY SOCIETY AND LAW — ITSL

Volume 14

Rolf H. Weber

Künstliche Intelligenz und rechtliche Normenordnung

CENTER FOR INFORMATION TECHNOLOGY SOCIETY AND LAW — ITSL

Schriften aus dem ITSL, herausgegeben
von Florent Thouvenin und Rolf H. Weber

Volume 14

Rolf H. Weber

Künstliche Intelligenz und rechtliche Normenordnung

© 2026 – CC BY-NC-ND (Werk), CC BY-ND (Text)

Verlag: EIZ Publishing (eizpublishing.ch), Bellerivestrasse 49, 8008 Zürich,
eizpublishing@europa-institut.ch

Herausgeber: Center for Information Technology Society and Law ITSL, Florent Thouvenin (Hrsg.)

Autor: Rolf H. Weber

ISBN:

978-3-03994-070-7 (Print – Softcover)

978-3-03994-071-4 (ePub)

DOI: <https://doi.org/10.36862/70VK-8CSN>

Version: 1.00 – 20260219

Dieses Werk ist als gedrucktes Buch sowie als E-Book (open access) in verschiedenen Formaten verfügbar. Weitere Informationen finden Sie unter der URL: <https://eizpublishing.ch/publikationen/kuenstliche-intelligenz-und-rechtliche-normenordnung/>.

Vorwort

Die schnellen und umfassenden technologischen Entwicklungen der letzten Jahre, die durch den Einsatz von Künstlicher Intelligenz (KI) einen zusätzlichen Schub erhalten haben und die in immer stärkerem Masse zu einer digitalen Transformation führen, verursachen im Recht grosse Herausforderungen. KI-Systeme und KI-Prozesse verändern das persönliche Lebensumfeld, die wirtschaftlichen Aktivitäten im Rahmen neuer Geschäftsmodelle sowie das Verhältnis zwischen Individuen und Staat.

Der Autor hat sich seit bald sieben Jahren mit verschiedenen Aspekten der durch KI-Systeme und KI-Prozesse hervorgerufenen Herausforderungen für die rechtliche Rahmenordnung beschäftigt. Die Entwicklungen sind noch längst nicht am Ende. Dennoch erscheint es als gerechtfertigt, im Sinne eines «Zwischenhaltes» einzelne Publikationen des Autors in einem Sammelband zusammenzuführen und der interessierten Öffentlichkeit auch «Open Access» zur Verfügung zu stellen.

Angesichts der Vielzahl kürzlich erschienener Publikationen zur Vereinbarkeit des KI-Einsatzes mit den Bestimmungen einzelner sektorieller Rechtsgebiete umfassen die vorliegend abgedruckten Beiträge vornehmlich Analysen von sich grundsätzlich stellenden Fragen der Einbettung von KI-Regulierungen in eine humanistische und menschenzentrierte rechtliche Gesamtordnung. Der Sammelband nimmt deshalb insbesondere solche Überlegungen auf, die sich mit den normativen Grundprinzipien und den Menschenrechten in einer auf ethischen Verhaltensrichtlinien beruhenden Gesellschaftsordnung auseinandersetzen.

Der Autor dankt Sarah Montani und Simon Henseler für die Zustimmung zur erneuten Publikation des zuvor gemeinsam erarbeiteten Beitrages, verschiedenen (nahestehenden) Herausgebern bzw. Verlagen für die Genehmigung des Wiederabdrucks früher publizierter Aufsätze, Herrn Prof. Dr. Florent Thouvenin für die Aufnahme des Sammelbandes in die gemeinsame ITSL-Reihe sowie

Herrn Dr. Tobias Baumgartner, Leiter des Verlags EIZ Publishing und Vizedirektor am Europa Institut Zürich, und Frau Dr. Petra Bitterli, Verlag EIZ Publishing, für die umsichtige Begleitung der Entstehung dieses Buches.

Zürich, Januar 2026

Rolf H. Weber*

* Dr. iur. Rolf H. Weber, Professor für Wirtschaftsrecht an der Universität Zürich, daselbst Co-Leiter des interdisziplinären «Blockchain Center» und des «Center for Information Technology, Society, and Law», sowie Rechtsanwalt in Zürich (Bratschi AG).

Inhaltsverzeichnis

Vorwort	V
Einleitung und Regulierungsthematik	1
Rechtliche Entwicklungen im Kontext der Künstlichen Intelligenz	3
Rolf H. Weber	
Künstliche Intelligenz: Regulatorische Überlegungen zum «Wie» und «Was»	17
Rolf H. Weber	
Künstliche Intelligenz im Spiegel rechtlicher Grundprinzipien	35
Dürfen Maschinen über Menschen entscheiden? Eine rechtliche Auslegeordnung im Lichte neuer Technologien	37
Rolf H. Weber	
Automatisierte Entscheidungen: Perspektive Grundrechte	47
Rolf H. Weber	
Implementation of Normative Principles in the AI Environment	63
Rolf H. Weber	
Rule of Law and Ethics-based Equality in the Age of Artificial Intelligence	87
Rolf H. Weber	
Künstliche Intelligenz: Einzelthemen	101
Regulierung von Algorithmen in der EU und in der Schweiz, Überlegungen zu ausgewählten Regulierungsthemen	103
Rolf H. Weber / Simon Henseler	
Künstliche Intelligenz und Datenschutz	129
Rolf H. Weber	
Juristische Sprachmodelle zwischen Transparenz und Datensicherheit	149
Rolf H. Weber	

Human-Machine Cooperation through AI Agents?	163
Rolf H. Weber	
Herausforderungen beim Einsatz künstlicher Intelligenz in der Justiz	177
Rolf H. Weber	
Ausblick	199
Digitalisierung und der Kampf ums Recht	201
Rolf H. Weber	
Synthetische Persönlichkeit und der Schutz des Selbst – Wie künstliche Intelligenz die Grenzen des Menschseins verwischt	223
Sarah Montani / Rolf H. Weber	

Einleitung und Regulierungsthematik

Rechtliche Entwicklungen im Kontext der Künstlichen Intelligenz

Inhaltsübersicht

I.	Künstliche Intelligenz als neues Phänomen	3
1.	Wesen und Tragweite der Künstlichen Intelligenz	3
2.	Leitbilder im KI-Zeitalter	5
II.	Entwicklung der rechtlichen Rahmenordnung	7
1.	Globale Richtlinien	7
2.	Europarat	8
3.	Europäische Union	8
3.1.	Frühe Ethik-Leitlinien	8
3.2.	Konzeptioneller Ansatz der KI-Verordnung	9
3.3.	Auswirkungen auf die Schweiz	10
4.	Regulierungsansatz in der Schweiz	11
III.	Künstliche Intelligenz als Querschnittsmaterie	12
1.	Gewährleistung einer sachgerechten Rahmenordnung	12
2.	Wichtige Einzelthemen	13

I. Künstliche Intelligenz als neues Phänomen

1. Wesen und Tragweite der Künstlichen Intelligenz

Die schnell voranschreitenden und inhaltlich umfassenden technologischen Entwicklungen der letzten Jahre, die – nun verstärkt durch den Einsatz von Künstlicher Intelligenz (KI) – zu einer digitalen Transformation geführt haben, verursachen im Recht grosse Herausforderungen. KI-Systeme und KI-Prozesse verändern das persönliche Lebensumfeld, die wirtschaftlichen Aktivitäten im Rahmen neuer Geschäftsmodelle sowie das Verhältnis zwischen Individuen und staatlichen Stellen.¹

¹ Angesichts der Tatsache, dass KI-Systeme und KI-Prozesse vielfältige Disziplinen betreffen, sind im vorliegenden Sammelband bewusst auch mehrere Beiträge zu ethischen, gesellschaftlichen und ökonomischen Fragestellungen abgedruckt.

Die Zahl der KI-Definitionen sowie die KI-Leitbilder sind kaum mehr überblickbar.² Die von der Schweiz massgeblich mitentwickelte KI-Konvention des Europarats von 2024 verwendet folgende Umschreibung: Ein KI-System ist «ein maschinengestütztes System, das aufgrund expliziter oder impliziter Zielvorgaben ableitet, wie es Ergebnisse wie Vorhersagen, Inhalt, Empfehlungen und Entscheidungen erzeugen kann, die sich auf die physische und virtuelle Umgebung auswirken können» (Art. 2).³ Ähnlich ist die Umschreibung auch in Art. 3 der KI-Verordnung der Europäischen Union.⁴

Das «Design» von verantwortungsvoller KI muss die Interessen verschiedenster Akteure berücksichtigen und gestützt auf eine entsprechende Analyse diese Interessen sachgerecht gegeneinander abwägen. Das erwähnte, in der Technologiebranche oft verwendete Wort «Design» erinnert an das Buch «Der kleine Prinz» von Antoine de Saint-Exupéry aus dem Jahre 1943.⁵ Der kleine Prinz bittet den Erzähler nach dessen Notlandung mit dem Flugzeug in der Sahara: «Zeichne mir ein Schaf:». ⁶ Zuerst zeichnet der Erzähler eine Riesenschlange (mit verschlungenem Elefanten), die der kleine Prinz beide erkennt, doch wegen des fehlenden Schafes unzufrieden bleibt. Hernach erstellte Zeichnungen befriedigen den kleinen Prinzen ebenfalls nicht, weil die Schafe zu klein, zu wild oder zu alt sind. Zur letzten Zeichnung, die eine Kiste abbildet, bemerkt der Erzähler: «Das Schaf, das du willst, steckt drin»; diese Zeichnung stellt den kleinen Prinzen zufrieden. Ähnlich wird bei Verwendung von KI-Modellen ein sog. «Prompt» formuliert, der im KI-System das gesuchte, vorerst nicht sichtbare «Resultat» generiert.

Die Verwendung juristischer Sprachmodelle ist in den letzten drei Jahren exponentiell angestiegen; in weiteren Bereichen nimmt der Einsatz von KI-Systemen ebenfalls stark zu.⁷ Die durch das Internet eingeleitete Digitalisierung erhält damit einen zusätzlichen massiven Schub; die rechtliche Rahmenord-

² Bereits im Jahre 2020 hat das AI Ethics Guidelines Global Inventory von AlgorithmWatch über 160 Dokumente mit KI-Leitlinien erwähnt; vgl. <https://inventory.algorithmwatch.org/> (besucht am 12.1.2026).

³ Konvention des Europarates zu Künstlicher Intelligenz (inkl. Erläuterungsbericht), SEV 2025 vom 5.9.2024.

⁴ Verordnung (EU) 2024/1689 vom 13.6.2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung verschiedener Richtlinien und Verordnungen, ABl. 2024 vom 12.7.2024.

⁵ So auch PASCAL PICHONNAZ, Avant-propos, ZSR 2025 I 273.

⁶ Vgl. ANTOINE DE SAINT-EXUPÉRY, Der kleine Prinz, 1943, Kap. 2.

⁷ RICHARD SUSSKIND, How To Think About AI. A Guide For The Perplexed, Oxford 2025, 13.

nung muss deshalb verhindern, dass die Situation einer «technological myopia»⁸ (Kurzsichtigkeit) eintritt.

2. Leitbilder im KI-Zeitalter

KI-Systeme und KI-Prozesse stellen nicht nur Herausforderungen für die Rechtsordnung dar, sondern sie betreffen die Gesellschaft insgesamt. Vorhandene humanistische und sozio-ökonomische Leitbilder benötigen eine neue Orientierung mit Blick auf zukunftsfähige Lösungen. Aus gesellschaftlicher Sicht erscheint es als unerwünscht, wenn die Technologien allein die Rahmenbedingungen des Zusammenlebens determinieren.

Zwar sind schon früher die Vertreter des sog. transhumanistischen Menschenbildes davon ausgegangen, Technologien seien zur Verbesserung und Optimierung der Individuen heranzuziehen, und zwar mit unterschiedlichen Theorieansätzen:⁹ (i) Der technologische Posthumanismus geht z.B. konkret davon aus, dass der Mensch von einer artifiziellen Superintelligenz abgelöst werde (Übertragung des menschlichen Bewusstseins auf ein anderes Medium).¹⁰ (ii) Der kritische Posthumanismus versteht den Menschen als Teil eines komplexen Netzwerks, das in der Lage ist, den Anthropozentrismus zu überwinden.¹¹ Ungeachtet dieser Gedankenansätze zeigen die neuesten Diskussionen, dass die Technologie-Optimisten sich im Kontext der KI-Diskussionen nicht durchsetzen konnten.¹²

Auch im KI-Zeitalter muss das Individuum in der Lage sein, selbstbestimmt zu entscheiden, das eigene Potenzial zu entfalten und seine Tätigkeiten autonom zu gestalten, d.h. im Vordergrund steht die gezielte Erweiterung des menschlichen Handlungsraums und der menschlichen Fähigkeiten durch KI.¹³ Schon Aristoteles ist vom Menschen als «Zoon Politikon» ausgegangen, Immanuel

⁸ SUSSKIND (Fn. 7), 30 f.

⁹ JANINA LOH, Trans- und Posthumanismus zur Einführung, 4. Aufl. Hamburg 2023, 32 und 58.

¹⁰ STEFAN L. SORGNER, Beyond Humanism: Reflections on Trans- and Posthumanism, 21 Journal of Evolution and Technology 2010, 1 ff.; RAY KURZWEIL, The Singularity is Near: When Humans Transcend Biology, East Rutherford 2005, 125 f.

¹¹ Für einen detaillierten Überblick zu diesen Theorieansätzen vgl. VANESSA RÜEGGER/KATHARINA SCHÖBI, Künstliche Intelligenz und Menschenrechte – Menschenwürde als Leitidee im Umgang mit KI, ZSR 2025 I 209, 220 ff.

¹² Eingehender RÜEGGER/SCHÖBI (Fn. 11), 226 ff.

¹³ Vgl. MAX SENEGES/MONIKA ILVES/DETLEF GÜRTLER/MANUEL DOLDERER, KI zur Potentialentfaltung. Ein sozio-ökonomisches Leitbild für ein humanistisches KI-Zeitalter, Berlin 2025, 1, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5501838 (besucht am 12.1.2026).

Kant hat mit seinem Autonomiekonzept den Grundstein für die menschliche Selbstbestimmung gestützt auf die moralische Urteilkraft gelegt und Hannah Arendt sieht das wahre Menschsein in der Öffentlichkeit (d.h. im Handeln und im Diskurs) entstehen.¹⁴

Abgesehen von der Erarbeitung einer sachgerechten normativen Rahmenordnung ist die Herausbildung ethischer Standards von grosser Bedeutung.¹⁵ Schon seit Jahrhunderten haben Philosophen verschiedenster Herkunft (mit Aristoteles beginnend und später z.B. Locke und Rousseau, der das Konzept der Transformation vom natürlichen zum sozialen Menschen entwickelte, einschliessend) ethische Verhaltensrichtlinien entwickelt; in neuerer Zeit sind ihnen internationale Organisationen und Branchenverbände gefolgt. Als wichtigste Attribute fallen in Betracht:¹⁶

- *Gerechtigkeit und Gleichheit*: Jeder Mensch verfügt über eine unentziehbare Würde und über gleiche Rechte (z.B. mit Bezug auf Informationen, auf Zugang zu Leistungen und auf ein faires Verfahren).
- *Freiheit*: Die Ethik verlangt die Zuerkennung verschiedenster Menschenrechte an die Individuen.
- *Sorge und gegenseitiges Verständnis («compassion»)*: Die Empathie für die Mitmenschen führt zu Solidarität und gegenseitiger Unterstützung.
- *Partizipation*: Das Recht und die Fähigkeit zur Teilnahme am sozialen Leben und in Entscheidungsverfahren sind wichtige Werte.
- *Nachhaltigkeit und Verantwortung*: Nachhaltigkeit ist bedeutungsvoll für eine lebenswerte Umwelt und die Verantwortung für eigene Tätigkeiten erweist sich als eine wichtige Verhaltensmaxime.

Die verschiedenen ethischen Werte sind miteinander verwoben und betreffen nicht nur die persönlichen Verhaltensweisen, sondern auch die beruflichen Aktivitäten. Bei der Entwicklung ethischer Standards ist ein Gleichgewicht zwischen der technischen Höchstleistung und der Einhaltung einer men-

¹⁴ Vgl. SENGES/ILVES/GÜRTLER/DOLDERER (Fn. 13), 2 f.

¹⁵ ROLF H. WEBER, Ethics as Pillar of Internet Governance, 23 Annual Review of Law and Ethics 2015, 95, 100.

¹⁶ Die nachfolgende Liste basiert gekürzt auf WEBER (Fn. 15), 100 f.; vgl. auch ROLF H. WEBER, Rule of Law and Ethics-based Equality in the Age of Artificial Intelligence, in: Erich Schweighofer et al. (Hrsg), Der Mensch im Zentrum – KI, Ethik & Recht, Bern 2025, 23, 28 (vgl. [Text 6, S. 87 ff.](#)).

schenzentrierten Perspektive anzustreben.¹⁷ Die algorithmischen Funktionen müssen im Sinne von «best practices» so ausgestaltet sein, dass sich negative Auswirkungen auf die verfassungsrechtlich geschützten Menschenrechte und auf ethische Verhaltensweisen vermeiden lassen.¹⁸

Die Realisierung eines humanistischen Leitbildes im KI-Zeitalter setzt eine interdisziplinäre Zusammenarbeit verschiedener Wissensbereiche (z.B. Jurisprudenz, Ökonomie, Ethik, Politik, Soziologie, usw.) im Sinne der vorerwähnten Überlegungen voraus, selbst wenn die hernach im Sammelband abgedruckten Beiträge schwergewichtig die rechtlichen Themenstellungen in den Vordergrund der Diskussion rücken.

II. Entwicklung der rechtlichen Rahmenordnung

In den letzten Jahren sind auf globaler, regionaler und nationaler Ebene viele KI-Regulierungen entstanden, die kurz zu erläutern sind.

1. Globale Richtlinien

Auf der globalen Ebene fehlen derzeit verbindliche Staatsverträge, die KI-Regulierungen enthalten. Hingegen haben verschiedene internationale Organisationen Leitlinien erarbeitet. Die OECD ist bereits früh (2019) mit zwischenzeitlich revidierten (2024) KI-Leitlinien, die laufend weiterentwickelt werden, an die Öffentlichkeit getreten; der Leitfaden baut auf einem breit abgestützten und etablierten Konzept der Sorgfaltsprüfung zu Risiken für Mensch, Umwelt und Planet auf und erleichtert damit dessen Rezeption in der Praxis.¹⁹

Die UNESCO hat 2021 detaillierte Ethik-Empfehlungen zur Künstlichen Intelligenz verabschiedet; ihr Kernbereich betrifft die Menschenwürde in ihrem

¹⁷ Zutreffend hat das Institute of Electrical and Electronics Engineers (IEEE), heute der wichtigste Berufsverband im Bereich der Informationstechnologien, bereits im Jahre 2017 den (überarbeiteten) Entwurf für ein Dokument mit dem Titel «Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Artificial Intelligence and Autonomous Systems» veröffentlicht und auf die Grundsätze der «Auditability, Accessibility, Meaningfulness and Readability» hingewiesen: vgl. IEEE, Ethically Aligned Design, 68, https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (besucht am 12.1.2026).

¹⁸ WEBER (Fn. 16), 28.

¹⁹ Im Einzelnen dazu CHRISTINE KAUFMANN, Neue OECD-Instrumente zu künstlicher Intelligenz – Wege zu vertrauenswürdiger künstlicher Intelligenz, Jusletter IT, 4. Juli 2024, Rz. 1 ff., 30.

normativen Sinngehalt.²⁰ Weitere Organisationen innerhalb der Vereinten Nationen sind mit ethisch geprägten KI-Leitlinien gefolgt.²¹

2. Europarat

Unter massgeblicher Mitwirkung (und Leitung) der Schweiz hat der Europarat am 17. Mai 2024 das Rahmenübereinkommen über Künstlichen Intelligenz und Menschenrechte, Demokratie und Rechtsstaatlichkeit verabschiedet.²² Den Funktionen des Europarats entsprechend, ist der Anwendungsbereich der KI-Konvention breit (Menschenrechte, Demokratie) und nicht vornehmlich auf die Geschäftswelt konzentriert.

Nach allgemeinen Bestimmungen (Art. 1–3) und allgemeinen Verpflichtungen (Art. 4–5) sieht die KI-Konvention insbesondere allgemeine Prinzipien mit Bezug auf Tätigkeiten innerhalb des Lebenszyklus von KI-Systemen vor (Art. 6–13). Risiken und nachteilige Auswirkungen von KI-Systemen sind zu identifizieren, bewerten, vermeiden und mindern (Art. 16). Gestützt auf das «Impact Assessment»-Werkzeug des Europarates müssen die Mitgliedstaaten die konkreten Massnahmen umsetzen, was signifikante Investitionen in die Sicherheit und Transparenz von KI-Systemen zur Folge haben wird.²³

3. Europäische Union

3.1. Frühe Ethik-Leitlinien

Im Jahre 2019 hat die EU-Kommission erste Ethik-Leitlinien, ausgearbeitet von einer hochrangigen Expertengruppe, publiziert; diese Leitlinien umschreiben das Fundament einer vertrauenswürdigen KI, d.h. einer rechtmässigen, ethischen und robusten KI.²⁴ Der Mensch soll, auch im KI-Zeitalter, eine «einzigartige und unveräusserliche moralische Vorrangstellung einnehmen»; KI-Systeme sind so auszugestalten, dass die Menschen als moralische Subjekte anerkannt sowie ihre Freiheit und Autonomie gefördert werden; die körper-

²⁰ UNESCO KI-Erklärung, November 2021.

²¹ Im Einzelnen dazu RÜEGGER/SCHÖBI (Fn. 11), 215 m.w.V.

²² Vgl. vorne Fn. 3.

²³ DAVID MARTI, Die KI-Konvention des Europarats: Ursprung, Inhalt, Ausblick, Jusletter IT, 4. Juli 2024, Rz. 37 ff., 40 und 41.

²⁴ Hochrangige Expertenkommission für künstliche Intelligenz, Ethik-Leitlinien für eine vertrauenswürdige KI, Brüssel 2019.

liche und geistige Unversehrtheit des Menschen sowie seine persönliche und kulturelle Identität sind zu schützen.²⁵

Im Jahre 2023 haben die EU-Organen als eine Fortführung der Ethik-Leitlinien die Europäische Erklärung zu den digitalen Rechten und Grundsätzen für die digitale Dekade angenommen; im Mittelpunkt dieser Erklärung stehen der Schutz der Menschenwürde und die Gewährleistung der Menschenrechte.²⁶

Parallel dazu ist in intensiven Diskussionen die detaillierte KI-Verordnung ausgearbeitet worden,²⁷ die im Rahmen eines sog. Omnibus-Paketes nun aber eine gewisse Vereinfachung erfahren soll.

3.2. Konzeptioneller Ansatz der KI-Verordnung

Die Verordnung (EU) 2024/1689 vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz und zur Änderung verschiedener Richtlinien und Verordnungen, die nach einem vielschichtigen Gesetzgebungsprozess am 1. August 2024 in Kraft getreten ist, verfolgt das erklärte Ziel, einen einheitlichen Rechtsrahmen für die Entwicklung, das Inverkehrbringen, die Inbetriebnahme und die Verwendung von KI-Systemen innerhalb der EU zu schaffen. Das Konzept der KI-Verordnung orientiert sich am Produktsicherheitsrecht und ist als «New Legislative Framework» (NLF) ausgestaltet.²⁸ Im Vordergrund stehen deshalb die Sicherheits-Anforderungen für KI-Systeme, die aber im Laufe der Ausarbeitung der KI-Verordnung um Regelungen zu KI-Modellen mit allgemeinem Verwendungszweck (General Purpose AI, GPAI) ergänzt worden sind.²⁹ Konzeptionell bemüht sich die KI-Verordnung, einen sog. risikobasierten Ansatz zu verwirklichen, anlehnend an die in den KI-Leitsätzen der OECD verankerte Differenzierung nach Risikostufen.³⁰

²⁵ Hochrangige Expertenkommission (Fn. 24), Rz. 38, 41 und 153.

²⁶ Europäische Erklärung vom 23.1.2023, 2023/C 23/01.

²⁷ Vgl. vorne Fn. 4.

²⁸ Vgl. STEPHANIE VOLZ, Die KI-Verordnung und die Schweiz, sic! 2025, 656, 657 m.w.V.

²⁹ Die Begründung für diese relativ spät dazugekommenen Ergänzungen liegt darin, dass zu Beginn der Arbeiten an der KI-Verordnung (2020) die GPAI noch nicht verbreitet Anwendungen gefunden haben, d.h. der Regulierungsbedarf erst in einem späteren Zeitpunkt eingetreten ist.

³⁰ Zum risikobasierten Ansatz sowie zum Risikomanagement-System eingehender vgl. MARTINA ARIOLI, Risikomanagement nach der EU-Verordnung über Künstliche Intelligenz, Jusletter IT, 4 Juli 2024, Rz. 11 ff. und 15 ff.; JOANNA KULESZA, Europe's Regulatory Sovereignty in the Age of Artificial Intelligence: A Human-Centric Response to the US-China Technological Rivalry, 78/4 Sprawy Międzynarodowe 2025, 167, 177 ff.

Nach ihrem Wortlaut soll die KI-Verordnung technologieneutral sein. Effektiv sind die detaillierten Regelungen aber auf spezifische KI-Systeme ausgerichtet, d.h. die EU hat darauf verzichtet, allgemein unzulässige Verhaltensweisen zu umschreiben und diese unabhängig von der eingesetzten Technologie für anwendbar zu erklären.³¹ Ob dieser Ansatz auch in einigen Jahren, falls sich die technologischen Rahmenbedingungen stark ändern sollten, noch trägt, bleibt zu sehen.

Die KI-Verordnung regelt nicht alle Rechtsfragen, die sich ggf. im Kontext von KI-Systemen stellen. Spezifische Rechtsgebiete wie insbesondere das Haftungsrecht sowie das Immaterialgüter- bzw. Urheberrecht werden besonderen Rechtsakten vorbehalten.³² Die mit Bezug auf verstärkte Regulierungen bei der EU eingetretene Trendwende hat vorerst indessen zu einem Aufschub der weiteren Arbeiten an angepassten Regeln für das Haftungsrecht geführt.

3.3. Auswirkungen auf die Schweiz

Die KI-Verordnung ist für die Schweiz nicht unbeachtlich; deren Art. 2 Abs. 1 statuiert das sog. Auswirkungsprinzip, das ein Angebot von Produkten oder Dienstleistungen aus der Schweiz an Adressaten in EU-Mitgliedstaaten (d.h. auch in den umliegenden Ländern) nur in Übereinstimmung mit deren Vorschriften zulässt.³³ Grenzüberschreitend tätige Unternehmen müssen somit für entsprechend gelieferte Produkte und getätigte Dienstleistungen die Anforderungen der KI-Verordnung erfüllen, sofern nicht eine spezifische Ausnahme vorliegt (Art. 2 Abs. 2 i.V. m. Anhang IB).

Der Vollständigkeit halber ist noch darauf hinzuweisen, dass im grenzüberschreitenden Geschäftsverkehr zwischen Schweizer Unternehmen und Vertragspartnern in einem EU-Land weitere Rechtserlasse zur Anwendung kommen können, etwa das Abkommen über technische Handelshemmnisse (MRA), das derzeit zwar von der EU nicht aktualisiert wird, sowie die Datenschutz-Grundverordnung von 2016, die in Art. 3 Abs. 2 ebenfalls das Auswirkungsprinzip verankert, doch hat die EU durch den Angemessenheitsbeschluss von 2024 die Gleichwertigkeit des Schweizer Datenschutzrechts anerkannt.³⁴

³¹ CHRISTIANE WENDEHORST, in: Mario Martini/Christiane Wendehorst (Hrsg.), KI-VO: Verordnung über Künstliche Intelligenz, Kommentar, München 2024, Art. 1 N 5; VOLZ (Fn. 28), 657.

³² GEORG BORGES, Die Europäische KI-Verordnung (AI Act), CR 2024, 497, 506.

³³ BORGES (Fn. 32), 498.

³⁴ Im Einzelnen zu MRA und DSGVO vgl. VOLZ (Fn. 28), 669.

4. Regulierungsansatz in der Schweiz

Die Schweiz kennt kein allgemeines KI-Gesetz und dürfte die Ausarbeitung eines solchen Gesetzes in nächster Zeit auch nicht planen. Vielmehr hat der Bundesrat im Februar 2025 gestützt auf einen sehr umfangreichen Bericht des Bundesamtes für Kommunikation (BAKOM) und weiterer Bundesämter (z.B. Bundesamt für Justiz) beschlossen, einerseits die Ratifikation der vorerwähnten KI-Rahmenkonvention des Europarates voranzutreiben und andererseits spezifische Anpassungen des Schweizer Rechts in Betracht zu ziehen.³⁵

Der Bundesrat vertritt die (zutreffende) Auffassung, die punktuelle und sachbezogene Anpassung bzw. Ergänzung des geltenden Schweizer Rechts sei angemessener als die Ausarbeitung einer umfassenden KI-Gesetzgebung; die Bundesverwaltung hat bis Ende 2026 entsprechende Vorschläge auszuarbeiten. Betroffen sind etwa das Haftungsrecht (teilweise auch das Vertragsrecht), das Arbeitsrecht, das Urheberrecht (bzw. allgemein das Immaterialgüterrecht) sowie ebenfalls verschiedene sektorielle Gebiete wie z.B. das Datenschutz- und das Lauterkeitsrecht.³⁶

Angesichts der in der KI-Konvention des Europarats angesprochenen gesellschaftlichen Themen der Demokratie und der Rechtsstaatlichkeit ist nicht nur der Bund gefordert, Umsetzungsnormen zu erlassen, sondern deren Bestimmungen betreffen nach der Ratifikation auch die Kantone; insoweit sind noch einige Rechtsunsicherheiten auszuräumen.³⁷ Mit entsprechenden Klärungen ist Ende 2026 (Vorlage des Berichts der Bundesverwaltung) zu rechnen.

Der Bundesrat hat sich – wie erwähnt mit zutreffenden Überlegungen – dafür entschieden, auf die Ausarbeitung einer umfassenden KI-Gesetzgebung zu verzichten und vielmehr das heutige Recht spezifisch anzupassen, soweit die KI-Systeme und KI-Prozesse einer konkreten rechtlichen Rahmenordnung bedürfen. Dieses Konzept entspricht nicht nur der erfolgreichen Vorgehensweise beim Erlass des DLT-Gesetzes von 2020, das eine sinnvolle Regelungsdichte für den Handlungsbedarf im Blockchain-Bereich eingeführt hat, sondern auch den detaillierten Ausführungen des Autors in einem Aufsatz, der anfangs 2022

³⁵ Vgl. Medienmitteilung des Bundesrates vom 12.2.2025, <https://www.news.admin.ch/de/nsb?id=104110> mit Links zu den weiteren Dokumenten (besucht am 12.1.2026).

³⁶ Vgl. dazu ROLF H. WEBER/ADRIAN BIERI, KI-Regulierung in der Schweiz und ihre Relevanz für Kanzleien, in: Karin Mülchi/Stéphanie Fuchs (Hrsg.), Recht^{KI}, Anwendung künstlicher Intelligenz in der Rechts- und Steuerrechts-Praxis, Zürich 2026, 65, 67 ff.

³⁷ Dazu NADJA BRAUN BINDER/NINA LAUKENMANN/ANNA KUHN, Rechtliche Umsetzung der KI-Konvention in den Kantonen, ZSR 2025 I 313 ff.

publiziert worden ist; angesichts der Tatsache, dass die Überlegungen dieses Beitrages weiterhin als bedenkenswert erscheinen, wird der Aufsatz gleich nach der vorliegenden Einführung – und vor der Diskussion allgemeiner und spezifischer Themenstellungen zur KI – nachfolgend abgedruckt.³⁸

III. Künstliche Intelligenz als Querschnittsmaterie

1. Gewährleistung einer sachgerechten Rahmenordnung

Wie einleitend ausgeführt, sind rechtliche Regelungen in eine humanistische und menschenzentrierte Gesamtordnung einzubetten. KI-Regulierungen müssen deshalb zwingend die Grundwerte der Gesellschaft berücksichtigen. Gestützt auf diese Einschätzung enthält der Sammelband im ersten Hauptteil vier Beiträge, die sich mit den durch KI-Systeme für die Grundprinzipien der Rechtsordnung verursachten Herausforderungen beschäftigen.

Eine wesentliche Ausgangsfrage betrifft das Thema, ob Maschinen über Menschen entscheiden dürfen, d.h. inwieweit Technologien an die Stelle des selbstbestimmten Menschen zu treten vermögen.³⁹ Ebenfalls in einer schon sehr frühen Phase hat der Autor die Perspektive der Grundrechte angesichts der immer verbreiteter angewendeten KI-Systeme analysiert; der Beitrag zeigt auf, dass Technologien durchaus in Menschen- und Grundrechte einzugreifen vermögen und dass die rechtliche Rahmenordnung gewisse Schrankensetzungen im Interesse der selbstbestimmten Individuen vornehmen muss.⁴⁰

Normative Verhaltens-Leitlinien spielen in allen Segmenten der Gesellschaft eine Rolle; gewisse Grundprinzipien wie z.B. die Transparenz, die Rechenschaft/Verantwortung, der Schutz von Treu und Glauben sowie ein ethisches Vorgehen sind bereits in anderen Kontexten (z.B. Internet Governance) eingehend thematisiert worden, doch erweist sich eine spezielle Analyse mit Bezug auf KI-Systeme dennoch als sachgerecht. Der Autor hat sich in drei englisch verfassten Beiträgen mit diesen Fragestellungen beschäftigt; die Essenz der drei Aufsätze wird in diesem Sammelband in einem neu formulierten und

³⁸ Vgl. [Text 2, S. 17 ff.](#)

³⁹ Dazu [Text 3, S. 37 ff.](#)

⁴⁰ Vgl. [Text 4, S. 47 ff.](#); zur Meinungsäußerungsfreiheit im Besonderen vgl. nun Council of Europe, Guidance Note on the Implications of Generative Artificial Intelligence for Freedom of Expression, 3 December 2025, CDMSI(2025)15rev.

angepassten Text unter dem Titel «Implementation of Normative Principles in the AI Environment» zusammengefasst.⁴¹

Nur im Rahmen von auf andere Themen ausgerichteten Beiträgen kommt im Sammelband ein weiteres Prinzip, nämlich das Gebot der Nichtdiskriminierung, zur Sprache; die Pflicht zur Gleichbehandlung ist aber von grosser Bedeutung und mag im KI-Kontext ein (ggf. der einzige) Anlass sein, über die Einführung eines allgemeinen Gesetzes zur Nichtdiskriminierung (bzw. zur Gleichbehandlung) auf einer horizontalen Ebene nachzudenken.⁴²

Einen zentralen Stellenwert muss die Vereinbarkeit der Anwendung von KI-Systemen und KI-Modellen mit allgemeinen ethischen Leitlinien einnehmen; über die vorerwähnten allgemeinen Grundsätze hinaus⁴³ hat der Autor kürzlich das Spannungsfeld zwischen der «Rule of Law» und der ethik-bezogenen Gleichbehandlung einerseits sowie den technologischen Herausforderungen durch KI-Systeme andererseits analysiert und Vorschläge zu einem möglichen sachgerechten Abbau der entsprechenden Spannungen formuliert.⁴⁴

2. Wichtige Einzelthemen

Gewisse Rechtsbereiche sind durch die KI-Entwicklungen stärker betroffen als andere Teile der Rechtsordnung; aus diesem Grunde bedarf es für die betroffenen Rechtsgebiete besonderer legislatorischer Überlegungen. Zwischenzeitlich gibt es bereits eine fast unüberblickbare Zahl von Erläuterungen und Analysen in der Schweizer Literatur. Selbstredend hat sich der Autor nur mit einzelnen wichtigen Themen beschäftigen können.

Das zweite Hauptkapitel beginnt mit einem Aufsatz, der sich bereits in einem sehr frühen Stadium des Aufkommens von KI-Systemen mit der Regulierung von Algorithmen in der EU und in der Schweiz beschäftigt, d.h. allgemeine Überlegungen zu ausgewählten Regulierungsthemen analysiert.⁴⁵ Das wohl am intensivsten diskutierte Thema im KI-Kontext betrifft den Datenschutz, der sich in der digitalen Welt und der Anwendung von KI-Systemen im Beson-

⁴¹ Vgl. nachfolgend [Text 5, S. 63 ff.](#)

⁴² Vgl. [Text 5, S. 63](#), 79 ff. und [Text 7, S. 103](#), 109 f., 122 ff.; eingehender zum Diskriminierungsrisiko vgl. FLORENT THOUVENIN/STEPHANIE VOLZ/SORAYA WEINER/CHRISTOPH HEITZ, Diskriminierung beim Einsatz von Künstlicher Intelligenz (KI). Technische Grundlagen für Rechtsanwendung und Rechtsentwicklung, Jusletter IT, 4. Juli 2024, Rz. 4 ff.

⁴³ Vgl. vorne Kap. [I.2.](#)

⁴⁴ Vgl. nachfolgend [Text 6, S. 87 ff.](#)

⁴⁵ Vgl. nachfolgend [Text 7, S. 103 ff.](#)

deren ganz spezifischen Herausforderungen ausgesetzt sieht; einen Überblick mit weiteren Literaturhinweisen gibt der Beitrag «Künstliche Intelligenz und Datenschutz».⁴⁶

Im Zusammenhang mit dem Datenschutz spielt auch die Datensicherheit bei KI-Systemen eine grosse Rolle; juristische Sprachmodelle müssen deshalb eine Interessenabwägung zwischen (i) der Schaffung angemessener Transparenz und sachgerechter Zugangsrechte einerseits sowie (ii) der Einhaltung der Datensicherheitsvorschriften andererseits erreichen.⁴⁷

Neuere Entwicklungen haben gezeigt, dass sich KI-Systeme auch als «Agenten» in Vertragsbeziehungen einsetzen lassen. Solche technologische Massnahmen rufen nicht nur die Herausforderungen des aus dem Gesellschaftsrecht seit Jahrzehnten bekannten «Principal-Agent»-Problems hervor, sondern stellen überdies wichtige verbraucherrechtliche Fragen; die in den letzten drei Jahrzehnten im Verbraucherrecht neu eingeführten Schutzvorkehrungen zugunsten der schwächeren Vertragspartei dürfen nicht durch KI-Systeme ausgehebelt werden.⁴⁸

Schliesslich ergeben sich vielfältige Problemstellungen beim Einsatz von KI-Systemen im öffentlichen Sektor. Die Verwaltung greift schon intensiv auf die Verwendung von Algorithmen zurück.⁴⁹ Der Autor hat sich mit den Herausforderungen beim Einsatz Künstlicher Intelligenz in der Justiz beschäftigt.⁵⁰

Der Sammelband wird abgerundet mit zwei Beiträgen, die sich möglichen künftigen Entwicklungen und den damit verbundenen Herausforderungen widmen. Die Digitalisierung lässt sich zum Teil auch als «Kampf ums Recht» sehen, und zwar insoweit, dass die Datafizierung als technologischer Trend (mit Erscheinungen wie Big Data, Reg Tech, Legal Tech, usw.) viele traditionelle Rechtsverständnisse grundsätzlich hinterfragt und berufliche Tätigkeiten in

⁴⁶ Vgl. nachfolgend [Text 8, S. 129 ff.](#); aus der seither erschienenen Literatur vgl. ROLF H. WEBER, Datenschutz-Compliance beim Einsatz von KI-Systemen, in: Sonja Janisch/ Sebastian Krempelmeier/Peter Mader/Elisabeth Staudegger/ClemensThiele (Hrsg.), Am Glasfaserpuls, Festschrift für Dietmar Jähnel, Wien 2025, 749 ff.; FLORENT THOUVENIN, Datenschutzrechtliche Herausforderungen bei der Entwicklung von KI-Modellen, ZSR 2025 I 343 ff.; WEBER/BIERI (Fn. 36), 70 ff.

⁴⁷ Vgl. nachfolgend [Text 9, S. 149 ff.](#)

⁴⁸ Dazu nachfolgend [Text 10, S. 163 ff.](#)

⁴⁹ Eingehender zu den entsprechenden Fragestellungen vgl. die verschiedenen Beiträge im Sammelband von NADJA BRAUN BINDER/FLORENT THOUVENIN (Hrsg.), Ein Rechtsrahmen für KI in der Schweiz: Perspektive öffentliches Recht, Zürich 2025.

⁵⁰ Vgl. dazu nachfolgend [Text 11, S. 177 ff.](#)

neue «Umwelten» stellt. Disruptive Effekte im vertrauten «Rechtsgebäude» lassen sich nicht völlig ausschliessen, weshalb es wichtig ist, sich sinnvolle Überlebensstrategien des Rechts zu überlegen, solange eine Mitgestaltung der Rahmenbedingungen möglich ist.⁵¹

Der Sammelband schliesst mit einem Ausblick, der relativ provokativ den Herausforderungen nachgeht, die eintreten können, wenn nicht nur die textliche Information, sondern vermehrt auch Bilder (z.B. automatisierte Schaffung von Videos [etwa SORA von OpenAI]) die selbstbestimmten Individuen beeinträchtigen. Menschen, die in solchen Videos optisch nachvollziehbar vorkommen (Gesichter, Körper, Bewegungen), sind nicht mehr nur reale Personen, sondern ein mögliches Ergebnis statistischer Synthesevorgänge (als Muster-Neukombinationen von Daten). Der klassische Persönlichkeitsschutz scheitert dann am Fehlen eines realen Individuums; die generative KI vermag zudem «synthetische Doppelgänger» zu produzieren, was die Grenzen zwischen Realität und Fiktion weiter verwischt.⁵² Im Schöpfungsprozess zeichnet sich damit die Kreativität nicht mehr durch das Entwerfen eines Werks aus, sondern durch die Anweisung, wie das Werk zu entwickeln ist (sog. «Prompt»-Kompetenz), d.h. das Handwerkliche verliert an Bedeutung, während das Konzeptionelle in den Vordergrund tritt.

Die verschiedenen auf die Verwirklichung einer menschenzentrierten Ausgestaltung der rechtlichen KI-Rahmenordnung sowie auf die Lösung konkreter Einzelprobleme ausgerichteten Beiträge sollen die weitere Diskussion bereichern und Anhaltspunkte für künftige Forschungsfelder vermitteln.

⁵¹ Vgl. dazu nachfolgend [Text 12, S. 201 ff.](#)

⁵² Vgl. dazu nachfolgend [Text 13, S. 223 ff.](#)

Künstliche Intelligenz: Regulatorische Überlegungen zum «Wie» und «Was»

Erstveröffentlichung in: EuZ, Ausgabe 1/2022, 29–49

Neue technologische Entwicklungen stellen das Recht regelmässig vor Herausforderungen; diese Einschätzung gilt auch für die Künstliche Intelligenz. Die EU hat mit einem sehr detaillierten risikoorientierten Regulierungsvorschlag reagiert. Ein solches Konzept dürfte für die Schweiz nicht zielführend sein. Erfolgversprechender ist die Ausarbeitung eines technologieneutralen Rahmengesetzes, unter dessen «Schirm» die regulierungsbedürftigen Themen einer Normierung durch Anpassungen bestehender Gesetze zugeführt werden.

Inhaltsübersicht

A.	Einleitung	18
B.	Regulierungskonzepte und -modelle	19
I.	Gestaltungsaufgabe des Rechts	19
II.	Staatliche und kooperative Regulierung	20
III.	Horizontale und vertikale Regulierung	21
IV.	Regulierungstiefe und -dichte	22
C.	Struktureller und themenbezogener Regulierungsansatz	23
I.	Ausgangslage: Kooperative Regulierung mit Rahmengesetz	23
II.	Probleme bei grenzüberschreitenden Sachverhalten	25
III.	Identifikation der relevanten Themen	26
D.	Regelungspunkte bei KI-relevanten Themen	27
I.	Nichtdiskriminierung und Fairness	27
II.	Transparenz und Erklärbarkeit	28
III.	Manipulation	30
IV.	Datenschutz und Datensicherheit	31
V.	Haftung und Verantwortlichkeit	32
E.	Ausblick	33

A. Einleitung

Die Künstliche Intelligenz (KI) ist zu einem regulatorischen «Hot Topic» geworden. Der technologische Fortschritt im Digitalisierungsbereich rief eine Debatte über die Notwendigkeit der Regulierung von Algorithmen hervor; in einer ersten Phase haben insbesondere internationale Organisationen (OECD, Europarat, EU, UNESCO) mit Leitlinien und politischen Erklärungen die Diskussionen geprägt,¹ nunmehr liegen angesichts des Rufes nach weiteren Normierungen auch erste gesetzliche Vorschläge vor.

Phänomenologisch lassen sich verschiedene Erscheinungsformen von Algorithmen unterscheiden:² Die ursprünglichen deterministischen Algorithmen sind mehr und mehr durch modifizierende selbstlernende Algorithmen, die darauf abzielen, statistische Muster in Datensätzen zu erkennen und sich diese zunutze zu machen, verdrängt worden. Solche Algorithmen, die zur Technologie der Künstlichen Intelligenz gehören, bereiten die menschlichen Entscheidungen vor oder ersetzen sie durch technische Abläufe, obschon ein Algorithmus (im ethisch gehaltvollen Sinne) selbst keine Entscheidung treffen kann.³

Mit dem Entwurf für einen «Artificial Intelligence Act» (AIA) ist die EU-Kommission im April 2021 regulatorisch vorgeprescht.⁴ Das sehr detaillierte Normengefüge basiert auf einem differenzierten risikobasierten Ansatz und reguliert vornehmlich die risikoreichsten Anwendungen.⁵ Die Schweiz ist dadurch auch unter Handlungsdruck gekommen, selbst wenn eine Interdepartementale Kommission des Bundes im Dezember 2019 noch die Meinung vertrat, die

¹ Zwischenzeitlich ist die Zahl der KI-Deklarationen von Institutionen, Organisationen und Verbänden auf über 150 gestiegen; angesichts des schon vorhandenen Schrifttums wird dieses Thema vorliegend nicht vertieft.

² Vgl. Rolf H. Weber/Simon Henseler, Regulierung von Algorithmen in der EU und in der Schweiz, EuZ 2020, 28, 29; soweit nachfolgend themenspezifische Regulierungsaspekte zur Diskussion gelangen, wird auf diesen materialreichen Beitrag verwiesen, ohne die zitierten Quellen umfassend zu wiederholen.

³ Weber/Henseler, 29 m.w.V.

⁴ 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, April 21, 2021, COM(2021) 206 final.

⁵ Vgl. Angela Müller, Der Artificial Intelligence Act der EU: Ein risikobasierter Ansatz zur Regulierung von Künstlicher Intelligenz – mit Auswirkungen auf die Schweiz, EuZ 01/2022, A 3 ff.; Philipp Roos/Caspar Alexander Weitz, Hochrisiko-KI-Systeme im Kommissionsentwurf für eine KI-Verordnung, MMR 2021, 844 ff.

Schweiz sei an sich gut gerüstet, um die neuen technologischen Herausforderungen normativ zu bewältigen.⁶

Das Ziel der nachfolgenden Überlegungen besteht nicht darin, die Chancen und Risiken, die sich für die Gesellschaft aus dem Einsatz von KI ergeben, erneut im Einzelnen darzustellen; die entsprechenden Einschätzungen im publizierten Schrifttum weisen keine erheblichen Diskrepanzen auf. Ebenso wenig soll der AIA-Vorschlag einer detaillierten Analyse der umfangreichen Bestimmungen unterzogen werden.⁷ Vielmehr geht der Beitrag der Frage nach, welcher Regulierungsansatz im Kontext der Künstlichen Intelligenz sich für die Schweiz als sachgerecht erweist.⁸ Aus diesem Grunde sind vorerst die wesentlichen Elemente möglicher Regulierungskonzepte und -modelle aus einer theoretischen Perspektive darzustellen bzw. deren Charakteristiken gegeneinander abzuwägen.

B. Regulierungskonzepte und -modelle

I. Gestaltungsaufgabe des Rechts

Das Recht ist ein strukturelles System, das auf organisierten oder untereinander verbundenen regulatorischen Elementen beruht. Die Funktion des Rechts besteht darin, durch verhaltensbezogene Vorgaben und Institutionen das Zusammenleben in der Gesellschaft zu normieren und Konflikte zu lösen, die

⁶ Interdepartementale Arbeitsgruppe Künstliche Intelligenz, Herausforderungen der Künstlichen Intelligenz, Bericht an den Bundesrat, 13. Dezember 2019, 10 ff. Verabschiedet hat der Bundesrat immerhin Leitlinien «Künstliche Intelligenz» für den Bund, Orientierungsrahmen für den Umgang mit künstlicher Intelligenz in der Bundesverwaltung, 25. November 2020; weiter hat der Bundesrat am 25. August 2021 beschlossen, ein «Kompetenznetzwerk Künstliche Intelligenz, KNW KI» in der ersten Jahreshälfte 2022 ins Leben zu rufen.

⁷ Vgl. dazu statt Vieler etwa Müller, A 5 ff.; Roos/Weitz, 844 ff.; Gerald Spindler, Der Vorschlag der EU-Kommission für eine Verordnung zur Regulierung der Künstlichen Intelligenz (KI-VO-E), CR 2021, 361 ff.; Michael Veale/Frédéric Zuiderveen Borgesius, Demystifying the Draft EU Artificial Intelligence Act, Comp. L. Rev. Int. 2021, 97 ff.

⁸ Diese Frage ist auch der Ausgangspunkt des Beitrags von Nadja Braun Binder/Thomas Burri/Melinda Florina Lohmann/Monika Simmler/Florent Thouvenin/Kerstin Noëlle Vokinger, Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht, Jusletter 28. Juni 2021, Rz. 4; detailliert diskutiert werden die relevanten Themenbereiche, nicht aber das regulierungstheoretische Konzept, weshalb sich der vorliegende Beitrag als komplementäre Ergänzung verstehen lässt.

sich bei der Anwendung der Regeln ergeben.⁹ Die Schaffung von Recht vermag dabei auf verschiedenen, nachfolgend anzusprechenden Quellen zu beruhen.

Gewisse Grundfragen stellen sich aber für jede Rechtsordnung:¹⁰ Wer darf regulieren? In welchem Interesse? Durch welche Mechanismen? Mit welchen Zielen? Ungeachtet der konkreten Beantwortung dieser Fragen erscheint indessen als unbestritten, dass dem Recht eine Gestaltungsaufgabe mit Blick auf die zwischenmenschlichen Beziehungen in der Gesellschaft zukommt.¹¹

II. Staatliche und kooperative Regulierung

Die traditionellen regulatorischen Instrumente finden sich – für grenzüberschreitende Erscheinungen – in den internationalen Staatsverträgen und in den einzelstaatlichen Rechtsordnungen (Verfassung, Gesetz). Diese Rechtsquellen kommen in einem geordneten (oft demokratischen) Verfahren zustande; die einzelnen Normen lassen sich auch obrigkeitlich mit Zwang durchsetzen.¹² Gerade mit Blick auf schnelle technologische Veränderungen hat sich in den letzten Jahrzehnten aber gezeigt, dass staatliches Recht meist nur relativ langsam entsteht und hernach zur Anwendung gelangt.¹³ Andere normative Instrumente erweisen sich deshalb zur Ergänzung des Regulierungsumfelds als notwendig.

Seit Jahren ist anerkannt, dass neben dem staatlichen «Hard Law» auch das «Soft Law» eine gewichtige Rolle zu spielen vermag. Unter den Begriff des Soft Law fallen die verschiedenartigsten Formen von Selbstregulierungen, die auf Initiative der betroffenen Branchen, Unternehmen oder zivilgesellschaftlichen Vereinigungen entstehen.¹⁴ Der Vorteil von Selbstregulierungen liegt darin, dass die entsprechenden Bestimmungen gestützt auf Know How und Erfahrungen der Betroffenen zeitgerecht und zu tiefen Kosten entwickelt werden, der Nachteil hingegen darin, dass sie sich obrigkeitlich nicht durchsetzen

⁹ Für einen neueren Überblick vgl. Rolf H. Weber, Internet Governance at the Point of No Return, Zürich 2021, 13 f. m.w.V.; Amnon Reichman/Giovanni Sartor, Algorithms and Regulation, in: H.-W. Micklitz et al. (eds.), Constitutional Challenges in the Algorithmic Society, Cambridge 2022, 131 ff.

¹⁰ Vgl. Rolf H. Weber, Realizing a New Global Cyberspace Framework, Zürich 2014, 34–36.

¹¹ Grundlegend dazu Herbert L.A. Hart, The Concept of Law, 2nd ed., Oxford 1997, 55 ff.

¹² Allgemein zum Recht des Cyberspace vgl. Chris Reed, Making Laws for Cyberspace, Oxford 2012, 70–73, 105–06.

¹³ Im Schrifttum wird oft von «regulatory lag» gesprochen: vgl. Weber/Henseler, 32.

¹⁴ Für einen ganz neuen Überblick vgl. Rolf H. Weber, Integrity in the 'Infinite Space' – New Frontiers for International Law, ZaöRV 81 (2021), 601, 619 f.

lassen.¹⁵ Soft Law als Form einer kooperativen Regulierung schneidet indessen qualitativ keineswegs schlechter ab als Hard Law.¹⁶

Als Zwischenform hat in den letzten Jahren das Modell der «Co-Regulierung» (bzw. der «regulierten Selbstregulierung») an Bedeutung gewonnen. Dieses Regulierungskonzept verstärkt Selbstregulierungen durch eine staatliche «Anerkennung» bzw. sogar eine Überwachung der Beachtung von Verhaltenspflichten.¹⁷ Vielfältige Beispiele sind aus den Finanz-, Medien- und Internetmärkten bekannt.¹⁸ Dank der Mitwirkung des traditionellen Gesetzgebers bei der Durchsetzung der normativen Bestimmungen verstärkt sich deren Wirkung.¹⁹ Das Modell der Co-Regulierung entspricht dem im Kontext der Internet Governance intensiv diskutierten und teilweise auch angewendeten Multistakeholder-Konzept; die Entwicklung eines erfolgreichen normativen Umfelds setzt voraus, dass alle betroffenen Akteure (z.B. Regierungen/Verwaltungen, Unternehmen, Zivilgesellschaft, Akademie) gemeinsam an der Regelbildung mitwirken.²⁰ Diese Vorzüge sind gerade für den KI-Bereich fruchtbar zu machen.

III. Horizontale und vertikale Regulierung

Mit dem Entwurf für einen Artificial Intelligence Act (AIA) hat die EU-Kommission einen horizontalen Regulierungsansatz gewählt; grundsätzlich ist das risikobasierte Modell für alle Wirtschaftssegmente und Branchen relevant.²¹ Der Vorteil der horizontalen Regulierung liegt darin, dass eine normative Segmentierung vermieden werden kann; nachteilig wirkt sich hingegen aus, dass die Berücksichtigung von branchenspezifischen Besonderheiten nur schwer möglich ist.²²

¹⁵ Rolf H. Weber, *Artificial Intelligence ante portas: Reactions of Law*, J 2021(4), 486, 488.

¹⁶ Weber/Henseler, 33 m.w.V.

¹⁷ Eingehender dazu Weber, 621 f.

¹⁸ Vgl. dazu Chris T. Marsden/Trisha Meyer/Ian Brown, *Platform values and democratic elections: How can the law regulate digital disinformation?* CLSR 36 (2020), 105373, 1, 9 ff.

¹⁹ Vgl. auch Myriam Senn, *Non-State Regulatory Regimes, Understanding Institutional Transformation*, Berlin/Heidelberg 2011, 43, 139-148, 230.

²⁰ Für eine umfassende Darstellung des Multistakeholder-Konzepts vgl. Rolf H. Weber, *Legal foundations of multistakeholder decision-making*, ZSR 135 (2016) I 247-267; zur Anwendung des kooperativen Regulierungsansatzes im KI-Kontext vgl. Weber (Fn. 15), 487 f. und Weber/Henseler, 33.

²¹ Vgl. zum horizontalen risikobasierten Ansatz auch Müller, A 6 ff.

²² Im Finanzmarktrecht hat die FINMA bzw. später der Bundesrat beabsichtigt, die traditionelle vertikale Regulierung durch ein horizontales Modell zu ersetzen. Bis zu einem ge-

Ein vertikaler Regulierungsansatz würde es erlauben, für die betroffenen Wirtschaftssegmente eine möglichst massgeschneiderte Regulierung zu verwirklichen. Der Gesundheitssektor bedarf z.B. nicht zwingend derselben Vorgaben wie die Autobranche. Angesichts der vielfältigen KI-Einsatzmöglichkeiten weist eine sektorspezifische Regulierung deutliche Vorzüge auf; sie vermag den normativen Herausforderungen präziser Rechnung zu tragen und vermeidet, dass allgemeine Anordnungen teilweise ungeeignet sind. Für den KI-Bereich hat dieses Konzept im Vergleich zu einem nur horizontalen Ansatz nicht zu unterschätzende Vorteile.

IV. Regulierungstiefe und -dichte

Der AIA-Entwurf, den die EU-Kommission vorgelegt hat, zeichnet sich durch eine sehr hohe Regulierungstiefe und -dichte aus. Auf über 100 Seiten finden sich viele detaillierte Einzelregeln, deren Umsetzung sich nicht immer leicht vornehmen lässt, was für die Rechtsanwendung und -durchsetzung als nicht unproblematisch erscheint.

Ein anderer Regulierungsansatz würde darin bestehen, konkret zu prüfen, welche Sachverhalte im Interesse des Staates und der Zivilgesellschaft tatsächlich einer Regulierung bedürfen. Dieses Vorgehen hat der Bundesrat mit Bezug auf den durch die Distributed Ledger Technology (DLT) verursachten Handlungsbedarf gewählt. Zwar erklärte der Bundesrat in der Botschaft zum DLT-Gesetz (November 2019), bestmögliche Rahmenbedingungen für DLT-Geschäftsmodelle schaffen zu wollen, damit sich die Schweiz als ein führender, innovativer und nachhaltiger Standort für Blockchain-Unternehmen etablieren und weiterentwickeln könne.²³

Materiell hat der Bundesrat indessen auf den Vorschlag für ein umfassendes Blockchain-Gesetz, wie z.B. im Fürstentum Liechtenstein realisiert, verzichtet, und den bewährten und ausgewogenen Rechtsrahmen der Schweiz nicht

wissen Grade ist dies mit dem FinfraG und dem FIDLEG auch geschehen; im Laufe der Ausarbeitung der neuen Finanzmarktarchitektur hat sich aber gezeigt, dass einzelne vertikale Regulierungen in Kraft bleiben müssen (z.B. neben dem FiniG das BankG, das VAG, das KAG), um branchengerechte Normierungen zu erreichen (vgl. *Thomas Jutzi/Ksenia Wess/Damian Sieradzki, Die neue Finanzmarktarchitektur im europäischen Regulierungskontext*, AJP 2020, 572, 583 m.w.V.).

²³ Botschaft des Bundesrates vom 27. November 2019 zum Bundesgesetz zur Anpassung des Bundesrechts an Entwicklungen der Technik verteilter elektronischer Register, BBl 2020, 233, 239.

grundsätzlich in Frage gestellt.²⁴ Vielmehr nimmt das zwischenzeitlich vom Parlament (einstimmig) verabschiedete DLT-Gesetz lediglich diejenigen Anpassungen in einzelnen Gesetzen vor, die erforderlich gewesen sind, um den DLT-Anwendungen nicht normative Lücken oder Hindernisse in den Weg zu stellen.²⁵

Eine ähnliche Vorgehensweise erscheint auch mit Bezug auf die Künstliche Intelligenz als denkbar und sachgerecht. Nach einer Identifikation der relevanten Themen und der Analyse des konkreten Normierungsbedarfs ist zu beurteilen, welcher theoretische Regulierungsansatz sinnvoll ist. In der Umsetzung liesse sich dann – ähnlich wie im Fall des DLT-Gesetzes – ein «Bundesgesetz zur Anpassung des Bundesrechts an Entwicklungen der Künstlichen Intelligenz» ausarbeiten. Im Gegensatz zum DLT-Gesetz dürfte es indessen nicht ausreichen, in der Einleitung lediglich festzuhalten, die «nachstehenden Erlasse werden wie folgt geändert», sondern zu einzelnen Themen (z.B. Nichtdiskriminierung, IT-Sicherheit) sind ggf. «neue» Gesetze vorzubereiten.

C. Struktureller und themenbezogener Regulierungsansatz

I. Ausgangslage: Kooperative Regulierung mit Rahmengesetz

Die Künstliche Intelligenz ist eine moderne und sich noch stark im Fluss befindliche Technologie. Ein gewisser Handlungsbedarf lässt sich aber bereits identifizieren. Die Wahrscheinlichkeit ist überdies erheblich, dass zusätzliche Fragen und Probleme erst im Laufe der nächsten Monate und Jahre auftreten. Aus diesem Grunde drängt sich ein flexibles normatives Regelwerk auf. Im Sinne der vorangehenden theoretischen Überlegungen sind auch die Vorteile der vertikalen (sektorspezifischen) gegenüber der horizontalen Regulierung fruchtbar zu machen.

Wie erwähnt ist die Erarbeitung und Inkraftsetzung von staatlichen Rechtsinstrumenten regelmässig zeitintensiv; neue Regelungen hinken deshalb den Technologien nach, wie die bereits einsetzende Diskussion zur sog. technologieneutralen Umschreibung des Begriffs der Künstlichen Intelligenz im

²⁴ Botschaft, 239 f.

²⁵ Für einen Überblick vgl. Hans Kuhn/Rolf H. Weber, Einleitung, in: R.H. Weber/H. Kuhn, Entwicklungen im Schweizer Blockchain-Recht, Basel 2021, Kap. 1, Rz. 1 ff.

AIA-Vorschlag zeigt.²⁶ Aus diesem Grunde erscheint es als wünschenswert, Soft Law-Komponenten im KI-Regulierungsumfeld ebenfalls zu berücksichtigen. Diese Vorgehensweise lässt sich am ehesten verwirklichen, wenn sich die staatliche Mitwirkung an der Normsetzung auf ein Rahmengesetz beschränkt und ein kooperativer Regulierungsansatz verwirklicht wird.

In überzeugender Weise werden im Europarat mit seinen 47 Mitgliedern mögliche Modelle für eine Regulierung der Künstlichen Intelligenz diskutiert. Das am 11. September 2019 eingesetzte Ad hoc Committee on Artificial Intelligence (CAHAI) hat im umfangreichen Bericht vom 17. Dezember 2020²⁷ nicht nur die Chancen und Risiken der Künstlichen Intelligenz analysiert und insbesondere die Relevanz von Menschenrechten und Demokratie für KI-Regulierungen diskutiert, sondern auch die möglichen Formen von rechtlichen Instrumenten erläutert. Das Kapitel «Mapping of Instruments Applicable to Artificial Intelligence»²⁸ hält zwar dafür, dass die bestehenden Ethik-Leitlinien regulatorisch nicht ausreichen würden, aber doch verschiedene Optionen verbleiben, um eine rechtliche Ordnung für KI sinnvoll festzulegen.

Insbesondere bringt der CAHAI-Report die Idee einer «Framework Convention» auf.²⁹ Im Gegensatz zum AIA-Vorschlag der EU, der eine umfassende und detaillierte Regulierung postuliert, öffnet der Europarat zutreffend den Weg zu einer differenzierten rechtlichen Strukturierung der normativen Vorgaben. Dieser Vorschlag überlässt den Mitgliedstaaten einen nicht unerheblichen Freiraum, den auch die Schweiz mit einem eigenen Ansatz ausnutzen könnte,³⁰ und weist zudem Ähnlichkeiten zum erwähnten Modell eines «Rahmengesetzes» auf, den der Bundesrat mit Blick auf die DLT-Entwicklungen gewählt hat.³¹

²⁶ Zu diesem vorliegend nicht zu vertiefenden Problem vgl. Jonas Fischer/Mathias Fuchs, Brauchen wir eine Legaldefinition für künstliche Intelligenz? Jusletter 8. November 2021, Rz. 18 ff.

²⁷ Council of Europe, Ad hoc Committee on Artificial Intelligence, Feasibility Study, CAHAI(2020) 23, <https://www.coe.int/en/web/artificial-intelligence/cahai> (CAHAI-Report).

²⁸ CAHAI-Report, 18 ff.

²⁹ CAHAI-Report, 46 ff. CAHAI hat an diesem Konzept im Final Meeting vom 29.11.–2.12.2021 ausdrücklich festgehalten («Possible elements of a legal framework on artificial intelligence»).

³⁰ So auch Florent Thouvenin/Markus Christen/Abraham Bernstein/Nadja Braun Binder/Thomas Burri/Karsten Donnay/Lena Jäger/Mariela Jaffé/Michael Krauthammer, Melinda Lohman/Anna Mätzener/Sophie Mützel/Liliane Obrecht/Nicole Ritter/Matthias Spielkamp/Stephanie Volz, Digital Society Initiative, Positionspapier – Künstliche Intelligenz, November 2021, 2, <https://www.itsl.uzh.ch/en/Knowledge-transfer/Publications.html>.

³¹ Vgl. vorne [Ziff. B.IV](#).

Ein entsprechendes Regulierungskonzept erweist sich auch für den künftig in der Schweiz normativ zu bewältigenden KI-Bereich als sachgerecht. Das Modell des Rahmengesetzes wird im Schrifttum als mögliche Regulierungsform anerkannt und gerade in komplexen Materien für sinnvoll gehalten.³² Statt ein horizontales umfassendes «KI-Gesetz» zu erlassen, können in ein KI-Rahmengesetz allgemeine und spezifische Normen, jeweils – soweit möglich – in Abänderung und Ergänzung bestehender Rechtsquellen, aufgenommen werden.³³

II. Probleme bei grenzüberschreitenden Sachverhalten

Die Angebote für durch Künstliche Intelligenz geprägte Produkte und Dienstleistungen bleiben nicht auf die Schweiz beschränkt. Vielmehr ist vorauszusehen, dass auch EU-Märkte bedient werden sollen. Der Entwurf der EU für die AIA-Verordnung sieht ausdrücklich vor, dass bei Angeboten an EU-Verbraucher deren Bestimmungen zu beachten sind.³⁴

Die AIA-Verordnung wird also das Schutzniveau in der Schweiz mitprägen und der Schweizer Bevölkerung einen grundrechtlichen Schutz anbieten, sofern sich Unternehmen in der EU einzelner AI-Vorkehren in der Produktion von Gütern und der Bereitstellung von Dienstleistungen bedienen.³⁵ Eine gewisse materielle Kohärenz in der Ausgestaltung der Regulierungen wäre deshalb erwünscht, um zu vermeiden, dass der Zugang von Produkten und Dienstleistungen zum EU-Binnenmarkt erschwert wird.³⁶ Eine gesetzgeberische Lösung in der Schweiz müsste somit versuchen, sachlich vergleichbare bzw. EU-kompatible KI-Standards zu realisieren.

Äquivalenz bedeutet aber nicht Identität der Regulierungen. Diese Einschätzung zeigt sich seit Jahren in der Beurteilung des vergleichbaren Datenschutzniveaus; ausländische Datenschutzgesetze müssen nicht genau gleiche Schutzprinzipien und -mechanismen aufweisen, sondern «lediglich» in einer «äquivalenten» Weise die Vertraulichkeit der Daten von Schweizer Personen

³² Kuhn/Weber, Rz. 17-19; für das Finanzmarktrecht vgl. Pascal Zysset, Selbstregulierung im Finanzmarktrecht, Diss. Bern, Zürich 2017, 207 f. m.w.V.

³³ Dieser Regulierungsansatz wird nun auch von Thouvenin et al., 2, vertreten.

³⁴ Ähnlich wie bei der EU-DSGVO kommt nach Art. 2(1) des AIA-Vorschlags das EU-Recht zur Anwendung, wenn ein KI-System innerhalb der EU eingesetzt wird oder wenn das vom System hervorgebrachte Ergebnis innerhalb der EU verwendet wird.

³⁵ Vgl. auch Müller, A 5 f.

³⁶ Ob solche Handelshemmnisse ggf. einen Verstoß des WTO-Rechts zur Folge hätten, wäre gesondert zu prüfen; zum Ganzen auch Braun Binder et al., Rz. 4.

sicherstellen.³⁷ Zudem sollte die Schweiz nicht vorschnell auf einen durch ein weniger dichtes Regulierungsmodell bewirkten Wettbewerbsvorteil verzichten, nicht zuletzt auch mit Blick auf die vielen Abnehmer ausserhalb der EU, denn KI-Systeme lassen sich global vermarkten.

III. Identifikation der relevanten Themen

Wie erwähnt ist ein horizontaler Regulierungsansatz regelmässig gezwungen, eine hohe Regulierungsdichte zu realisieren, und hat dennoch den Nachteil, dass sektorspezifische Besonderheiten nicht berücksichtigt werden können. Wird von einem solchen umfassenden horizontalen Regulierungsansatz abgesehen, erweist es sich indessen als notwendig, die relevanten Themen der Künstlichen Intelligenz, die einen Handlungsbedarf auslösen, zu identifizieren. Nach zwischenzeitlich verbreiteter Meinung stehen folgende Themen im Vordergrund:³⁸

- Nichtdiskriminierung und Fairness;
- Transparenz und Erklärbarkeit;
- Manipulation;
- Datenschutz und Datensicherheit;
- Haftung und Verantwortlichkeit.

Abgesehen von der Identifikation der relevanten Themen hat der Gesetzgeber auch die regulatorischen Instrumente festzulegen, die zum Einsatz kommen sollen. Der stärkste Eingriff würde in einem Verbot bestehen; eine Alternative besteht in der Festlegung von Zulassungsbedingungen, die der Anbieter von KI-Produkten oder -Dienstleistungen einzuhalten hat. Verbote sind grundsätzlich nur gerechtfertigt, wenn die Sicherheitsrisiken als so erheblich erscheinen, dass eine sachgerechte Minimierung mit möglichen Vorsichtsmassnahmen nicht in Frage kommt;³⁹ jedes Verbot führt nämlich dazu, dass die Angebote für entsprechende KI-Produkte und -Dienstleistungen völlig vom Markt verschwinden. Zulassungsverfahren sind insbesondere geeignet, die

³⁷ Art. 16 Abs. 1 des neuen DSGVO verlangt einen «angemessenen Schutz» im Ausland, nicht das Vorliegen identischer Datenschutzgrundsätze.

³⁸ Zu den einzelnen Themen vgl. nachfolgend [Ziff. D.](#)

³⁹ Mögliche Beispiele für ein Verbot sind der Einsatz von Gesichtserkennung und anderen biometrischen Fernerkennungsverfahren im öffentlichen Raum zum Zwecke der Massenüberwachung oder der Einsatz von Social Scoring mit dem Ziel, eine Zugangsregulierung zu grundlegenden sozialen Diensten zu bewirken; vgl. *Thouvenin et al.*, 6 f.

Realisierung von Sicherheitsvorkehrungen zu gewährleisten; im Einzelnen ist jeweils zu prüfen, ob bereits bestehende Zulassungsverfahren erweitert werden können oder ob es sich aufdrängt, neue Zulassungsverfahren zu schaffen.⁴⁰

D. Regelungspunkte bei KI-relevanten Themen

Die nachfolgend angesprochenen, im KI-Kontext relevanten Themen betreffen besondere politik- und marktrelevante Bereiche. Potentielle KI-Risiken vermögen sich aber auch auf grundlegende Rechtsprinzipien auszuwirken; zu treffend thematisiert der Europarat im CAHAI-Report z.B. die Menschenwürde, die Grundfreiheiten, die Demokratie und die «Rule of Law».⁴¹ Die materielle Diskussion der «ausgewählten» fünf wichtigen Themenbereiche erfolgt bewusst knapp, weil vertiefendes Schrifttum bereits vorhanden ist⁴² und die rechtliche Rahmenordnung in einem Forschungsprojekt während der nächsten drei Jahre detailliert analysiert werden soll.⁴³

I. Nichtdiskriminierung und Fairness

Die zentralste Herausforderung bei der KI-Regulierung betrifft wohl den Bereich der Nichtdiskriminierung bzw. Gleichbehandlung. Beispiele, die zeigen, dass gewisse menschliche Merkmale zu unfairen oder gar diskriminierenden Behandlungen führen können, weil mit einzelnen Merkmalen ein «Bias» verbunden sein kann, dürfen als weitherum bekannt gelten.⁴⁴ Betroffen vom Grundsatz der Nichtdiskriminierung sind nicht nur private Unternehmen, sondern ebenso die öffentliche Verwaltung.⁴⁵

⁴⁰ Vgl. Thouvenin et al., 6.

⁴¹ CAHAI-Report, 27 ff.; zur Menschenwürde, zu den Grundfreiheiten und zu den rechtsstaatlichen Prinzipien eingehend Rolf H. Weber, *Automatisierte Entscheidungen: Perspektive Grundrechte*, SZW 2020, 18 ff. m.w.V.

⁴² Für weiterführende Überlegungen vgl. Weber/Henseler, 33 ff. und Braun Binder et al., Rz. 8 ff.; beide Beiträge enthalten umfangreiche weitergehende Hinweise, auf deren erneute Erwähnung im Interesse einer «schlanken» Darstellung verzichtet wird.

⁴³ Für einen Bericht über die Eröffnungsveranstaltung vom 10. November 2021 vgl. Fabienne Graf/Liliane Obrecht, *Rechtliche Rahmenbedingungen für Künstliche Intelligenz in der Schweiz*, Jusletter 29. November 2021, Rz. 1 ff.

⁴⁴ Im Einzelnen vgl. Weber/Henseler, 31 f. und 39 ff. sowie Braun Binder et al., Rz. 21 ff., je m.w.V.

⁴⁵ Für ein (österreichisches) Beispiel vgl. Braun Binder et al., Rz. 23.

Das Gebot der rechtsgleichen Behandlung ist verfassungsrechtlich verankert (Art. 8 Abs. 1 BV); immerhin bedarf dieser Grundsatz einer Verfeinerung, um sinnvoll anwendbar zu sein.⁴⁶ Zu beachten bleibt weiter, dass «Fairness» und «Bias» in einzelnen Disziplinen weiter gefasst werden.⁴⁷ Besonders problematisch ist die indirekte bzw. verdeckte Diskriminierung, die sich ggf. lediglich in einzelnen Auswirkungen von Einschätzungen zeigt.⁴⁸

Im Gegensatz zu anderen Themenbereichen dürfte sich zur Sicherstellung von Nichtdiskriminierung und Fairness ein horizontaler (allgemeiner) Gesetzeserlass, der öffentliche Verwaltungen und weitergehend als heute insbesondere private Unternehmen betrifft, als notwendig erweisen,⁴⁹ da die Ergänzung einer grossen Zahl von bestehenden Gesetzen kaum sinnvoll zu bewerkstelligen wäre.

II. Transparenz und Erklärbarkeit

Der Einsatz von Künstlicher Intelligenz muss transparent erfolgen; die betroffenen Personen sind zudem in die Lage zu versetzen, die KI-Vorgehensweise zu verstehen, was deren Erklärbarkeit und Interpretierbarkeit voraussetzt. Die Transparenz und Nachvollziehbarkeit von KI-Vorgängen sind adressatengerecht auszugestalten, abhängig vom Verständnishorizont des Betroffenen und der Bedeutung der Entscheidung für die erfasste Person; die der automatisierten Entscheidung zugrunde liegende Logik muss verständlich sein und die notwendigen Informationen enthalten, um einen ausreichenden Grad an Nachvollziehbarkeit zu erreichen.⁵⁰

Denkbar wäre weiter die Einführung einer Kennzeichnungspflicht bei der Verwendung von algorithmischen Systemen. Um die Erkennbarkeit des KI-Einsatzes für die interessierte Öffentlichkeit zu gewährleisten, würde auch die Möglichkeit bestehen, ein öffentlich zugängliches Register zu schaffen, das

⁴⁶ Zum sekundärrechtlichen bzw. einfachgesetzlichen Diskriminierungsschutz vgl. Weber/Henseler, 39 f.

⁴⁷ Braun Binder et al., Rz. 25 f.

⁴⁸ Braun Binder et al., Rz. 24 und 27 ff. m.w.V.

⁴⁹ In diese Richtung gehen wohl auch Thouvenin et al., 4; ob eine «organische Weiterentwicklung» des heutigen Rechts ausreicht (so Braun Binder et al. [Fn. 8], Rz. 30), erscheint mit Bezug auf den Nichtdiskriminierungsgrundsatz indessen als zweifelhaft.

⁵⁰ Thouvenin et al., 3; zu den spezifischen Transparenzanforderungen des EU-Rechts, das weiter entwickelt ist als das Schweizer Recht (z.B. im Gesundheitsbereich), vgl. Weber/Henseler, 37 f.

aufzeigt, in welchen Bereichen die öffentliche Verwaltung einzelne KI-Systeme einsetzt.⁵¹

Im Datenschutz- und im Verbraucherrecht (stärker in der EU als in der Schweiz) sind erste Regulierungsansätze bereits vorhanden. So enthalten z.B. die DSGVO und das neue DSG der Schweiz eine spezifische Informationspflicht und ein Auskunftsrecht bei automatisierten Entscheidungen; offenzulegen sind aussagekräftige Informationen über die involvierte Logik sowie die Tragweite und die angestrebten Auswirkungen einer KI-Verarbeitung.⁵² Zwar sind Inhalt und Ausmass der gesetzlichen Informationspflichten bzw. Auskunftsrechte im Einzelnen umstritten, doch hat sich zwischenzeitlich die Meinung durchgesetzt, die Grundannahmen des Algorithmus seien offenzulegen, nicht aber der Algorithmus selbst.⁵³

Entsprechende rechtliche Regelungen erweisen sich insbesondere im Gesundheitsbereich als angebracht. Der AIA-Vorschlag erklärt den Einsatz von KI in medizinischen Instrumenten als mit hohem Risiko behaftet.⁵⁴ Für solche Systeme gelten spezifische Transparenzpflichten; offenzulegen ist, mit welchen Daten die KI trainiert wurde und wie sie funktioniert und wirkt.⁵⁵ Die Transparenzschaffung kann aber in Widerspruch zu den Geheimhaltungsinteressen der die KI-Systeme entwickelnden Unternehmen geraten, was einen konkreten Interessenabwägungsprozess im Lichte der gegebenen Umstände erforderlich macht.⁵⁶

Einschränkungen mit Bezug auf die Transparenzschaffung sind auch beim raum- bzw. zeitbezogenen «Predictive Policing» unumgänglich, um zu vermeiden, dass die vorausschauende Polizeiarbeit schon frühzeitig offengelegt werden muss.⁵⁷ Mit dem Predictive Policing lassen sich anhand statistischer Prognosen wahrscheinliche Vorfälle identifizieren, was der Polizei das Ergreifen präventiver (aus Sicherheitsgründen als sinnvoll erscheinender) Massnahmen ermöglicht.⁵⁸

⁵¹ Thouvenin et al., 3.

⁵² Art. 13 Abs. 2 lit. f und Art. 14. Abs. 2 lit. g DSGVO; Art. 21 Abs. 1 und Art. 25 Abs. 2 lit. f rev. DSG.

⁵³ Weber/Henseler, 35, und Braun Binder et al., Rz. 12, je m.w.V.

⁵⁴ Art. 5 i.V.m. Anhang 2A des AIA-Vorschlags.

⁵⁵ Im Einzelnen dazu Braun Binder et al., Rz. 9 und 15.

⁵⁶ Braun Binder et al., Rz. 10 m.w.V.

⁵⁷ Vgl. auch Weber (Fn. 41), 24 f.

⁵⁸ Braun Binder et al., Rz. 13.

III. Manipulation

Künstliche Intelligenz lässt sich einsetzen, um menschliches Verhalten zu beeinflussen. Ein solcher Eingriff in die Autonomie der betroffenen Person vermag unbewusst auf das Denken und Handeln einer Person einzuwirken. Besonders problematisch ist die verdeckte Beeinflussung (z.B. durch Verwendung von Empfehlungsalgorithmen auf Plattformen der sozialen Medien).⁵⁹

Manipulationen können auch im Kontext der demokratischen Prozesse vorkommen und damit z.B. den Schutz der freien Willensbildung beeinträchtigen.⁶⁰ Weltweit diskutiert wird die Manipulation insbesondere im Vorfeld von Wahlen und Volksabstimmungen; immerhin lässt sich nicht übersehen, dass bereits heute gewisse rechtliche Rahmenbedingungen bestehen, um – unabhängig von den genutzten Informationskanälen – gegen schwerwiegende Desinformationen vorzugehen.⁶¹ Allgemein betrachtet steht das Recht indessen bei der Erfassung von Manipulationen und Desinformationen noch am Anfang.

Gewisse rechtliche Regelungen, deren Zweck darin besteht, Manipulationen zu vermeiden, sind hingegen im privatrechtlichen Bereich vorhanden: Das Verbreiten von manipulativer Information fällt in den Anwendungsbereich des strafrechtlichen Ehrverletzungsschutzes (Art. 173 ff. StGB) und des allgemeinen Persönlichkeitsrechts (Art. 28 ZGB). Diese rechtlichen Regelungen dürften geeignet sein, die wesentlichsten Desinformationen durch KI zu erfassen.⁶²

Weiter enthält das Gesetz gegen den unlauteren Wettbewerb verschiedene Bestimmungen, die im Rahmen der Verbreitung marktrelevanter Desinformationen von Bedeutung sind (insbesondere Art. 3 Abs. 1 lit. b, lit. d und lit. i UWG sowie allgemein Art. 2 UWG).⁶³ Dennoch bleibt genauer zu analysieren, inwieweit eine Konkretisierung des Manipulationstatbestandes im UWG nicht sinnvoll wäre.

Der AIA-Vorschlag sieht ein Verbot für gewisse Formen des manipulativen Gebrauchs von KI und für KI-Anwendungen vor (Art. 5 Abs. 1 lit. a und b). Die Formulierung des Tatbestandes ist aber sehr offen umschrieben und auch unge-

⁵⁹ Vgl. auch Thouvenin *et al.*, 4.

⁶⁰ Thouvenin *et al.*, 4 f.

⁶¹ Vgl. BGE 140 I 338, E.5.3; Braun Binder *et al.*, Rz. 36.

⁶² Vgl. dazu den kürzlich erschienenen BAKOM-Bericht, Intermediäre und Kommunikationsplattformen – Auswirkungen auf die öffentliche Kommunikation und Ansätze einer Governance, Bericht vom 17. November 2021.

⁶³ Braun Binder *et al.*, Rz. 35 und 38.

nügend zielgerichtet, weshalb sich eine Übernahme in das Schweizer Recht nicht aufdrängt; vielmehr erscheint die hiesige Rechtsordnung als hinreichend flexibel, um die relevanten Probleme zu erfassen, die durch Manipulationen mittels KI entstehen könnten.⁶⁴

IV. Datenschutz und Datensicherheit

Der Schutz der Privatsphäre war historisch betrachtet das erste Thema, das im Kontext der Künstlichen Intelligenz heiss diskutiert worden ist.⁶⁵ Die Schaffung der DSGVO in der EU sowie des neuen DSG in der Schweiz scheint den Herausforderungen aber an Brisanz genommen zu haben.

Insbesondere die bereits erwähnten Bestimmungen zu den automatisierten Entscheidungen, aber auch die Anordnungen zu den Datenschutz-Folgeabschätzungen, nehmen den Regelungsbedarf, der durch den Einsatz von Algorithmen entsteht, detailliert auf. Einzelne Verfeinerungen lassen sich diskutieren⁶⁶ und sind auch angebracht; deren Realisierung dürfte aber nicht auf unüberwindbare Hindernisse stossen.

Unabhängig von diesen (theoretischen) Überlegungen bleibt indessen im Auge zu behalten, dass insbesondere die Vorgaben des Datenschutzrechts bereits sehr weit gediehen sind und es voraussichtlich in vielen Einzelkonstellationen weniger um legislatorische Ergänzungen wegen Regelungslücken als um die konkrete Umsetzung vorhandener Normierungen geht.⁶⁷

Die Datensicherheit ist schon heute im digitalen Umfeld von grosser Bedeutung. Mit KI-Anwendungen liegt die Messlatte für Sicherheitsstandards aber noch höher. Insbesondere stellt sich die Frage, ob die Schweiz – ähnlich wie andere Länder – ein allgemeines IT-Sicherheitsgesetz erlassen sollte.⁶⁸ Im Sinne einer kooperativen Regulierung ist es indessen durchaus denkbar, dass private Organisationen sich bemühen, angemessene IT-Sicherheitsstandards zu entwickeln, wie dies insbesondere durch die International Standardisation Organisation (ISO) und auch schweizerische Vereinigungen bereits geschehen ist.⁶⁹

⁶⁴ Vgl. auch Braun Binder *et al.*, Rz. 39.

⁶⁵ Für einen Überblick vgl. Weber/Henseler, 34 ff. m.w.V.

⁶⁶ Vgl. dazu Braun Binder *et al.*, Rz. 20.

⁶⁷ Vgl. Weber/Henseler, 35 f.

⁶⁸ Thouvenin *et al.*, 6.

⁶⁹ Vgl. dazu schon Rolf H. Weber/Annette Willi, IT-Sicherheit und Recht, Zürich 2006, 67 ff. und 74 ff. m.w.V.

V. Haftung und Verantwortlichkeit

In der Öffentlichkeit intensiv diskutiert wird die Problematik der zivilrechtlichen Haftung und der strafrechtlichen Verantwortlichkeit für die Verursachung von Schäden im Falle des Einsatzes von KI im Strassenverkehr.⁷⁰ Ohne Zweifel ergibt sich in diesem Bereich ein Handlungsbedarf, und zwar nicht zuletzt angesichts der Tatsache, dass die traditionellen Haftungsregeln (insbesondere diejenigen des Produkthaftungsrechts) an die physischen Güter und nicht an die körperlosen Dienstleistungen bzw. an die Software anknüpfen.⁷¹

Die EU ist mit ihren Vorbereitungsarbeiten zur Anpassung der Produkthaftungspflicht-Richtlinie schon recht weit gediehen; zu klären ist immerhin noch die Einordnung digitaler Dienste.⁷² Im Vordergrund steht die Frage, ob eine isolierte Steuerungssoftware als (Teil)Produkt qualifiziert werden kann, eine Annahme, die auch in der Schweiz vermehrt Zustimmung findet.⁷³ In dieser Betrachtungsweise liesse sich Software in jeglicher Form als «typische Erscheinung der fortschreitenden Technisierung» unter den Anwendungsbereich des Produkthaftungspflichtgesetzes subsumieren.⁷⁴ Weiter klärungsbedürftig ist die Konkretisierung des Begriffs der Fehlerhaftigkeit im Falle von KI-Anwendungen.⁷⁵ Immerhin lässt sich nicht übersehen, dass die Schweizer Rechtsprechung die deliktische Produzentenhaftung durch eine extensive Auslegung der ausservertraglichen Norm von Art. 55 OR bereits seit Jahrzehnten erheblich erweitert hat.⁷⁶

Ein besonderes Thema ist die Risikoprävention. Nach dem schon seit Jahren bekannten Konzept der Zuordnung von Risikosphären ist zu prüfen, welche Partei in der Lieferkette für gewisse Risiken am besten einzustehen vermag.⁷⁷ Grundsätzlich lässt sich sagen, dass die KI-Herstellerin am ehesten die Gefährdung durch sorgfältiges Programmieren zu beherrschen vermag; indessen

⁷⁰ Vgl. Melinda F. Lohmann, *Automatisierte Fahrzeuge im Lichte des Schweizer Zulassungs- und Haftungsrechts*, Baden-Baden 2016, 211 ff.

⁷¹ Melinda F. Lohmann, *Ein zukunftsfähiger Haftungsrahmen für Künstliche Intelligenz. Warum die Schweiz ihr Produkthaftungsrecht aktualisieren muss*, HAVE 2021, 111 ff.

⁷² Vgl. auch Braun Binder et al., Rz. 41 f. m.w.V.

⁷³ Walter Fellmann, *Haftpflchtrecht im Zeichen der Digitalisierung*, HAVE 2021, 105, 109.

⁷⁴ Fellmann, 107; Braun Binder et al., Rz. 43.

⁷⁵ Vgl. dazu Braun Binder et al., Rz. 44 f. m.w.V.

⁷⁶ Diese Rechtsprechung hat vor fast 40 Jahren mit dem sog. Schachtrahmen-Entscheid des Bundesgerichts (BGE 110 II 456 ff.) ihren Anfang genommen.

⁷⁷ Zum Konzept der Risikosphären vgl. Rolf H. Weber, *Smart Contracts: Vertrags- und verfügungsrechtlicher Regelungsbedarf? sic!* 2018, 291, 297 f.

verringert sich bei lernfähigen KI-Anwendungen, die nach dem Inverkehrbringen aufgrund der Nutzung trainiert werden, die Kontrolle.⁷⁸ Entlastungsmöglichkeiten ergeben sich aus der unsachgemässen Änderung eines KI-Systems in der kundenseitigen Anwendung oder im Falle von sog. Entwicklungsrisiken.⁷⁹

Gesetzgeberischer Handlungsbedarf besteht somit im Haftungsrecht, selbst wenn einzelne bestehende Haftungsnormen (z.B. Art. 55 OR) durchaus analog auch im KI-Kontext angewendet werden können und die Umschreibung von *ex ante* Sorgfaltspflichten im traditionellen rechtlichen Umfeld als möglich erscheint. Für eine Schweizer Regulierung lässt sich indessen auf die bereits schon weit vorangeschrittenen Diskussionen im EU-Raum zurückgreifen: In Frage kommen z.B. die Einführung sektorspezifischer Gefährdungshaftungsnormen aufgrund der Herstellung oder Nutzung von KI-Systemen oder die Schaffung einer allgemeinen Gefährdungshaftung, kombiniert mit einer Versicherungspflicht.⁸⁰

E. Ausblick

Der Gesetzgeber steht auch in der Schweiz vor neuen Herausforderungen im KI-Kontext. Die Notwendigkeit der Schaffung konkreter Regulierungen, die als unausweichlich erscheinen, um die KI-Risiken sachgerecht in den Griff zu bekommen, bedeutet aber nicht, dass zwingend der sehr detaillierte und teilweise problematische AIA-Vorschlag der EU «ungekürzt» übernommen werden muss. Vielmehr erweist es sich für die Schweiz als sinnvoll, gestützt auf einen kooperativen Regulierungsansatz in der Form eines Rahmengesetzes punktuell Anpassungen der bestehenden Normen in den betroffenen Rechtsbereichen vorzunehmen, soweit ein Handlungsbedarf vorliegt.⁸¹

Das Modell eines Rahmengesetzes, das – ähnlich wie das DLT-Gesetz – als «Schirm» die verschiedenartigen normativen Anpassungen zusammenhält, erweist sich als eine sinnvolle Vorgehensweise. Konkrete gesetzliche Ergänzungen mit horizontaler Wirkung sind im Datenschutz-, Persönlichkeits- und Lauterkeitsrecht erforderlich, doch wäre deren Ausmass gut überblickbar. Genauere vertikale Regelungen drängen sich im Gesundheitsrecht auf. Etwas komplexer ist die Situation im Haftungsrecht (erweiterte «Produkthaftung»),

⁷⁸ Lohmann (Fn. 71), 120; Braun Binder et al., Rz. 46.

⁷⁹ Art. 5 Abs. 1 lit. b und lit. e PrHG; Braun Binder et al., Rz. 47 f.

⁸⁰ Im Einzelnen dazu Braun Binder et al., Rz. 50 m.w.V.

⁸¹ Vgl. auch Braun Binder et al., Rz. 55.

weil eine Kombination von horizontalen und vertikalen Regeln ins Auge zu fassen ist.

In zwei Bereichen wird – weitergehendender als im DLT-Gesetz – zu analysieren sein, ob nicht eine umfassende horizontale Regulierung anzustreben wäre, nämlich beim Thema der Nichtdiskriminierung im privaten Bereich und beim Thema der IT-Sicherheit. Zwar wäre dieser Regulierungsansatz aufwendiger, aber im gesamtgesellschaftlichen Interesse an diesen beiden Materien wohl sachgerecht.

Bei der konkreten Ausgestaltung der neuen Gesetzesbestimmungen sind die erwünschte Äquivalenz zum EU-Recht und das Erfordernis der technologie-neutralen Erfassung von KI-System im Auge zu behalten. Zur Verwirklichung dieser beiden Anliegen muss aber nicht der AIA-Vorschlag der EU «kopiert» werden, sondern es bleibt ausreichender Spielraum für eine den Schweizer Bedürfnissen entsprechende KI-Gesetzgebung, z.B. gestützt auf ein materielles, in kooperativer Regelbildung zu verwirklichendes Konzept, das stärker auf die Vermeidung negativer Entwicklungen (Diskriminierung, Manipulation) als auf die Einhaltung risikodeterminierter Standards ausgerichtet ist.⁸²

⁸² Vgl. vorne [Ziff. C.II.](#)

Künstliche Intelligenz im Spiegel rechtlicher Grundprinzipien

Dürfen Maschinen über Menschen entscheiden?

Eine rechtliche Auslegeordnung im Lichte neuer Technologien

Erstveröffentlichung in: Schweizer Monat, Februar 2019, 72–75

In der öffentlichen Diskussion um die Grenzen des Einsatzes von selbst lernenden Maschinen und künstlicher Intelligenz (KI) steht oft die Frage nach der Verantwortung oder der Haftung im Vordergrund, namentlich wenn es um populäre Beispiele von Unfällen mit Robotern oder selbstfahrenden Autos geht. Dabei wird vergessen: die Haftungsproblematik als Folge einer Schadensverursachung stellt sich eigentlich erst ganz am Ende der Kausalkette. Es wäre stattdessen angezeigt, die Debatte zu den rechtlichen Leitplanken algorithmischer Entscheidungen mit der Frage zu beginnen, wo und *wie das Rechtssystem die Entscheidungen und Handlungen definiert*. Denn: Dank neuer Technologien ist es möglich geworden, immer mehr – früher von Menschen getroffene – Entscheidungen an Maschinen zu delegieren. Und die Dynamik dieser Technologien, einschliesslich der Tatsache, dass ihre Entscheidungsgeschwindigkeit zunimmt und auch ihre Entscheidungsqualität (d.h. die Effizienz bei der Entscheidungsfindung) bei Maschinen oft höher ist als bei Menschen, kann in Konflikt geraten mit dem eher statischen Recht, das zuvorderst eine stabile und Verlässlichkeit bringende Ordnung verwirklichen will.

Diese Entwicklung stellt die Menschheit vor Fragen, die sie in ihrer bisherigen Geschichte nicht beantworten musste, und die nun plötzlich sehr aktuell werden. Etwa: Sind automatisierte Entscheidungen anders geartet als menschliche Entscheidungen? Und wenn ja, wie reflektiert das Recht diesen Umstand? Haben die Maschinen schon heute ein Recht, Entscheidungen zu fällen – oder sollen sie ein solches erhalten? Und wenn ja: In welchen Lebensbereichen könnte dies der Fall sein? Um diese Fragen zu beantworten, erscheint es als sinnvoll, ein wenig auszuholen.

Recht als Rahmenordnung für Technologien

Klar ist: Das Recht setzt die Leitplanken für das menschliche Verhalten und die gesellschaftliche Organisation. Mit seinen Anordnungen kann das Recht etwas *erlauben* oder aber *verbieten*, ausserdem *Bedingungen* für die Ausübung von Tätigkeiten vorschreiben. Wie sieht das im Hinblick auf Maschinen konkret aus?

Einen klaren *Verbots*-Weg hat die nach dem Zweiten Weltkrieg unter dem Eindruck der Industrialisierung geschriebene Verfassung der Freien Hansestadt Bremen gewählt, die festlegt: «Der Mensch steht höher als Technik und Maschine (...).»¹ Wörtlich genommen würde diese Vorschrift die Übertragung von Entscheidungen auf Maschinen also schlicht ausschliessen. Diese Vorschrift dürfte weltweit einzigartig sein und ist (soweit ersichtlich) in der Rechtsprechung auch noch nie zur Anwendung gekommen. Zahlreicher als die ausdrücklichen Verbote sind im aktuellen Recht aber die impliziten Förderungen neuer Technologien: So verpflichtet etwa die Behindertenrechts-Konvention der Vereinten Nationen (UNO) die Staaten zur *Unterstützung* der Forschung und Entwicklung neuer Technologien, die behinderten Menschen eine geeignete Hilfe sein könnten². Trifft diese Voraussetzung z.B. für automatisierte, mit künstlicher Intelligenz betriebene Geräte zu, wäre deren Einsatz *geboten*. Ausserdem haben die Bürgerinnen und Bürger nach der Europäischen Grundrechtecharta ein Recht auf eine gute Verwaltung³. Was harmlos klingt, bedeutet im Zeitalter von KI aber auch: Falls Algorithmen die Verwaltungsaufgaben besser, schneller und/oder für die Betroffenen einfacher erledigen, *müssten* sie eingesetzt werden.

Bereits aus diesen Beispielen wird ersichtlich, dass ein grundsätzliches Problem, dem sich das Recht ausgesetzt sieht, wenn neue Technologien die Umweltbedingungen verändern (wie z.B. auch beim Aufkommen der Internet- oder später der Blockchain-Infrastruktur), darin besteht, dass einstmals klar ersichtliche und sinnvoll erscheinende Regelungen plötzlich ihren Charakter verändern, nicht mehr denselben Sinn haben wie *vor* der Entwicklung der Technologie. Der regelmässig langsame Anpassungsprozess der Regulierungen ist sachimmanent, und eine Beschleunigung des Gesetzgebungsprozesses würde bedingen, die Kompetenzen vom Parlament auf einzelne Verwaltungseinheiten zu verschieben, was unter Demokratie-Gesichtspunkten nicht ideal

¹ Art. 12 Abs. 1.

² Art. 4 Abs. 1 lit. g.

³ Art. 41.

ist⁴. Einfache Verbote, Erlaubnisse und Definitionen von Bedingungen nützen der Zivilgesellschaft also im Umgang mit neuen Technologien nur so lange, wie letztere einen klar definierten Tätigkeitsbereich (etwa bei der Fertigung in der Industrie) haben, der reglementiert wird. Spätestens wenn es um den Umgang mit Menschen und ihren direkten Schicksalen – ihrer Würde – geht, hört dieser Kreis aber auf. Was also tun?

Menschenwürde und informationelle Selbstbestimmung als Ausgangspunkt

Die Verfassungen vieler kontinentaleuropäischen Länder, so auch die Bundesverfassung (BV) in der Schweiz, enthalten den Grundsatz der Menschenwürde⁵ und der informationellen Selbstbestimmung⁶. Die Menschenwürde ist (auch bei einem Roboter-Einsatz) unantastbar und schützt vor Eingriffen in die physische oder psychische Integrität des Einzelnen. Nach der sogenannten «Objektformel» dürfen Menschen nicht (oder nur unter bestimmten Voraussetzungen) den Entscheidungen von Maschinen unterworfen sein, weil dadurch die Menschenwürde gefährdet werden könnte. In einer solchen, eher abstrakten Betrachtungsweise spielt auch *keine* Rolle, ob automatisierte Maschinen-Entscheidungen qualitativ schlechter oder besser sind als menschliche Entscheidungen.

Weitere Bestimmungen sekundieren diesen Grundsatz: Das zentrale Gleichbehandlungsgebot bzw. *Diskriminierungsverbot*⁷ will verhindern, dass Fehlentscheidungen zulasten von möglicherweise benachteiligten Individuen eintreten. Dieser Grundsatz kommt etwa zum Tragen, wenn die von intelligenten Systemen verwendeten Daten zu Personen und Entscheiden nicht repräsentativ sind oder sie keinen relevanten Bezug zur Auswertung aufweisen (z.B. im Rahmen eines mit einem Roboter geführten Anstellungsgesprächs). Weiter räumt der Anspruch auf rechtliches Gehör dem Einzelnen ein Recht auf Anhörung ein: Behörden und Gerichte sind zu einer Begründung ihrer Entschei-

⁴ Ein in letzter Zeit vermehrt zur Anwendung kommendes Instrument ist die sog. «Experimentier-Regulierung», in der Schweiz schon anfangs der 1980er Jahre mit der Verordnung zu den Lokalradio-Konzessionen getestet und heute z.B. mit dem «Sandbox-Freiraum» für FinTech-Unternehmen erneut fruchtbar gemacht. Experimente müssen aber zeitlich beschränkt sein, stabile Verhältnisse lassen sich damit nicht schaffen.

⁵ Art. 7 BV.

⁶ Art. 13 Abs. 2 BV.

⁷ Art. 8 BV.

dungen verpflichtet⁸ – bei Algorithmen, die Recht sprechen sollen, würde das wohl schwierig.

Das Grundrecht auf *Schutz der Privatsphäre* schützt zudem jeden Umgang mit persönlichen Daten, d.h. mit solchen Daten, die einen gewissen Bezug zu einer bestimmten oder zu einer (weit ausgelegt) bestimmbar Person haben. Staatliche Verfügungen, auch erlassen gestützt auf einen Algorithmus, stellen regelmässig eine Bearbeitung personenbezogener Daten dar und fallen deshalb in den Schutzbereich der informationellen Selbstbestimmung. Moderne Datenschutzgesetze enthalten deshalb konkrete Regelungen bei automatisierten Einzelfallentscheidungen. Ungeachtet dessen bleibt die aber interessanteste Frage im Umgang mit den neuen Maschinen offen: einzig sind wir uns wohl alle dahingehend, dass KI und ihre Handlungen reguliert werden müssen, dass man ihnen Grenzen setzt und «Pflichten» auferlegt. Aber: wollen wir ihnen auch Rechte⁹ zugestehen?

Rechtspersönlichkeit von Maschinen?

Den Anfang genommen hat die Diskussion, ausgelöst von staatlichen Behörden, im Steuerrecht: weil immer mehr Arbeiten von Maschinen und nicht mehr von Menschen erledigt werden, möchten die Steuerämter eine Anknüpfung der Steuererhebung bei den Maschinen ins Auge fassen (sog. Roboter-Steuer). Das Konzept würde darin bestehen, dass Steuerausfälle wegen des Verlustes an Arbeitsstellen dadurch kompensiert werden, dass der Gebrauch von Robotern bei den Unternehmen besteuert wird. Nötig wäre die Entwicklung einer spezifischen *steuerlichen Rechtspersönlichkeit* und eines angemessenen Steuersystems. Von ebenso grosser Bedeutung ist die Frage der Rechtssubjekteigenschaft mit Blick auf Verantwortung und Haftung. Wer soll für den Schaden eintreten, den eine automatisiert handelnde Maschine verursacht? Steht das die Algorithmen verwendete Unternehmen oder der Software-Entwickler in der Pflicht?

⁸ Art. 29 Abs. 2 BV.

⁹ Einer individuellen (und meist auch einer juristischen) Person kommt in jeder Rechtsordnung eine gesetzlich umschriebene Rechtspersönlichkeit zu, d.h. die Person ist ein Rechtssubjekt. Mit dieser Charakterisierung sind regelmässig viele Rechte, aber auch Pflichten verbunden. Zur Frage, ob den Maschinen auch eine Rechtspersönlichkeit bzw. die Rechtssubjekteigenschaft zukommen soll, finden sich in der juristischen Lehre lediglich erste Arbeiten. Immerhin beginnen sich die Spezialistinnen und Spezialisten nun mit dieser auf den ersten Blick «ungewöhnlichen» Themenstellung zu beschäftigen.

Bisher ist die juristische Lehre überwiegend, aber nicht einstimmig, der Auffassung gewesen, dass den Maschinen keine Rechtspersönlichkeit zukommt, was auch mit der Tatsache zusammenhängen mag, dass es sich bei der künstlichen Intelligenz, die algorithmische Entscheidungen ermöglicht, um eine von der Technikgestaltung abhängige «Querschnittsmaterie» handelt. Nicht auszuschliessen ist aber, dass sich diese Meinungsbildung künftig, insbesondere wenn die Quantität von an Maschinen delegierten Entscheidungen (und deren Qualität) zunimmt, ändert.

In ähnlicher Weise wird über eine «Ethik der Algorithmen» bzw. über eine «Roboter-Ethik» debattiert, die danach fragt, ob die autonomen Maschinen selber über irgendeine Form von ethikrelevantem Bewusstsein verfügen. Eine solche Annahme würde bedingen, dass Maschinen auch eine *moralische* Verantwortung für ihr Tun zu übernehmen in der Lage wären; mehrheitlich wird deshalb heute in der Ethik davon ausgegangen, dass anstelle der Schaffung einer Roboter-Ethik vielmehr eine Erweiterung der auf den Menschen bezogenen Ethik sachgerecht wäre.

Verfassungsrechtliche Rahmenbedingungen für algorithmische Entscheidungen

Auch künftig gilt: Automatisierte Einzelentscheidungen von Behörden sind nur zulässig, wenn – wie für jedes staatliche Handeln – eine *gesetzliche Grundlage* vorliegt (Legalitätsprinzip). Für schwerwiegende Grundrechtseingriffe ist sogar ein formelles Gesetz erforderlich. Heute fehlt es verbreitet am Vorliegen dieser Voraussetzung, obwohl in vielen Bereichen der Verwaltung eine Digitalisierung der Verfahren durchaus denkbar wäre. Im Vordergrund stehen Steuerveranlagungen (in Deutschland schon weiter vorangetrieben als in der Schweiz), aber auch in anderen Bereichen (z.B. Zollabrechnungen) ist der Einsatz von Algorithmen durchaus denkbar.

Um das Manko der fehlenden gesetzlichen Grundlage zu beheben, will der Bundesrat mit dem neuen, im Parlament zur Diskussion anstehenden (nachfolgend genauer angesprochenen) Datenschutzgesetz – wenn zwar systematisch nicht überzeugend – Abhilfe schaffen: Im Anhang zum Gesetzesentwurf ist vorgesehen, dass die Festsetzung der Schwerverkehrsabgabe, des Steuerbetrags nach Tabaksteuergesetz und Mineralölsteuergesetz sowie die Veranlagung nach Zollgesetz und nach Biersteuergesetz automatisiert bzw. durch einen Algorithmus erfolgen darf. Weiter sollen die mit der Durchführung der

Unfall- und der Militärversicherung betrauten Organe pauschal ermächtigt werden, automatisierte Verfügungen zu erlassen.

Rechtsstaatlich weiter zu beachten sind die allgemeinen Verfahrensgarantien, insbesondere das rechtliche Gehör. Beim automatisierten Erlass einer Verfügung ist gesetzlich in den meisten Fällen ein vorgängiges Recht des Betroffenen auf Äusserung *nicht* vorgesehen; sofern eine Verfügung zu seinem Nachteil ergehen soll, ist das Recht auf Gehör aber jedenfalls zu gewährleisten. Eine weitere Anforderung betrifft das Recht auf Begründung staatlichen Handelns; ob die Entscheidung eines Algorithmus, der mit künstlicher Intelligenz ausgestattet ist, eine Begründung zu liefern vermag, ist zumindest ungewiss. Spezifische Kausalketten einer individuellen Konstellation wird ein Algorithmus regelmässig auch nicht erkennen können. Aus diesen Gründen dürften automatische Entscheidungen nur, aber immerhin in stark *standardisierten Verfahren* zur Anwendung kommen können.

Automatisierte Einzelentscheidungen und Datenschutz

Dem Datenschutz kommt in der digitalisierten Welt eine ohnehin immer grössere Bedeutung zu. Die Aufrechterhaltung des Schutzes der Personendaten ist auch sicherzustellen, wenn Maschinen die Entscheidungen automatisiert fällen. Die Datenschutz-Grundverordnung der Europäischen Union (DSGVO) und der Entwurf zu einem Schweizer Datenschutzgesetz (E-DSG) anerkennen dieses Anliegen durch Schaffung besonderer rechtlicher Vorgaben.

Schon das französische Datenschutzgesetz von 1978 hatte ein grundsätzliches Verbot automatisierter Einzelentscheide enthalten. Die DSGVO geht in Art. 22 ebenfalls von deren Unzulässigkeit aus, enthält aber Ausweichklauseln: Ausnahmen sind etwa vorgesehen mit Blick auf den Abschluss oder die Erfüllung eines Vertrages und im Falle einer ausdrücklichen Einwilligung der betroffenen Person, die auf einer ausreichenden Information zu beruhen hat. Obwohl in der Schweiz der E-DSG grundsätzlich dem Konzept der DSGVO folgt, wird in Art. 19 nicht ein Verbot von automatisierten Einzelentscheidungen verankert, sondern lediglich eine (relativ detaillierte) *Informationspflicht* und ein Recht auf das sog. menschliche Gehör vorgeschlagen. Ähnlich wie in der EU sind zudem Ausnahmen vorgesehen. Wie sich das Parlament zu diesem Vorschlag des Bundesrates stellt, bleibt abzuwarten.

Ganz allgemein nehmen moderne Datenschutzgesetze (auch in der Schweiz) einen Paradigmenwechsel vor: Die Anliegen des Datenschutzes und der Datensicherheit sollen durch Technologiegestaltung umgesetzt werden. Daten-

schutzfreundliche Voreinstellungen und *Datenschutzfolgeabschätzungen* bezwecken, die persönlichen Daten vor Kenntnisnahme durch unberechtigte Dritte abzuschotten. Je besser die Technikgestaltung funktioniert, umso höher ist das Schutzniveau. Diese allgemeine Überlegung gilt auch für automatisierte Einzelentscheidungen.

Aspekte der Verantwortung und Haftung

Nun zu Verantwortung und Haftung: Das (den umliegenden Ländern ähnliche) Schweizer Haftungsrecht kennt eine Vielzahl allgemeiner Prinzipien, die im Falle von Schadensverursachungen zur Anwendung kommen können: Da wären, *erstens*, die allgemeinen Regeln der Deliktshaftung. Sie verpflichten die Anbieter von Gütern (also auch von Software), die Produkteentwicklungen zu beobachten und regelmässige Risikoabschätzungen vorzunehmen, um festzustellen, welche Fehlentwicklungen dem Technologiedesign inhärent sein könnten, sowie hernach proaktiv Schutzmassnahmen zu treffen, um die Risiken zu mindern. Der anwendbare Haftungsmassstab ist offen und objektiviert; eine Verantwortung entfällt nur dann, wenn unvorhersehbare Systemreaktionen eintreten.

Das besondere Produkthaftungs- und Produktsicherheitsrecht (in der Schweiz und in der EU) begründet, *zweitens*, sogenannte Gefährdungshaftungen, die den Anbieter von Gütern kausal haften lassen, ausser wenn ihm – etwa im Falle eines «Ausreissers» – der Exkulpationsbeweis gelingt. Die schon etwas in die Jahre gekommenen Gefährdungshaftungsgesetze erfassen aber Dienstleistungen grundsätzlich nicht; in der Europäischen Union ist deshalb anerkannt, dass sich eine Anpassung der Regulierungen aufdrängt. Das EU-Parlament hat überdies die Kommission aufgefordert, einen Vorschlag für eine Richtlinie zu zivilrechtlichen Regelungen im Bereich Robotik («Civil Law Rules on Robotics») auszuarbeiten.

Abseits dieser Neuentwicklungen sind, *drittens*, auch konkrete Gesetzesanpassungen, wie dies Deutschland vor knapp 2 Jahren mit einer Revision des Strassenverkehrsrechts getan hat, angezeigt. Nach § 1a des Strassenverkehrsgesetzes (StVG) dürfen zwar Fahrzeuge auf hoch- oder vollautomatisierte Fahrfunktionen ausgelegt werden, doch bleibt die Rolle des Fahrzeugführers

als verantwortliche Person erhalten¹⁰. Wenn also der Computer das Fahrzeug steuert, kommt dem Fahrzeugführer eine veränderte Aufgabe zu, nämlich die unmittelbare Überwachung des Fahrzeuges, weshalb er «wahrnehmungsbereit»¹¹ zu sein hat, um bei entsprechenden Signalen des Fahrzeugs oder bei «offensichtlich» eine Fehlfunktion indizierenden Umständen die Kontrolle über das Fahrzeug wieder zu übernehmen¹². Dieses Konzept einer Verantwortungszuordnung im Mensch-Maschine-Verhältnis wird durch den «Event Data Recorder»¹³ auf der Beweisebene abgesichert. Mit einer solchen Regelung hat Deutschland an sich Neuland betreten, auch wenn einige Details noch klärungsbedürftig sein werden.

Die Gretchenfrage: Was tun?

Die im Titel gestellte Frage, ob Maschinen über Menschen entscheiden dürfen, lässt sich nicht mit einem einfachen «ja» oder «nein» beantworten. Vielmehr geht es vorerst darum, in der Gesellschaft zu diskutieren, welche Technologien im Interesse der Menschen einsetzwürdig sind und welche nicht. Diese Debatte ist breit, nicht allein in der Rechtswissenschaft, zu führen und muss neben ökonomischen Effizienzüberlegungen auch die ethisch-moralische und die soziale Dimension mitumfassen. Im Kontext der Internet Governance wird meist vom sog. «Multistakeholder»-Vorgehen gesprochen.

Der technologische Fortschritt lässt sich nicht aufhalten; es wäre auch wenig sinnvoll, auf den Nutzen und die Chancen des Einsatzes von Algorithmen zu verzichten. Die Geschichte lehrt, dass ein «Hände weg»-Ansatz nicht zukunftsgerichtet ist; so hat etwa das Verbot des Autofahrens im Kanton Graubünden sich nicht sehr lange halten lassen (1900–1925). Das Recht ist vielmehr aufgerufen, Rahmenbedingungen festzulegen, die zu beachten sind, wenn in bestimmten Situationen Maschinen über Menschen entscheiden dürfen.

Grundpfeiler einer solchen rechtlichen Rahmenordnung müssen die Menschenwürde, das Diskriminierungsverbot und der Schutz der Privatsphäre (informationelle Selbstbestimmung) sein. Wenn es gelingt, diese drei verfas-

¹⁰ «Fahrzeugführer ist auch derjenige, der eine hoch- oder vollautomatisierte Fahrfunktion im Sinne des Absatzes 2 aktiviert und zur Fahrzeugsteuerung verwendet, auch wenn er im Rahmen der bestimmungsgemässen Verwendung dieser Funktion das Fahrzeug nicht eigenhändig steuert» (§ 1 Abs. 4 StVG).

¹¹ § 1b Abs. 1 StVG

¹² § 1b Abs. 2 StVG

¹³ § 63a StVG

sungsrechtlichen Prinzipien im Kontext algorithmischer Entscheidungen zu operationalisieren und noch weitere spezifische Gesetzesanpassungen kleineren Ausmasses (z.B. im Haftungsbereich) zu realisieren, verliert die negativ behaftete Vorstellung, dass Maschinen über Menschen entscheiden, erheblich an Brisanz.

Weiterführende Literatur

Beck Susanne, Der rechtliche Status autonomer Maschinen, AJP 2017, 183 ff.

Djeffal Christian, Künstliche Intelligenz in der öffentlichen Verwaltung, Berichte des NEGZ, Nr. 3, Berlin 2018.

Glaser Andreas, Einflüsse der Digitalisierung auf das schweizerische Verwaltungsrecht, SJZ 2018, 181 ff.

Noto La Diega Guido, Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information, JIPITEC 2018, 1 ff.

Oberson Xavier, Taxer les robots?, AJP 2017, 232 ff.

Rechsteiner David, Der Algorithmus verfügt. Verfassungs- und verwaltungsrechtliche Aspekte automatisierter Einzelentscheidungen, Weblaw Jusletter vom 26. November 2018.

Thouvenin Florent/Früh Alfred/George Damian, Datenschutz und automatisierte Entscheidungen, Weblaw Jusletter vom 26. November 2018.

Wildhaber Isabelle, Robotik am Arbeitsplatz: Robo-Kollegen und Robo-Bosse, AJP 2017, 213 ff.

Automatisierte Entscheidungen: Perspektive Grundrechte*

Erstveröffentlichung in: SZW 2020, 18–26

Automated decisions affect the legal system in a wide variety of areas. In a State governed by the rule of law, fundamental rights are of great importance; their validity can lead to a conflict with «results» produced by algorithms. The article examines five areas (employment relationships, biomedical «environment», moderation of online content, financial markets, procedural guarantees) and, based on the insights gained, develops new points of reference in order to give greater emphasis to fundamental rights.

Inhaltsübersicht

I.	Einleitung	48
II.	Betroffene Grundrechte und deren Ziele	49
1.	Betroffene Grundrechte	49
2.	Ziele der einschlägigen Grundrechte	51
III.	Phänomenologisch identifizierbare Spannungsfelder	53
1.	«Human Resources» und Arbeitsplatz	53
2.	Biomedizinische «Umgebung»	54
3.	Moderation von Online-Inhalten	55
4.	Kundenbetreuung in den Finanzmärkten	57
5.	Einhaltung von Verfahrensgarantien	57
IV.	Anknüpfungspunkte für einen nachhaltigen Grundrechtsschutz	60

* Frau Prof. Dr. Christine Kaufmann, Universität Zürich, danke ich für wertvolle Hinweise bei der Ausarbeitung des Textes.

I. Einleitung

Intelligente Maschinen und Systeme, die auf selbstlernender Software aufbauen, ermöglichen in immer weiter gehender Weise das Füllen automatisierter Entscheidungen. Die gebräuchlichen Stichworte lauten Algorithmen, künstliche Intelligenz (Artificial Intelligence) und Machine Learning.¹ Dass die neuen Technologien an sich von Menschen zu fällende Entscheidungen vermehrt an Maschinen bzw. automatisierte Datenauswertungsanlagen übertragen, führt (auch) zu rechtlichen Herausforderungen: Lernende Systeme oder künstliche neuronale Netze ermöglichen algorithmisch herbeigeführte Entscheidungen, die sich nicht ohne weiteres in die geltende Rechtsordnung einfügen vermögen,² weil der Grundsatz, dass das Recht die Leitplanken für das menschliche Verhalten und die gesellschaftliche Organisation festlegt,³ ins Wanken gerät. Zudem tritt die Analyse von datenbasierten Korrelationen anstelle der rechtlichen Kausalitäten in den Vordergrund.⁴

In besonderem Masse im Fokus stehen die Grundrechte; sie konstituieren die Verfassungsordnung eines Gemeinwesens und vermitteln den Individuen spezifische Rechte und Ansprüche. Die nachfolgenden Ausführungen erläutern, welche Grundrechte und welche ihrer spezifischen Schutzbereiche durch automatisierte Entscheidungen von staatlichen Behörden oder privaten Akteuren beeinträchtigt werden könnten. Anzusprechen sind die konkreten kritischen Aspekte automatisierter Entscheidungen sowie der Handlungsbedarf, der sich ergibt, um die Grundrechte sachgemäss zu schützen. Abschliessend wird analysiert, wie den Grundrechten besser Nachdruck verliehen werden kann.

¹ Für einen Überblick zu den Begriffen vgl. Florent Thouvenin/Alfred Früh, Automatisierte Entscheidungen: Grundfragen aus der Perspektive des Privatrechts, SZW 1/2020, 3,6 f., sowie Sebastian Höfer, Algorithmen, maschinelles Lernen und die Grenzen der KI, Jusletter vom 26. November 2018, Rz. 2.

² Rolf H. Weber, Digitalisierung und der Kampf ums Recht, in: A. Dal Molin-Kränzlin/A. M. Schneuwly/J. Stojanovic (Hrsg.), Digitalisierung – Gesellschaft – Recht, Zürich/St. Gallen 2019, 3, 16 (vgl. [Text 12, S. 201 ff.](#); für einen neuen Überblick zum internationalen Umfeld vgl. Rikke Frank Jørgensen (Hrsg.), Human Rights in the Age of Platforms, Cambridge MA 2019; Thomas Wischmayer/Timo Rademacher (Hrsg.), Regulating Artificial Intelligence, Cham 2020.

³ Rolf H. Weber, Dürfen Maschinen über Menschen entscheiden?, Schweizer Monat, Februar 2019, 72 ff. (vgl. [Text 3, S. 37 ff.](#)).

⁴ Vgl. Florent Thouvenin et al., Towards Principled Regulation of Automated Decision-Making (ADM) – A Workshop Report, abrufbar unter <<https://www.it-sl.uzh.ch/dam/jcr:edba006c-8452-4ffc-bbb0-fd8fde62d114/Workshop%20Report%20Lavin.pdf>>, 3.

II. Betroffene Grundrechte und deren Ziele

1. Betroffene Grundrechte

Automatisierte Entscheidungen, basierend auf Algorithmen, tangieren eine grosse Zahl von Grundrechten, weil es um das Verhältnis zwischen Mensch und Maschine geht. Zudem lässt sich auch die – vorliegend nicht diskutierte und bisher in der Regel verneinte – Frage stellen, ob Algorithmen eine neue Art von Rechtspersönlichkeit bzw. E-Persönlichkeit zukommen soll.⁵ Abgesehen davon, dass die Rechtsordnung allgemein die Priorität von Mensch oder Maschine festzulegen vermag,⁶ hat der Gesetzgeber auch konkret anzuordnen, in welchem Ausmass automatisierte staatliche Entscheidungen in den Grundrechtsbereich der Einzelnen einzudringen berechtigt sind.

Im Vordergrund steht die Würde des Menschen, die vor Eingriffen in die physische und psychische Integrität des Einzelnen zu schützen ist (Art. 7 BV). Automatisierte Entscheidungen sind verfassungsrechtlich nicht zu legitimieren, wenn sie den Schutzbestand der Menschenwürde beeinträchtigen.⁷

Weil automatisierte Entscheidungen regelmässig auch Personendaten beinhalten bzw. betreffen, kann die verfassungsrechtlich geschützte informationelle Selbstbestimmung (Art. 13 Abs. 2 BV) bzw. der Datenschutz eine Beeinträchtigung erfahren.⁸ In der Informationsgesellschaft muss der Einzelne die

⁵ Zu dieser vorliegend nicht zu vertiefenden Frage vgl. Weber (Fn. 2), 19 f.; Susanne Beck, Der rechtliche Status autonomer Maschinen, AJP 2017, 183, 186 ff.; Xavier Oberson, *Taxer les robots?*, AJP 2017, 232, 236 ff.; Martin Zobl/Michael Lysakowski, *E-Persönlichkeit für Algorithmen?*, *digma* 2019, 42 ff.

⁶ Die nach dem Zweiten Weltkrieg unter dem Eindruck der Industrialisierung geschriebene Verfassung der Freien Hansestadt Bremen statuiert in Art. 12 Abs. 1 das Prinzip, dass der Mensch höher steht als Technik und Maschine. Demgegenüber verpflichten etwa die Behindertenrechts-Konvention der Vereinten Nationen und Art. 41 der Europäischen Grundrechte-Charta die Staaten dazu, die Forschung und Entwicklung neuer Technologien zu Gunsten der Bevölkerung zu fördern; im Einzelnen dazu Weber (Fn. 2), 16; Christian Djeflal, *Künstliche Intelligenz in der öffentlichen Verwaltung*, Berichte des NEGZ, Nr. 3, Berlin 2018, 17 f.

⁷ Die Zahl der Verweise auf die verfassungsrechtliche Literatur muss im vorliegenden Beitrag knapp bleiben; statt vieler vgl. allgemein Matthias Mahlmann, *Die Garantie der Menschenwürde in der Schweizerischen Bundesverfassung*, AJP 2013, 1307 ff., 1311, und Philip Laternser, *Der Gehalt von Art. 7 BV: zur Begründung der bundesgerichtlichen Menschenwürdekonkretisierung*, Diss. Zürich 2016 sowie spezifisch auch Weber (Fn. 2), 17.

⁸ Ulrich Häfelin/Walter Haller/Helen Keller/Daniela Thurnherr, *Schweizerisches Bundesstaatsrecht*, 9. Aufl. Zürich 2016, § 12, Rz. 380 ff.; Florent Thouvenin/Alfred Früh/Damian

Funktion des zumindest «faktischen Kontrolleurs» seiner Personendaten einnehmen können. Eine Regelung hat dieser Aspekt insbesondere in den datenschutzrechtlichen Anforderungen im Falle automatisierter Einzelentscheidungen gefunden, sei es durch ein Verbot mit Ausweichklauseln (Art. 22 DSGVO), sei es durch Statuierung einer detaillierten Informationspflicht (Art. 19 E-DSG).⁹

Automatisierte Entscheidungen (und allgemein Algorithmen, die automatisch Inhalte generieren oder mit Menschen auf sozialen Netzwerken interagieren¹⁰) vermögen die Kommunikationsgrundrechte (z.B. Meinungsäußerungsfreiheit, Medienfreiheit) zu beeinträchtigen (Art. 16 und 17 BV), wenn die automatisierte Moderation von Online-Inhalten den Einzelnen in seiner informationellen oder kommunikativen Entfaltung einschränkt.¹¹

Von ganz grundsätzlicher Bedeutung ist das Diskriminierungsverbot (Art. 8 Abs. 2 BV). Automatisierte Entscheidungen können erhebliche Diskriminierungsprobleme nach sich ziehen.¹² Algorithmen widerspiegeln nämlich den Bias derjenigen Person(en), der(die) ihn geschaffen hat(haben), mit der Folge, dass unsoziale Entscheidungen zu resultieren vermögen.¹³ Das Risiko, dass es wegen der Automatisierung zu Fehlentwicklungen zulasten von potentiell benachteiligten Personen kommt, beruht nicht selten auf der Tatsache, dass die von intelligenten Systemen verwendeten Daten ggf. nicht repräsentativ sind oder keinen relevanten Bezug zur konkreten Datenauswertung aufweisen.¹⁴

George, Datenschutz und automatisierte Entscheidungen, Jusletter vom 26. November 2018, Rz. 19 ff. m.w.V.

⁹ Thouvenin/Früh/George (Fn. 8), Rz. 23 ff.

¹⁰ Franziska Oehmer, Meinungsfreiheit für Social Bots?, Jusletter vom 29. April 2019, Rz. 5.

¹¹ Häfelin/Haller/Keller/Thurnherr (Fn. 8), § 15, Rz. 447 ff.; Oehmer (Fn. 10), Rz. 16 ff.; Filippo A. Raso/Hannah Hilligoss/Vivek Krishnamurthy/Christopher Bavitz/Levin Kim, Artificial Intelligence & Human Rights: Opportunities & Risks, Berkman Klein Center for Internet and Society at Harvard University, Research Publication No. 2018-6, September 25, 2018, 37 ff.

¹² Weber (Fn. 2), 17; Frederik Zuiderveen Borgesius, Discrimination, artificial intelligence, and algorithmic decisionmaking, Study published by the Council of Europe, Strasbourg 2018, 10 ff.; aus der allgemeinen verfassungsrechtlichen Literatur vgl. Bernhard Waldmann, Das Diskriminierungsverbot von Art. 8 Abs. 2 BV als besonderer Gleichheitssatz, Bern 2003.

¹³ Bericht der Expertengruppe zur Zukunft der Datenbearbeitung und Datensicherheit, 17. August 2018, 38 f.; Council of Europe, Committee of Experts on Internet Intermediaries, Study on the Human Rights Dimension of Automated Data Processing Techniques (in particular Algorithms) and Possible Regulatory Implications, März 2018, 27.

¹⁴ Weber (Fn. 2), 17.

Wenn in staatlichen Verfahren auch Algorithmen eingesetzt werden, ist dem Recht auf Einhaltung der verfassungsrechtlichen Verfahrensgarantien besondere Beachtung zu schenken (Art. 29 BV). Mehrere Grundrechte (Anspruch auf rechtliches Gehör, Recht auf Begründung staatlichen Handelns) sind beim Einsatz von Algorithmen in den Prozessen der Entscheidungsfindung nicht leicht zu verwirklichen.¹⁵ Verfassungsrechtlich weiter in Betracht fallen Prinzipien im Strafverfahren, deren Durchsetzung zum Teil von der konkreten Ausgestaltung der gesetzlichen Rahmenbedingungen abhängt.

Abgesehen von den traditionellen, international anerkannten Menschenrechten, zu denen auch die bisher nicht erwähnte Religionsfreiheit gehört, haben in den letzten Jahrzehnten weitere «grundrechtsähnliche Rechtspositionen» zu Diskussionen geführt. Als mögliche Beispiele¹⁶ genannt werden – ohne dass der Grundrechtscharakter immer zweifelsfrei gegeben ist – etwa das Recht auf Eigentum und auf Berufsausübung, das Recht auf Personenfreizügigkeit und/oder auf Asyl (z.B. bei Vornahme algorithmischer Entscheidungen im Migrationsbereich), die Freiheit des häuslichen Zusammenlebens (z.B. im Kontext des algorithmisch beurteilten Familiennachzuges), das Recht auf Nationalitätenwechsel, das Recht auf einen angemessenen Lebensstandard und auf soziale Sicherheit (im Einzelnen sehr umstritten), das Recht auf Teilnahme an der politischen Willensbildung, das Recht auf Bildung, die Unschuldsvermutung in Strafverfahren, die Abwehr einer ungerechtfertigten Verhaftung und das Recht auf gerichtliche Wiedergutmachung.

2. Ziele der einschlägigen Grundrechte

Wie alle Grundrechte bezwecken die von automatisierten Entscheidungen potentiell betroffenen (vorerwähnten) Grundrechte vornehmlich die individuelle Selbstentfaltung des Einzelnen. Der Grundsatz, dass die verfassungsrechtlich verankerten Grundrechte zu beachten sind, gilt allgemein für Algorithmen verwendende Akteure, also für den Staat, z.B. beim Erlass von Verfügungen (Art. 35 Abs. 1 BV), für Private, die eine staatliche Aufgabe wahrnehmen (Art. 35 Abs. 2 BV), sowie darüber hinausgehend auch für Private, soweit die Grundrechte eine indirekte Drittwirkung aufweisen (Art. 35 Abs. 3 BV). Die Analyse des Ausmasses einer solchen indirekten Drittwirkung hängt von den konkreten Umständen ab und bedarf künftig auch noch einer weitergehenden wissenschaftlichen Durchdringung mittels vertiefender Untersuchungen.

¹⁵ Im Einzelnen dazu hinten Teil III.5.

¹⁶ Einen umfassenden Überblick zu möglicherweise betroffenen weiteren Grundrechten geben Raso et al. (Fn. 11), 6.

Im Zentrum der Grundrehtediskussion steht die Verwirklichung der normativen Nichtdiskriminierung. Deren Herausforderung besteht darin, dass ein Problem der «ungleichen» Betroffenheit durch automatisierte Entscheidungen einzutreten vermag. Eine direkte Nichtdiskriminierung genügt also nicht zwingend, um eine indirekte Diskriminierung auszuschliessen.¹⁷

Von grosser Bedeutung mit Bezug auf automatisierte Entscheidungen ist weiter, ein sachgerechtes Umfeld für die Prinzipien der Transparenz und der Vorhersehbarkeit zu schaffen. Auch und gerade wenn Algorithmen entscheiden, muss für die Betroffenen nachvollziehbar sein, wie eine Entscheidung gefallen ist. Zudem sollten die «Resultate» im Rahmen einer gewissen Bandbreite den vernünftigen Erwartungen entsprechen.¹⁸

Wer eine Aufgabe übernimmt und dabei auch für andere Personen entscheidet, muss Rechenschaft ablegen und die Verantwortung für sein Handeln tragen. Dieser Pflicht unterliegen, wie eingangs erwähnt, alle Akteure in der jeweils betroffenen Weise und Intensität. In neuerer Zeit hat sich dafür der Begriff der «Accountability» durchgesetzt, zuerst im Kontext der Internet Governance, später auch mit anderen Bezügen.¹⁹ Accountability bedeutet, dass die handelnde Person einsteht für ihre Aktivitäten und bei Auftreten von Problemen auch die entsprechende Verantwortung übernimmt.²⁰

Automatisierte Entscheidungen beruhen auf technischen Abläufen. Die Datensicherheit mit Bezug auf die eingesetzten Algorithmen muss deshalb gewährleistet sein. Vertrauen (Trust) und Vertraulichkeit sind zentrale Elemente automatisierter Entscheidungen.²¹ Nur wenn die Betroffenen darauf vertrauen können, dass die technologischen Vorgänge transparent und in sachlich zutreffender Weise ablaufen, vermag Vertrauen zu entstehen.²²

¹⁷ Vgl. BGE 139 I 169 E. 7.2.1; Beatrice Weber-Dürer, in: D. Thürer/J.-F. Aubert/J. P. Müller (Hrsg.), *Verfassungsrecht der Schweiz*, Zürich 2001, 41, Rz. 24 f.; Häfelin/Haller/Keller/Thurnherr (Fn. 8), § 24, Rz. 776a.

¹⁸ Zu den Aspekten der Transparenz und Nachvollziehbarkeit eingehender Rolf H. Weber, *Socio-ethical values and legal rules on automated platforms: The quest for a symbiotic relationship*, 36 CLSR 2020, 105308, 1, 6 f.

¹⁹ Rolf H. Weber, *Shaping Internet Governance: Regulatory Challenges*, Zürich/Basel/Genf 2009, 137 ff. m.w.V.; die Nichtbeachtung der Accountability-Grundsätze kann zur Haftung führen (dazu der Beitrag von Clara-Ann Gordon/Tanja Lutz, *Haftung für automatisierte Entscheidungen – Herausforderungen in der Praxis*, SZW 1/2020, 53, 56 ff.).

²⁰ Weber (Fn. 18), 7.; Council of Europe Committee (Fn. 13), 39.

²¹ Onora O'Neill, *A Question of Trust*, Cambridge 2002, 61 ff.

²² Onora O'Neill, *Justice, Trust and Accountability*, Cambridge 2005.

III. Phänomenologisch identifizierbare Spannungsfelder

Die Zahl möglicher Herausforderungen seitens der Grundrechte für automatisierte Entscheidungen ist gross; anschliessend lassen sich nur einige wichtige Bereiche, die phänomenologisch feststellbare wesentliche Spannungsfelder betreffen, adressieren.²³

1. «Human Resources» und Arbeitsplatz

Private Gesellschaften und der Staat können in der arbeitsteiligen Wirtschaft und Gesellschaft nicht ohne Arbeitnehmer/innen ihre Aktivitäten entfalten. Bei deren Auswahl sind nicht nur die Fähigkeiten und Kenntnisse zu prüfen, sondern viele andere Faktoren spielen auch eine Rolle, und oft ist ein Entscheid zwischen einer grossen Zahl von Bewerbern zu fällen. Algorithmen vermögen den Auswahlprozess nicht nur effizienter zu gestalten, sondern auch zielgenauer in grossen Datenbeständen (Big Data) die (zumindest vermeintlich) wichtigsten Kriterien herauszufiltern. Die Bewerbungsprozesse werden deshalb vermehrt automatisiert.²⁴

Die Sammlung und automatisierte Verwertung von Daten über Arbeitnehmer/innen während der Zeit der Beschäftigung ist ein weiterer wichtiger Bereich für den Einsatz digitaler Technologien. Anstelle einer Vielzahl von Notizblättern, die über die Zeit hinweg angelegt werden, vermögen die Algorithmen in der digitalisierten Welt die relevanten Informationen schnell zum Gebrauch bereitzustellen, ggf. können die Algorithmen den Arbeitnehmern/innen sogar konkrete Weisungen erteilen.²⁵

Algorithmen lassen sich weiter im Kontext der durch Sozialschutzvorschriften konkretisierten Bereiche der Beschäftigung (z.B. Berufskrankheiten, Betriebsunfälle) einsetzen; die Algorithmen bzw. die automatisierten Entscheidungen in diesem Kontext führen zu neuen Herausforderungen.²⁶

²³ Eine umfassende Aufarbeitung der Fragestellungen für das US-amerikanische Recht haben Raso et al. (Fn. 11), 17 ff. vorgenommen.

²⁴ Vgl. auch Isabelle Wildhaber, Robotik am Arbeitsplatz: Robo-Kollegen und Robo-Bosse, AJP 2017, 213, 214 ff.; Council of Europe Committee (Fn. 13), 29; zu den automatisierten Entscheidungen am Arbeitsplatz vgl. auch Matthias Glatthaar, Robot Recruiting, Datenschutzrechtliche Aspekte einer Automatisierung von Rekrutierungsentscheiden, SZW 1/2020, 43, 45 ff.

²⁵ Wildhaber (Fn. 24), 218 ff.

²⁶ Wildhaber (Fn. 24), 221 f.; zum Sozialversicherungsrecht im Besonderen vgl. Kurt Pärli, Sozialversicherungsrecht im Zeitalter der Robotik, AJP 2017, 225 ff.

Als potentiell betroffene Grundrechte, die beim Einsatz von Algorithmen für automatisierte Entscheidungen im Bereich der «Human Resources» zu beachten sind, fallen insbesondere in Betracht:²⁷

- Nichtdiskriminierungsgebot (z.B. mit Bezug auf Geschlecht oder Gesundheitszustand);
- Informationelle Selbstbestimmung (z.B. spezifischer Schutz von besonders schützenswerten Personendaten und Persönlichkeitsprofilen);
- Meinungsäusserungs- und Informationsfreiheit sowie Versammlungsfreiheit (z.B. zur Durchsetzung der Rechte von Arbeitnehmern/innen);
- Recht auf Fürsorge durch den Arbeitgeber (z.B. Einhaltung von Schutzvorschriften durch Algorithmen in Verfahrensabläufen).

Im arbeitsrechtlichen Kontext ist die Rechtsprechung relativ weit entwickelt; selbst wenn automatisierte Entscheidungen noch kaum thematisiert werden, lassen sich doch durch Analogieschlüsse aus den Entscheiden zu den Aspekten von digitalisierten Arbeitsverhältnissen sinnvolle Problemlösungen finden.²⁸

2. Biomedizinische «Umgebung»

Der Einsatz von Robotern, die auf Algorithmen gestützte automatisierte Entscheidungen fällen, nimmt schon heute in den Bereichen der Diagnostik und der Patientenbetreuung in vielen Ländern erheblich zu. Grundsätzlich verbessern solche Hilfsmittel die Arbeiten der Medizinalberufe. Algorithmen vermögen also die biomedizinische «Umgebung» zugunsten der Patienten zu stärken.²⁹

Ähnliche Überlegungen gelten für die Betreuung älterer und behinderter Menschen; wenn automatisiert gewisse Vorgänge zum Laufen gebracht werden, besteht keine Abhängigkeit mehr von möglicherweise unerwarteten Reaktio-

²⁷ Vgl. auch *Raso et al.* (Fn. 11), 42 ff.

²⁸ Vgl. zur Thematik auch *Isabelle Wildhaber/Silvio Hänsenberger*, Internet am Arbeitsplatz, ZBJV 2016, 307 ff.; *Isabelle Wildhaber/Melinda F. Lohmann/Gabriel Kasper*, Diskriminierung durch Algorithmen – Überlegungen zum schweizerischen Recht am Beispiel prädiktiver Analytik am Arbeitsplatz, ZSR 2019 1459 ff.

²⁹ Zu den Einsatzmöglichkeiten von Robotern im Gesundheits- und Betreuungsbereich vgl. *Tim Hagemann*, Digitalisierung und technische Assistenz im Sozial- und Gesundheitswesen, in: T. Hagemann (Hrsg.), *Gestaltung des Sozial- und Gesundheitswesens im Zeitalter von Digitalisierung und technischer Assistenz*, Baden-Baden 2017, 9 ff.

nen solcher Personen oder etwaigen Fehleinschätzungen des Betreuungspersonals.

Als potentiell betroffene Grundrechte in diesem Kontext sind zu beachten:³⁰

- Recht auf Leben und freie Bewegung der von automatisierten Entscheidungen betroffenen Person (z.B. mit Bezug auf die persönliche Entscheidung, gewisse Behandlungen nicht ausgeführt haben zu wollen);
- Recht auf informationelle Selbstbestimmung (z.B. hinsichtlich einer algorithmisch determinierten Verarbeitung von – oft sensiblen – Personendaten);
- Recht auf angemessene Beschäftigung und einen angemessenen Lebensstandard (z.B. durch Algorithmen verursachte Beschränkungen mit Bezug auf die Ausübung gewisser, an sich möglicher Tätigkeiten);
- Recht auf Erziehung und Bildung (z.B. mit Blick auf einen durch automatisierte Programme in Pflegeverhältnissen herbeigeführten Ausschluss kranker Menschen von Weiterbildungen).

Erste grundrechtliche Überlegungen sind im Rahmen der Diskussion zu E-Health erfolgt; eine Vertiefung der rechtlichen Fragestellungen erweist sich aber als notwendig.

3. Moderation von Online-Inhalten

Die «Verwaltung» von Informationen ist angesichts des heutigen Datenüberflusses zu einer schwierigen Aufgabe geworden.³¹ Algorithmen vermögen insofern wichtige Unterstützungshilfen zu bieten und durch automatisierte Entscheidungen den Empfänger insbesondere mit denjenigen Informationen zu beliefern, die für ihn als geeignet erscheinen.³²

Solche Vorteile aus der Moderation von Online-Inhalten sind aber gekoppelt mit dem Risiko, dass die Auswahl der Daten gestützt auf vorgegebene ein-

³⁰ Vgl. auch Raso et al. (Fn. 11), 32 ff.

³¹ Vgl. schon früh Rolf H. Weber, *Kassandra oder Wissensbroker – Dilemma im «Global Village»*, in: J. Becker/R. M. Hilty/J.-F. Stöckli/Th. Würtenberger (Hrsg.), *Recht im Wandel seines sozialen und technologischen Umfeldes*, Festschrift für Manfred Rehbinder, München 2002, 405 ff.

³² Die Zuteilung der Information an diejenigen Adressaten, die über die Information verfügen sollten, ist eine zentrale Aufgabe des Informationsrechts (vgl. Rolf H. Weber, *Allgemeiner Überblick*, in: R.H. Weber (Hrsg.), *Informations- und Kommunikationsrecht*, Schweizerisches Bundesverwaltungsrecht, Bd. V/1, 2. Aufl. Basel 2003, Rz. 71 ff.

seitige Kriterien des Entwicklers der entsprechenden Software erfolgt.³³ Automatisierte Online-Moderation vermag deshalb zur Beeinflussung der Informations-Rezipienten und damit zur Meinungsmanipulation beizutragen.³⁴ Anwendbare Standards (Selbstregulierungen) haben sich in der Realität noch kaum durchgesetzt.³⁵

Als Beispiele von potentiell betroffenen Grundrechten lassen sich nennen:³⁶

- Kommunikationsgrundrechte (z.B. faktische Einschränkung der Meinungsäusserungs- oder Medienfreiheit);
- Nichtdiskriminierungsgebot (z.B. ungerechtfertigter Ausschluss gewisser Bevölkerungsgruppen von der medialen Mitwirkung nach nicht transparenten Kriterien);
- Recht auf Leben und persönliche Freiheit (z.B. Risiko der Gefährdung durch Informationsfilterung gemäss politischen Vorgaben);
- Recht auf informationelle Selbstbestimmung (z.B. Verbreitung unzutreffender oder irreführender Online-Inhalte wegen fehlerhafter Verwendung von Personendaten).

Im Zusammenhang mit den Online-Inhalten stellt sich noch die besondere Frage der Haftung der Provider, die in der Europäischen Union eine spezifische Regelung erfahren hat (Art. 12–15 der E-Commerce-Richtlinie von 2000³⁷), in der Schweiz indessen auf den allgemein anwendbaren Rechtsgrundlagen bei Verletzung von Individualrechten sowie auf Selbstregulierungen der betroffenen Kreise beruht.³⁸

³³ Zur Diskussion, ob den Social Bots ggf. Rechtspersönlichkeit zukommt und sie somit Träger des Grundrechts der Meinungsfreiheit sein können, vgl. *Oehmer* (Fn. 10), Rz. 18 ff. und Rz. 27.

³⁴ Vgl. auch *Oehmer* (Fn. 10), Rz. 16 ff.; *Patricia Egli/David Rechsteiner*, Social Bots und Meinungsbildung in der Demokratie, *AJP* 2017, 249 ff.

³⁵ Vgl. auch *Raso et al.* (Fn. 11), 38 f.

³⁶ *Raso et al.* (Fn. 11), 37 ff.

³⁷ Richtlinie 2000/31/EG des Europäischen Parlaments und des Rates vom 8. Juni 2000 über bestimmte rechtliche Aspekte der Dienste der Informationsgesellschaft, insbesondere des elektronischen Geschäftsverkehrs, im Binnenmarkt («Richtlinie über den elektronischen Geschäftsverkehr»), *ABI* 2000 L 178 vom 17. Juli 2000.

³⁸ Vgl. Bericht des Bundesrates vom 11. Dezember 2015, Die zivilrechtliche Verantwortlichkeit von Providern, 27 ff.

4. Kundenbetreuung in den Finanzmärkten

Im Vordergrund stehen in den Finanzmärkten mannigfaltige Fragen des Zugangs zu Dienstleistungen. Der Einsatz von Algorithmen und von automatisierten Entscheidungsabläufen eignet sich besonders beim Vorgang der Kreditvergabe von Banken an Kunden; Informationen lassen sich durch Credit Scoring zu einem detaillierten Bild über den Kunden verdichten.³⁹ Kreditauskunfteien, die über Informationen zur Kreditwürdigkeit einzelner Personen verfügen, gibt es zwar seit Jahrzehnten, die digitalen und stark verfeinerten Credit Scoring Verfahren haben indessen das Potential, viel stärker in individuelle Grundrechte einzugreifen.

Ähnlich wie in der Arbeitswelt sind Algorithmen auch zur Sammlung von Daten über Kunden während der Laufzeit einer Kontobeziehung geeignet. Entsprechende Datenanalysen erweisen sich insbesondere bei stark auf digitalen Komponenten beruhenden Bankvertragsbeziehungen als bedeutsam, etwa im Falle von Robo-Advice.⁴⁰

Besondere Bedeutung haben Algorithmen und automatisierte Entscheidungen ebenso im Kontext von RegTech bzw. dem Informationsaustausch zwischen Finanzmarktintermediären und Aufsichtsbehörden.⁴¹ In aller Regel dürften diese Informationen aber keinen sehr grundrechtssensitiven Charakter aufweisen.

Als potentiell betroffene Grundrechte sind zu beachten:⁴²

- Nichtdiskriminierungsgebot und Recht auf Gleichbehandlung (z.B. Risiko der Marginalisierung einzelner Bevölkerungsgruppen oder des Ausschlusses bestimmter Personen von gewissen Finanzgeschäften oder der Implementierung einer automatisierten Kontosperrung);
- Recht auf informationelle Selbstbestimmung (z.B. Problem der übersteigerten Verarbeitung von Personendaten im Bankgeschäft);

³⁹ Raso et al. (Fn. 11), 27 f.; Markus Winkler, Credit Scoring, AML Software & Risk Profiling: Automatisierte Entscheidungen im Rahmen von Finanzdienstleistungen, SZW 1/2020, 62, 65 ff.; vgl. auch The Economist, Lending and data, Numbers gain, July 6th 2019, 48 f.

⁴⁰ Vgl. dazu Rolf H. Weber/Rainer Baisch, Regulierung von Robo-Advice, AJP 2016, 195 ff.; allgemein zum Risk Profiling Winkler (Fn. 39), Ziff. 5.

⁴¹ Rolf H. Weber, Regtech as a new legal challenge, Journal of Financial Transformation 2017, 10 ff.

⁴² Raso et al. (Fn. 11), 26 ff.; vgl. auch Council of Europe Committee (Fn. 13), 26 f.

- Kommunikationsgrundrechte und Vereinigungsfreiheit (z.B. Einschränkung in der Ausübung der freien Meinungsäusserung mit Bezug auf Vorgänge in der Bank).

Die für die Finanzmärkte angesprochenen Probleme des Credit Scoring treten auch in weiteren Bereichen ähnlich auf. Das Scoring spielt z.B. eine immer grössere Rolle im Rahmen der Ausbildung (z.B. automatisierte Beurteilung von schriftlichen Arbeiten⁴³), was Konsequenzen für das spätere Berufsleben hat. In den Medien intensiv diskutiert wird weiter das chinesische Social Scoring, das zu einer Überwachung und Evaluierung verschiedenster Lebensbereiche führt.⁴⁴

5. Einhaltung von Verfahrensgarantien

Für das staatliche Handeln enthält die Bundesverfassung eine Reihe von Prinzipien, die auch im Falle algorithmisch herbeigeführter automatisierter Entscheidungen zu beachten sind. Im Vordergrund steht das Legalitätsprinzip; automatisierte Einzelentscheidungen sind nur zulässig, wenn eine gesetzliche Grundlage vorliegt, im Falle eines schwerwiegenden Grundrechtseingriffs sogar in der Gestalt eines formellen Gesetzes (Art. 36 Abs. 1 BV).⁴⁵ Im heutigen Recht ist diese Voraussetzung nur in beschränktem Masse erfüllt. Der Entwurf zum neuen Datenschutzgesetz (E-DSG) will nun einzelne Lücken des Legalitätsprinzips füllen.⁴⁶ Im Anhang zum E-DSG ist vorgesehen, dass die Festsetzung der Schwerverkehrsabgabe, des Steuerbetrags nach Tabaksteuergesetz und Mineralsteuergesetz sowie die Veranlagung nach Zollgesetz und nach Biersteuergesetz automatisiert bzw. durch einen Algorithmus erfolgen darf. Weiter sollen auch die mit der Durchführung der Unfalls- und der Militärversicherung betrauten Organe grundsätzlich die Ermächtigung erhalten, automatisierte Verfügungen zu erlassen.⁴⁷

⁴³ Für eine Darstellung der Verhältnisse in den USA vgl. *Raso et al.* (Fn. 11), 47 ff.

⁴⁴ Vgl. Bericht Expertenkommission (Fn. 13), 169 f.

⁴⁵ Im vorliegenden Beitrag, der sich vornehmlich mit Grundrechtsfragen beschäftigt, kann auf die Einzelheiten der staats- und verwaltungsrechtlichen Voraussetzungen für automatisierte Einzelentscheidungen nicht eingegangen werden; einen guten Überblick gibt *David Rechsteiner*, *Der Algorithmus verfügt. Verfassungs- und verwaltungsrechtliche Aspekte automatisierter Einzelentscheidungen*, Jusletter vom 27. November 2018, Rz. 14 ff.

⁴⁶ Botschaft zum Bundesgesetz über die Totalrevision des Bundesgesetzes über den Datenschutz und die Änderung weiterer Erlasse zum Datenschutz vom 15. September 2017, BBl. 2017, 6941, 7138 ff.

⁴⁷ Vgl. auch *Weber* (Fn. 2), 17.

Für staatliches Handeln sind weiter die allgemeinen Verfahrensgarantien zu beachten, insbesondere der Anspruch auf rechtliches Gehör (Art. 29 Abs. 2 BV).⁴⁸ Beim automatisierten Erlass einer Verfügung ist es technisch nicht immer einfach, den Betroffenen im Laufe eines Verfahrens rechtzeitig miteinzubeziehen; sofern eine Verfügung aber zu seinem Nachteil erfolgen soll, ist das Recht auf Gehör vorgängig zu gewähren.⁴⁹ Ersetzen Algorithmen bzw. auf künstlicher Intelligenz beruhende Programme den Richter, stellt sich weiter das Problem, dass zumindest in einer Anfangsphase ausreichende «Trainingsdaten» fehlen.⁵⁰

Eine weitere Anforderung von Art. 29 Abs. 2 BV betrifft das Recht auf Begründung staatlichen Handelns. Ob eine automatisierte Entscheidung gestützt auf Algorithmen eine rechtsgenügende Begründung, welche die Nachvollziehbarkeit der urteilsrelevanten Überlegungen ermöglicht, zu liefern vermag, ist zumindest ungewiss, weil Algorithmen in der Regel nicht alle Kausalitäten zu erkennen vermögen.⁵¹ In einem Beschwerdeverfahren müsste die betroffene Behörde auch begründen können, wie ein Algorithmus programmiert wurde und was er konkret gemacht hat. Aus diesem Grunde eignen sich automatisierte Entscheidungen insbesondere in stark standardisierten Verfahren, die kaum auf individuelle Merkmale der involvierten Parteien Bezug nehmen.⁵²

Eine besondere Bedeutung kommt den Algorithmen bei der Risikoeinschätzung in Strafverfahren zu. In neuerer Zeit sind Polizei und Strafverfolgungsbehörden vermehrt dazu übergegangen, «gefährliche» Personen anhand von Risikoeinschätzungen zu «überwachen» («Predictive Policing»)⁵³. Das Ziel besteht insbesondere darin, die Risikoprävention zu verbessern und die Rückfallgefahr zu senken.⁵⁴ Diese Vorgehensweise berührt eine Reihe der eingangs aufgelisteten Grundrechte, so etwa das Nichtdiskriminierungsgebot, den Gleichbehandlungsgrundsatz, das Recht auf Leben und persönliche Freiheit, das Recht auf informationelle Selbstbestimmung sowie die rechtstaatlichen

⁴⁸ Nadja Braun Binder, Künstliche Intelligenz und automatisierte Entscheidungen in der öffentlichen Verwaltung, SJZ 2019, S. 467, 476; vgl. auch Nadia Braun Binder, Automatisierte Entscheidungen: Perspektive Datenschutzrecht und öffentliche Verwaltung, SZW 1/2020, 27, 28 f., 32 f.

⁴⁹ Weber (Fn. 2), 18; Rechsteiner (Fn. 45), Rz. 19 ff. m.w.V.

⁵⁰ Vgl. auch Stephan Dreyer/Johannes Schmees, Künstliche Intelligenz als Richter?, CR 11/2019, 758 ff.

⁵¹ Braun Binder, Künstliche Intelligenz (Fn. 48), 472 f. m.w.V.

⁵² Weber (Fn. 2), 18.

⁵³ Braun Binder, Künstliche Intelligenz (Fn. 48), 470; vgl. auch Raso et al. (Fn. 11), 22 f.

⁵⁴ Braun Binder, Künstliche Intelligenz (Fn. 48), 471 m.w.V.

Verfahrensgarantien und den Schutz vor nicht gerechtfertigter Verhaftung.⁵⁵ Eine vertiefte Diskussion der sich stellenden Herausforderungen ist angebracht.

IV. Anknüpfungspunkte für einen nachhaltigen Grundrechtsschutz

Die vorangegangenen Analysen einzelner Bereiche, in denen automatisierte Entscheidungen zum Einsatz kommen können, mit Blick auf deren Vereinbarkeit mit grundrechtlichen Anforderungen haben gezeigt, dass die Thematik vertiefter Untersuchungen bedarf und dass die heutigen Einschätzungen erst vorläufige Beurteilungen erlauben. Die Vielfalt der potentiell anwendbaren Grundrechte erschwert auch die Herausbildung von allgemein gültigen Anwendungsprinzipien. Einzelne Aspekte rechtlicher Weiterentwicklungen haben sich indessen bereits herauskristallisiert:

- Der Interaktion zwischen Mensch und Maschine ist regulatorisch grössere Beachtung zu schenken. Die Festlegung von Handlungsstandards hat einen Beitrag zur Verstärkung einer werteorientierten «Leadership» zu leisten.⁵⁶ Erforderlich ist somit ein erweiterter interdisziplinärer Ansatz der Regelbildung, der ethische und sozio-ökonomische Elemente mitbeinhaltet.⁵⁷
- Die Herausbildung von (rechtlich wirksamen) Standardisierungen kann nicht mit einem «Top-Down»-, sondern nur mit einem «Bottom-Up»-Ansatz erfolgreich sein.⁵⁸ Das Prinzip der Multistakeholder-Beteiligung, wie es etwa im Kontext der Internet Governance und der Klimaverhandlungen schon erprobt worden ist, erweist sich auch mit Blick auf Algorithmen und automatisierte Entscheidungen als sachgerecht.⁵⁹
- Von besonderer Bedeutung ist die Sicherheit und Robustheit der Systeme; wenn es Dritten gelingt, informationstechnologische Eingriffe in

⁵⁵ Vgl. auch den Überblick zum US-amerikanischen Recht bei *Raso et al.* (Fn. 11), 20 ff.

⁵⁶ Vgl. auch *Mathias Risse*, Human Rights and Artificial Intelligence: An Urgently Needed Agenda, Harvard Kennedy School, Carr Center for Human Rights Policy, Discussion Paper CCDP 2018-002, April 2018, 8 ff.

⁵⁷ *Weber* (Fn. 18), 2 f.; *Zuiderveen Borgesius* (Fn. 12), 10 ff.

⁵⁸ *Weber* (Fn. 18), 4 f. m.w.V.

⁵⁹ *Rolf H. Weber*, Legal foundations of multistakeholder decision-making, ZSR 2016 I 247, 249 f.

algorithmisch determinierte Verfahrensabläufe zu realisieren, droht das ganze Konzept automatisierter Entscheidungen einzustürzen.⁶⁰

- Unabhängige Experten-Panels, welche die Systeme von automatisierten Entscheidungen unter verschiedensten Aspekten (z.B. Sicherheit, Nachvollziehbarkeit, Einhaltung von ethischen Grundsätzen) testen und überprüfen, können einen wertvollen Beitrag zu einer verbesserten Zuverlässigkeit und Betriebssicherheit leisten.⁶¹
- Im Rahmen eines breit verstandenen Risikomanagement-Konzepts hat sich in der letzten Zeit das Instrument des «Impact Assessment» entwickelt. Der Staat und private Unternehmen, die eine öffentliche Aufgabe erfüllen, sind ohnehin gehalten, die Grundrechte zu beachten (Art. 35 Abs. 1 und 2 BV). Zudem ist heute regelmässig eine Rechtsfolgeabschätzung beim Erlass neuer Gesetze vorzunehmen. Für Private gilt das Prinzip der indirekten Drittwirkung, die auf die Corporate Social Responsibility (CSR) ausstrahlt, oft auch konkretisiert durch gesetzliche Vorgaben. So ist schon seit Jahrzehnten in der Rechtsordnung die Umweltverträglichkeitsprüfung bekannt.⁶² In den letzten Jahren hat sich im Datenschutzrecht das Konzept des «Data Protection Impact Assessment» (Datenschutzfolgeabschätzung) durchgesetzt, wenn Risiken für die Datensicherheit nicht auszuschliessen sind.⁶³ Dieses Konzept hat in weiteren Rechtsbereichen zwischenzeitlich auch Anwendung gefunden, insbesondere in der Form des «Human Rights Impact Assessment» (HRIA).⁶⁴ Entsprechende Bemühungen im Menschenrechtsbereich, die zum Teil auch die Begrifflichkeit der Due Diligence verwenden (z.B. in den Richtlinien der OECD⁶⁵) und die zur Entwicklung bzw. Ausarbeitung von zielgerichteten «HRIA-Typen» beitragen, verdienen im Kontext der CSR-Diskussionen grössere Unterstützung.

⁶⁰ Eingehender Weber (Fn. 18), 7 f. m.w.V.

⁶¹ Council of Europe, Report on Artificial Intelligence, T-PD 56 (2018)09 Ref. Strasbourg, 3 December 2018, 15; *Alessandro Mantelero*, Regulating Big Data. The Guidelines of the Council of Europe in the Context of the European Data Protection Framework, CLSR 2017, 584 ff.

⁶² Vgl. Art. 10a ff. USG (SR 814.01) und die Verordnung über die Umweltverträglichkeitsprüfung (SR 814.011).

⁶³ Vgl. Art. 35 DSGVO; Art. 20 E-DSG.

⁶⁴ Vgl. *Alessandro Mantelero*, AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment, CLSR 2018, 754 ff.

⁶⁵ Vgl. z.B. die OECD Due Diligence Guidance for Responsible Business Conduct, Paris 2018.

Gesamthaft lässt sich deshalb festhalten, dass im Bereich der Betroffenheit von Grundrechten durch automatisierte Entscheidungen ein nicht unerheblicher Handlungsbedarf besteht. Werden indessen die vorerwähnten Massnahmen in der Praxis umgesetzt, insbesondere das «Human Rights Impact Assessment», bedarf es nicht der Schaffung neuer Grundrechte, weil der heutige internationale und nationale Verfassungsrahmen als ausreichend erscheint, um die technologischen Herausforderungen abzufedern.

Implementation of Normative Principles in the AI Environment

The following text is newly written, but based on three previous publications that contain valuable thoughts meriting to be re-produced with some additional considerations and adaptations. The three articles concerned are:

- TEXT 1: Socio-ethical values and legal rules on automated platforms: The quest for a symbiotic relationship, 36 *Computer Law & Security Review* 2020, No. 105380, 1–8
- TEXT 2: Artificial Intelligence ante portas: Reactions of Law, in: J – Multi-disciplinary Scientific Journal MDPI 4/2021, 486–499
- TEXT 3: Artificial Intelligence: A Theoretical Analysis of Regulatory Models, in: Belli/Gaspar (eds.), *The Quest for AI Sovereignty, Transparency and Accountability*, IGF Kyoto, Rio de Janeiro 2023, 71–83.*

The parts taken over from these three articles are identified in each case with a respective footnote at the beginning of the reused (not comprehensively taken over) text.

Table of Contents

I.	Introduction	64
II.	Design of a Comprehensive Normative Framework	66
1.	Guardrails for a Regulatory Environment	66
2.	Implementation of a Co-Regulatory Approach	68
III.	Regulatory Principles	70
1.	Transparency	71
1.1.	Notion	71
1.2.	Challenges	72
1.2.1.	Lack of Comprehensibility	73
1.2.2.	Problems of Mandated Disclosure	73
1.2.3.	Information Overload	74

* An updated and adjusted version of this article and the whole book has just been published by Springer, Cham 2025, 59–70.

2.	Accountability	75
2.1.	Notion	75
2.2.	Challenges	75
3.	Auditability	76
4.	Observability	77
IV.	Further Regulatory Objectives	79
1.	Non-discrimination	79
2.	Safety and Robustness	83
3.	Compliance	84
V.	Outlook	86

I. Introduction

Artificial intelligence (AI) offers new informational interaction and data processing opportunities enabling the implementation of innovative communications and business models.¹ The use of intelligent “devices” and the availability of algorithms include the potential to replace human activities by software and/or machines. Instead of human intervention, the programming of the code that executes the tasks becomes important.

Artificial intelligence provides the possibility to establish a “regime” of automated decision-making, being conducted in a very timely and effective manner. Automation of this kind is mainly feasible in situations not requiring a specific human input, for example, in the case of an algorithm-driven search or in case of a standardized exchange platform. However, automated decision-making can also cause many socio-ethical and legal challenges.

Hereinafter, the potential normative framework for a trust-oriented environment with regulatory tools suitable to minimize the occurrence of technological risks is addressed. Therefore, different elements of an appropriate normative framework are considered; the notion that legalization requires a specific form of discourse no longer reflects the needs of society; moreover, the inclusion of all concerned stakeholders is justified (leading to co-regulation as a new model).

Since the following considerations are based on a theoretical analysis,² a case study will not be presented by purpose. Obviously, AI offers manifold new

¹ The following wording is mainly taken over from TEXT 2, 486.

² The following wording is mainly taken over from TEXT 1, 1.

opportunities which have an impact on business models and markets. The use of intelligent “devices” and the availability of algorithms on platforms include the potential to replace human activities by software and/or machines. Instead of human intervention, the programming of the code, which executes the tasks, becomes important; thereby, automation can take place for manifold activities. This fact calls for the implementation of fundamental socio-ethical values into AI within an appropriate legal framework.

Artificial intelligence allows implementing a “regime” of automated decision-making being conducted in a very timely and effective manner. Such kind of automation is mainly feasible in situations not requiring a specific human input, for example in case of an algorithm-driven search or in case of a standardized exchange platform. Hereinafter, the analysis conceptualizes the value dimension in respect of automated business models and assesses possible regulatory tools that could help implementing the appropriate safeguards for its practical realization.

AI-driven platforms do have an impact on civil society as well as on the competitive environment. To reconcile socio-ethical values with legal rules, the following questions derived from a normative concept of society are to be addressed:³

- Do the AI processes comply with the fundamental principles such as human rights and non-discrimination?
- Is the automated decision-making based on a sufficient legal basis, at least in respect of governance-related matters?
- Does automated decision-making comply with all applicable requirements of data protection laws?
- Who is responsible for the monitoring of socially responsible activities and liable in case of a failure caused by the algorithms?

The subsequent considerations cannot answer all questions in a detailed way, but general observations help to identify the relevant regulatory responses.

³ In this sense also ROLF H. WEBER, *Dürfen Maschinen über Menschen entscheiden? Eine rechtliche Auslegeordnung im Lichte neuer Technologien*, Schweizer Monat, Februar 2019, 72 et seq. (see [Text 3, pp. 37 et seq.](#)).

II. Design of a Comprehensive Normative Framework

1. Guardrails for a Regulatory Environment

The processes and mechanisms being best suited to implement the socio-ethical and legal values as well as social justice must be properly identified.⁴ This task does not only encompass substantive elements but also organizational and procedural factors.

Rulemaking issues can be addressed from the perspective of different disciplines; nevertheless, in private matters (such as in case of online business models) the discussions must concentrate on the appropriate allocation of duties and responsibilities as well as the proper structuring of the concerned “organization” (for example, the offeror of the platform). In other words: rulemaking, at whatever level of social organization it may take place, refers to setting norms for the conduct of the business in an appropriate way. In addition, some key questions are to be asked and answered:⁵ (i) Who is entitled to set the rules? (ii) in whose’ interest? (iii) by which mechanisms? and (iv) for which purposes?

In view of these questions, the need is given to develop overarching networks and negotiation systems between the different stakeholders, thus forming a cooperative approach to rulemaking that includes the whole society, hence dividing responsibilities between public and private actors.⁶ Thereby, governmental regulations must be supplemented by self-regulatory initiatives; in particular, guidelines designing the normative framework for the activities executed by the “owners” or “controllers” of automated business models appear to be an appropriate instrument.

Law constitutes a structural system that is composed of an organized or connected group of objects (terms, units, or categories) forming a complex network.⁷ The function of law is crystallized in a “regime” of rules and institutions that underpin civil society, facilitate orderly interaction, and resolve conflicts and disputes arising in spite of the rules.⁸ Law can be created through

⁴ The following wording is mainly taken over from TEXT 1, 3–4.

⁵ ROLF H. WEBER *Shaping Internet Governance: Regulatory Challenges*, Zürich 2009, 105 et seq.

⁶ To legitimacy aspects see WEBER (*supra* note 5), 112–113.

⁷ The following wording is mainly taken over from TEXT 2, 486–487.

⁸ ROLF H. WEBER, *Internet Governance at the Point of No Return*, Zürich 2021, 13 with further references.

different processes, for example by negotiations among the concerned norm addressees, by imposition of legal rules through the governing body, or by evolution of self-regulatory mechanisms.⁹ The legal system may be seen not as a predetermined construct but as being embedded by way of rulemaking in other socially relevant systems.

Applying a theoretical perspective, rulemaking issues can be addressed from the angle of different disciplines, i.e. an interdisciplinary approach should prevail. Furthermore, the traditional rulemaking approach in international matters, namely the conclusion of multilateral treaties, does not fully fit the objectives of a regulatory framework setting guidelines for AI systems. Moreover, other mechanisms must play a more important role. A new rulemaking model that has been developed (and partly also applied) in Internet governance (as well as in the climate change and sustainability) is the multistakeholder participation model. If all concerned people and organizations of the public and the private sphere are involved in the discussions and negotiations of the regulatory framework for AI processes, the chances increase that the developments are in the interest and to the benefit of the whole society.¹⁰

Practical experience conveys the lesson that some basic challenges need to be addressed to make the multistakeholder concept successful, particularly if several forms of cooperation based on a variety of actors involved are considered. Thereby, four fundamental questions must be tackled in detail:¹¹ (i) How do concerned groups best match challenges with the organizations and networks? (ii) How can governing bodies/entities be most able to help develop legitimate, effective, and efficient solutions? (iii) How should the flow of information and knowledge necessary for successful legal regime be structured? (iv) How can different groups approach the coordination between the available normative networks to avoid conflicting interest?

Answers to these questions need differentiated thinking depending on the given environment. As items of shortcomings, the identification of adverse effects of automated decision-making in the relevant policy field, the facilitation of networking opportunities, and the public support of civil society interventions are to be considered.

⁹ WEBER (supra note 8), 13.

¹⁰ See ROLF H. WEBER, Legal foundations of multistakeholder decision-making, ZSR 2016 I 247, 249 et seq.

¹¹ See URS GASSER/RYAN BUDISH/SARAH MYERS WEST, Multistakeholder as Governance Groups: Observations from Case Studies, Berkman Center for Internet & Society Research Publications, Cambridge MA 2015, 1 et seq.

2. Implementation of a Co-Regulatory Approach

International legal instruments (Conventions, Treaties) usually qualify as hard law. From a theoretical perspective, hard law constitutes an institutional normative order.¹² In contrast, technological standards as well as social and ethical guidelines (only) build an informal normative order. During the last years, however, the borderlines have blurred, and the regulatory bodies are becoming intertwined, i.e., the hard/soft law dichotomy does not anymore correspond to the reality.¹³

The often-expressed assumption that hard law is qualitatively better than soft law does not hold in the present normative environment. Soft law can entail functions previously tied to hard law, for example the notion of coordinating the concerned actors and of improving the transactional efficiency.¹⁴ So far, the term soft law has not found a generally acknowledged definition. Partly, soft law is seen as a normative model entailing opinions and recommendations of authorities (e.g., the European Union Aviation Safety Agency (EASA) in European Union (EU) civil aviation law or the European Data Protection Board (EDPB) in EU data protection law). However, soft law is also used as a term for rules created by the participants of a specific community; these rules are usually efficient since they respond to real needs, mirror the technology, and provide the opportunity to flexibly adapt the legal framework to the changing environment. Since such rules are negotiated by the communities involved, the likelihood is high that they enjoy broad acceptance.¹⁵ Furthermore, the concerned individuals/entities are incentivized to conduct permanent consultation processes to develop and implement the regulations.¹⁶ A private or inter-agency process laying down a set of rules might also more easily realize new solutions than hard law.¹⁷ In addition, this regulatory model develops and establishes rules independently of the principle of territoriality, i.e., soft law has a global reach at least in principle.

¹² ROLF H. WEBER, *Realizing a new Global Cyberspace Framework. Normative Framework and Guiding Principles*, Zürich 2014, 42–43.

¹³ See the title of ROLF H. WEBER, *Overcoming the Hard Law/Soft Law Dichotomy in Times of (Financial) Crises*, 1 J. Gov. Reg. 2012, 1.

¹⁴ See WEBER (supra note 13), 11; ANDREW T. GUZMAN/TIMOTHY L. MEYER, *International Soft Law*, J. Leg. Anal. 2010, 171, 188 et seq.

¹⁵ WEBER (supra note 12), 27.

¹⁶ GUZMAN/MEYER (supra note 14), 179 et seq.

¹⁷ TIMOTHY L. MEYER, *Soft Law as Delegation*, 32 Int. Law J. 2008, 888, 897.

Soft law also carries risks since the private rules are not always transparent and not every relevant group concerned is always involved. Furthermore, the active participants spending human and/or financial resources are also confronted with a “free-rider problem”.¹⁸ In addition, self-regulation has a lower democratic legitimization and, particularly, non-compliance with private rules might be confronted with problems of enforcement and does not mandatorily cause sanctions.¹⁹ Nevertheless, the role of the governmental regulation could be limited to situations in which the private actors do not find suitable solutions themselves.²⁰

In a nutshell, soft law must fit into the overall legal environment with the objective of realizing a non-discriminatory and socially acceptable legal framework.²¹ As experience shows, by overcoming the traditional dichotomy, the concepts of hard law and soft law have moved closer to each other or even intermingled (see Table 1).²²

Hard Law	Legitimized	←	Partly legitimized	Soft Law
	Top-down bodies	←	Bottom-up bodies	
	Stable / inflexible	←	Adaptable / flexible	
	Enforceable	←	Partly enforceable	

As expressed by the arrows, soft law is moving towards hard law, i.e., notwithstanding the term “soft”, the respective rules and guidelines are becoming (or have already become) quite robust and are relatively narrow to the characteristics of the traditional hard law. It is to be expected that this move will continue during the coming years.

¹⁸ WEBER (supra note 12), 28.

¹⁹ See DAMIAN TAMBINI/DANILO LEONARDI/CHRIS MARSDEN, *Codifying Cyberspace: Communications self-regulation in the age of Internet convergence*, London 2008, 282–284.

²⁰ See JACOB E. GERSEN/ERIC A. POSNER, *Lessons from Congressional Practice*, 61 *Stan. L. Rev.* 2008, 573, 589.

²¹ For further details see ROLF H. WEBER, *Sectoral Self-Regulation as Viable Tool*, in: Mathis/Tor (eds.), *Law and Economics of Regulation*, Cham 2021, 27 et seq.

²² WEBER (supra note 21), 29; see also the table with the pyramid (hierarchy) of the different legal instruments in WEBER (supra note 21), 30, with further references.

Rules considered by the “governed” individuals as adequate guidelines appear to be legitimate since private incentives lead to a need-driven rule-setting process.²³ Insofar, soft law contributes to more extensive capacity building, and its enabling functionality serves to easing the coordination process while at the same time providing directionality to the provision of cross-border standards. Soft law quality allows regulators to enter arrangements by varying scope and specificity and then to clarify (or change) the expectations of the involved actors.²⁴

If soft law does not cover the whole scope of a socially desirable legal framework, it must be backed by a governmental “regime”, for example in the case that some basic rules should not be left to the private actors. In such a situation, legal doctrine refers to the concept of “co-regulation” (also called regulated self-regulation).²⁵ In some circumstances, it makes sense to have soft laws approved by a public authority as experiences in certain regulatory ecosystems (financial markets, media markets, Internet markets) have evidenced.²⁶ Nevertheless, in view of the rapid growth, the complex inter-relationships, and the dynamic changes having taken place in the new technological environment, it appears to be broadly accepted that a more flexible and innovation-friendly model of regulation is required.²⁷

III. Regulatory Principles

Many regulatory principles known from traditional regulated markets also play a role in the AI environment but need some further refinement and adaptations.²⁸ As an overview, the different relevant principles can be designed as follows (Table 2):

²³ See WEBER (supra note 12), 22 et seq.

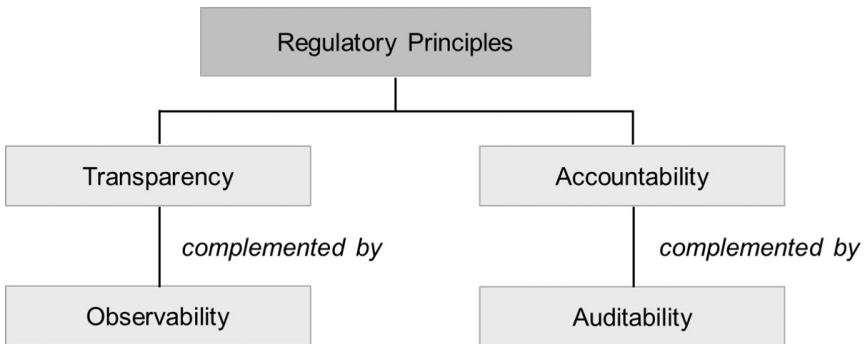
²⁴ WEBER (supra note 21), 28.

²⁵ See MYRIAM SENN, *Non-State Regulatory regimes, Understanding Institutional Transformation*, Berlin/Heidelberg, 2011, 43, 139–148, 230.

²⁶ CHRIS MARSDEN, *Internet Co-Regulation and Constitutionalism: Towards a More Nuanced View*, 2011, 9 et seq., <http://dx.doi.org/10.2139/ssrn.1973328>.

²⁷ JULIA BLACK, *Constitutionalizing Self-Regulation*, 59 Mod. L. Rev. 1996, 24, 32 et seq.

²⁸ This chapter is mainly (with the exception of Table 2) taken over from TEXT 3, 71 et seq.



1. Transparency

1.1. Notion

The use of AI offers many benefits but also causes some risks; therefore, the question arises how mitigation measures should be designed. In the past, transparency has often been judged as appropriate remedy. Indeed, already more than 100 years ago (in 1913), Brandeis wanted to make visible the opaque and hidden information, with the objective of creating truth that could enable control and serve as a “disinfectant”.²⁹

Transparency is usually assessed as encompassing characteristics such as clarity, accuracy, accessibility and truthfulness. These elements are also important in the context of artificial intelligence. As in other societal segments, transparency can enable access to the information necessary for the evaluation of opportunities and costs of operations and exchanges. Such an understanding of transparency links information disclosure to visibility, insight, and effective regulatory judgement.³⁰ But the assumption that transparency is sufficient to combat AI challenges and proposes to apply additional regulatory models might be questionable.

²⁹ LOUIS BRANDEIS, *The Other People's Money and How the Bankers Use It*, New York 1914, 92.

³⁰ The following text is based on ROLF H. WEBER, *Transparency on Digital Platforms*. Weblaw Jusletter, August 31, 2023; this publication analyzing the comparable digital platforms will not be cited anymore further on.

Often transparency is differentiated into three main pillars, namely (i) procedural transparency, (ii) decision-making transparency and (iii) substantive transparency.³¹

- *Procedural transparency* encompasses rules and procedures in the operation of legal entities that must be clearly stated, have an unambiguous character and are publicly disclosed. The rules should also make the process of governance and law-making accessible and comprehensible for the public.
- *Decision-making transparency* can be seen as reasoned explanations for decisions that, together with public scrutiny, are able to strengthen the institutional credibility and legitimacy of decisions.
- *Substantive transparency* is directed at the establishment of rules containing the desired substance of revelations, standards and provisions which avoid arbitrary or discriminatory decisions; substantive rules often include requirements of rationality and fairness.

In the AI context, all three elements are relevant. The concerned individuals need to know how the data processing is conducted (procedure), who is taking decisions and what material standards are applied. Compliance with the three elements also impacts the assessment of the below discussed accountability.

1.2. Challenges

But the concept of transparency having become an essential regulatory element mainly in financial markets and consumer laws, is increasingly exposed to challenges and critical analyses. Echoing these voices, transparency is partly seen as policy panacea.³² AI data processing is exposed to false binaries between secrecy and openness, to strategic occlusions and to market-dominant behavior of big enterprises; these factors influencing the algorithmic matching results can lead to power imbalances. Furthermore, some issues cause specific challenges and must be discussed in further detail, namely the lack of comprehensibility, the requirement of mandated disclosure, and the information overload.

³¹ See WEBER (supra note 5), 121.

³² For further details see BERNHARD RIEDER/JEANETTE HOFMANN, Towards platform observability, 9 Internet Policy Review 2020, 1, 3–6.

1.2.1. *Lack of Comprehensibility*

Over the last ten years, regulations in financial markets and consumer segments have substantially increased the scope of information duties to be observed by providers of goods and services.³³ Examples are the extensive information requirements for capital markets prospectuses and the specific (hardly understandable) descriptions for medical products. As mentioned, Brandeis attributed the characteristics of “sunlight” and “disinfectant” to the transparency principle; however, since the detailed disclosure often goes too far, the recipient does not anymore understand the key message.

Transparency should address the way information is delivered to optimize the outcome of the informational process. The basic objectives of transparency require robust and general rules; this principle is now enshrined in article 12 para. 1 GDPR; information must be given “in a concise, transparent, intelligible and easily accessible form, using clear and plain language”.³⁴ If the information is clear and straightforward, the addressee will be able to fully understand it (so-called comprehensibility).³⁵ The GDPR requirement is very appropriate, however, the reality in the AI environment shows that the chosen information approach often does not comply with the GDPR requirement (Article 12).

1.2.2. *Problems of Mandated Disclosure*

In the business-oriented context, the increasing number of information obligations has been mainly criticized by representatives of the law and economics discipline under the heading of “mandated disclosure paradigm”.³⁶ Apart from the hidden costs caused by such kind of disclosure (for example detailed information obligations in the Artificial Intelligence Act of the EU³⁷), Ben-Shahar & Schneider argue that the mandated disclosure would exacerbate inequality, impair consumers’ decisions and deter lawmakers from

³³ See ROLF H. WEBER, *The Disclosure Dream – Towards a New Transparency Concept in Consumer Law*, EuCML 2023, 67–68.

³⁴ General Data Protection Regulation (GDPR) 2016/679 of April 2016, OJ 2016 L 119 of 4 May 2016.

³⁵ See also ROLF H. WEBER, *From Disclosure to Transparency in Consumer Law*, in: K. Mathis/A. Tor (eds.), *Consumer Law and Economics*, Cham 2021, 73, 79–81.

³⁶ For a general assessment see OMRI BEN-SHAHAR/CARL E. SCHNEIDER, *More Than You Wanted to Know: The Failure of Mandated Disclosure*, Princeton 2014.

³⁷ See Art. 13 and Art. 20 of the Artificial Intelligence Act, Regulation 2024/1689 of 13 June 2024, OJ 2024 L 1–144 of 12 July 2024.

adopting better regulations;³⁸ and these authors add that the provided information whether individually aggregated or based on advice “will not adequately help the naïves in their dealings with the sophisticated”.³⁹

Even if these statements could be partly contested, it appears to be doubtful that the information addressees indeed read and understand the mandatorily provided information. Distributed ledger technologies aggravate the problem: platform users are often not able to understand the “IT codes” meaning that for example the disclosure of mathematical formulas constituting a smart contract do not lead to an informed addressee. Therefore, a potential way to assist individuals in making better decisions would be to direct choices through smart incentives without mandating a certain outcome.

1.2.3. *Information Overload*

The transparency principle is also confronted with the issue of information overload. Looking from a societal perspective, too detailed information requirements could have two negative effects:⁴⁰

- The sheer volume and the intensity of information lead to a confusion effect since the recipients are not anymore able to cope with all information details and lose the necessary overview in respect of the disclosed data.
- The permanent delivery of (similar) information causes a Cassandra effect; even if the recipients take note of the information, its content is no longer seen as being serious and reliable.

The general wisdom that overconsumption of information can have negative effects or even be risky also applies in respect of detailed disclosure requirements:⁴¹ (i) Over-information consumes working and leisure time on both sides of an informational relationship. (ii) Attention is a scarce resource; a person cannot dispose of this resource in an unlimited way. (iii) Over-information increases the risk that messages or data being spread out are redundant.

Notwithstanding the fact that detailed information provisions can constitute a certain value for individuals having broader expertise (academics, lawyers), it

³⁸ See BEN-SHAHAR/SCHNEIDER (supra note 36), *passim*, and WEBER (supra note 35), 75 and 77–78.

³⁹ OMRI BEN-SHAHAR/CARL E. SCHNEIDER, The Failure of Mandated Disclosure, 159 Univ. of Penn. L. Rev. 2011, 647, 748.

⁴⁰ See WEBER (supra note 35), 79–80 with further references.

⁴¹ See NIKLAS LUHMANN, Die Gesellschaft der Gesellschaft, 2 Vol., Frankfurt 1997, 1090, 1097 and 1102.

should not be underestimated that the balancing of interests remains difficult in relation to addressees not experienced in the AI services context. Incomplete disclosure leaves people ignorant, but complete disclosure creates overload problems;⁴² consequently, the regulator should recognize that “less is more” even if it cannot be excluded that “less is not enough”.⁴³

2. Accountability

2.1. Notion

Accountability, often being called a “twin” of transparency, encompasses the obligation of one person or legal entity to give account of, explain and justify the undertaken actions or decisions to another person in an appropriate way.⁴⁴ Accountability is a pervasive concept, including political, legal, philosophical, and other aspects, each of them casting a different shade on the meaning of the terms. Checks and balances as emanation of accountability constitute a prerequisite for legitimacy and a key element of any governance discussion.

As a fundamental principle, accountability concerns itself with power and power implies responsibility. Therefore, accountability can be framed among three elements,⁴⁵ namely (i) the provision of information in a timely manner, (ii) the introduction of standards that hold governing bodies accountable, and (iii) the implementation of mechanisms of sanction. In addition, accountability needs to include the democracy element of global governance if the outcome of the decision-making processes should be acceptable to civil society in general.⁴⁶

2.2. Challenges

Regulatory approaches seeking to create accountability in the AI context must tackle the relevant issues by “opening the black box” of algorithmic decision-making.⁴⁷ However, accountability should extend beyond oversight of algorithms and behavioral conduct. In addition, behavior cannot be reduced to

⁴² WEBER (supra note 35), 79–80.

⁴³ BEN-SHAHAR/SCHNEIDER (supra note 39), 647.

⁴⁴ WEBER (supra note 5), 133 with further references.

⁴⁵ See WEBER (supra note 8), 70.

⁴⁶ RIEDER/HOFMANN (supra note 32). 6.

⁴⁷ The black box problem is fundamentally described by FRANK PASQUALE, *The Black Box Society. The Secret Algorithms That Control Money and Information*, Cambridge MA 2015.

conduct since tight integration of data collections and targeted “intervention” in form of “surveillance capitalism”⁴⁸ has produced a market form that is unimageable outside the digital milieu. Consequently, the rising power of AI providers makes it necessary to assess what kind of accountability must be applied to understand the processes and their consequences in more detail.⁴⁹

A specific challenge to accountability is the occurrence of opacity⁵⁰ being an obvious concern that may stem from the increasing “mismatch between mathematical optimization in high-dimensionality characteristics of machine learning and the demands of human-scale reasoning”.⁵¹ The applicable techniques usually develop decision models inductively and learn programs from data.

Since many variables come into play, academics argue that the developed algorithms are not easily “legible” in daily life.⁵² Consequently, transparency in the sense of reconstructing the procedure of algorithmic decision-making often does not lead to an informative outcome. Even if regulators were given access to data centers and source code, the above discussed comprehensibility would not be straightforward in view of complex code designs and involved machine learning. In addition, the existence of different programming languages and execution environments adds further complications.⁵³

3. Auditability

An improvement of transparency and accountability can be achieved by extended auditability requirements if the respective provisions overcome an insufficient understanding of algorithms and platform architectures.⁵⁴ In order

⁴⁸ Term introduced by SHOSHANNA ZUBOFF, *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power*, New York 2019, 15.

⁴⁹ To the autonomy and power elements in the accountability context see also WEBER (supra note 8), 71.

⁵⁰ RIEDER/HOFMANN (supra note 32), 6–7.

⁵¹ JENNA BURRELL, *How the machine ‘thinks’: Understanding opacity in machine learning algorithms*, September 2015, 2, <http://dx.doi.org/10.2139/ssrn.2660674>.

⁵² ANSGAR KOENE ET AL., *A governance framework for algorithmic accountability and transparency*, European Parliamentary Research Service Study, April 2019, 31–32.

⁵³ PAUL DOURISH, *Algorithms and their others: Algorithmic cultures in context*, *Big Data & Society* 2/2016, 1, 4.

⁵⁴ See also MIKE ANNANY/KATE CRAWFORD, *Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability*, 20 *New Media & Society* 2018, 973–974.

to reach the (theoretical) transparency objective, it would be necessary to develop an institutionalized mechanism for the verification of AI-provided information or data. The respective efforts are done under the heading of “auditability”.

Several aspects need to be considered in the implementation of auditability principles:⁵⁵ (i) The creation of an intermediary (public or private sector entity) that audits data provided by large online platforms can ensure the accuracy of data. (ii) By bundling the auditing process through centralized auditing intermediaries, the exposure of sensitive private data to as few actors as possible is limited. (iii) By distancing the audit process from the regulator that is asking for data ensures that regulatory action does not overstep its bounds. (iv) By limiting the number of points through which the online platforms need to interact with outside intermediaries limits potential security risks that could arise from providing access to a wide variety of systems. (v) Having numerous regulators involved in auditing is likely to create unnecessary and redundant processes. (vi) Organizing auditing of transparency data through an external auditing intermediary ensures that even regulators without the capacity to organize audits themselves still may have access to such a system through auditing intermediaries.

The most important question about an auditing intermediary concerns the decision of whether such an intermediary would be public, private or somewhere in between.⁵⁶ Such an institution could be created within the context of the recently adopted EU Digital Services Act (DSA).⁵⁷ A further challenge raised by the proposal of auditing intermediaries is how much access to data these intermediaries would actually need. It must also be avoided that auditing intermediaries are misused by authoritarian countries for strategic national interests.⁵⁸

4. Observability

A further approach proposes to realize a concept of observability as a pragmatic way of thinking about the means and strategies necessary to hold

⁵⁵ For further details see BEN WAGNER/LUBOS KUKLIS, Establishing Auditing Intermediaries to Verify Platform Data, in: Moore/Tambini (eds.), *Regulating Big Tech*, Oxford 2021, 169, 172–173.

⁵⁶ WAGNER/KUKLIS (supra note 55), 174.

⁵⁷ Regulation 2022/2065 of 19 October 2022 on a Single Market for Digital Services, OJ 277 L 1 of 27 October 2022.

⁵⁸ WAGNER/KUKLIS (supra note 55), 174–175.

AI providers accountable.⁵⁹ Unlike transparency being normally described as a state that may exist or not, observability emphasizes the conditions for the practice of observing in each domain. While observability incorporates similar regulatory goals as transparency, it also partly deviates, most importantly by understanding accountability as a complex, dynamic “social relation”.⁶⁰ Observability should be a mechanism that can overcome the lack of sensitivity for fundamental power imbalances, strategic occlusions, and false binaries between secrecy and openness.

The challenges raised need to be addressed in a broad way, beginning with the question of how large-scale, transnational environments that heavily rely on technology as a mode of governance can be assessed.⁶¹ The concept of observability seeks to develop concrete actions in respect of (i) how people need to be treated in the digitized environment, (ii) how connections between participants are made and structured, and (iii) which outcomes should be achievable.⁶²

In the academic literature, the concept of observability starts with the recognition of a growing information asymmetry between AI services providers and civil society. The frequently given data monopoly situation deprives society of a crucial resource for producing knowledge about itself. The deep political and social repercussions reflect the need to implement broader forms of social accountability.⁶³ The concept of observability should be based on public interest as a normative horizon for assessing and regulating the societal challenges. In the context of the public sphere, public interest encompasses the protection of human rights such as the freedom of expression and the freedom of information, fostering cultural and political diversity throughout the whole society.⁶⁴

Furthermore, the principle of observability reflects the acknowledgment that the volatility of AI solutions requires continuous observation. If terms of

⁵⁹ A broad discussion of the observability concept is offered by RIEDER/HOFMANN (supra note 32), 9–18.

⁶⁰ See MARK BOVENS, *Analysing and Assessing Accountability: A Conceptual Framework*, 13 *European Law Journal* 2007, 447, 450.

⁶¹ RIEDER/HOFMANN (supra note 32), 9–10.

⁶² See also JOSÉ VAN DIJCK/THOMAS POELL/MARTIN DE WAAL, *The Values in a Connective World*, Oxford 2018, 158.

⁶³ RIEDER/HOFMANN (supra note 32), 9–10.

⁶⁴ See also JOSÉ VAN DIJCK, *Governing digital societies: Private platforms, public values*, 36 *CLSR* 2020, 1, 3.

services contracts would be made available as machine-readable documents, the ongoing observation and interpretation of AI-related activities could be facilitated.⁶⁵ Another factor concerns the availability of interfaces that provide continuous access to relevant data. Thereby, questions of how data and analytical capacities are made available, to whom, and for what purpose must be tackled.⁶⁶

Observability requires a critical audience. But the capacity for critique must be broader than “only” a critical attitude. Moreover, frameworks for data access should be linked to the cultivation of a robust civil society. Therefore, observability as a social relation makes scrutiny of realized transparency by a specific forum necessary.⁶⁷

Regulating AI providers with the objective of increasing observability does mean working towards structured information interfaces between them and society. Such kind of regulation requires engaging with the specific properties of algorithmic systems and the co-produced nature of AI results. The complex interactions between technical design, terms of service, and often large numbers of both users and “items/issues” have the consequence that the existing processes are conceptually insufficient.⁶⁸ AI providers should become subject to public interest requirements as a normative benchmark; elements could consist of risk control measures, auditability reviews, behavioral rules and strict responsibility obligations.

IV. Further Regulatory Objectives

1. Non-discrimination

Artificial intelligence systems cause a risk of discrimination.⁶⁹ Many examples already exist that show a certain degree of structural inequality or problems of foreseeable or unforeseeable biases, amongst others, in respect of gender

⁶⁵ RIEDER/HOFMANN (supra note 32), 13–14.

⁶⁶ VAN DIJCK (supra note 64), 3.

⁶⁷ BOVENS (supra note 60), 450.

⁶⁸ See also PHILIP M. NAPOLI, Social media and the public interest: Governance of new platforms in the realm of individual and algorithmic gatekeepers, 39 Telecommunications Policy 2015, 751 et seq.

⁶⁹ The following wording is mainly taken over from TEXT 2, 489–490.

considerations or the treatment of colored people.⁷⁰ Technologically, self-learning algorithmic programs do exercise the code implemented in the software; however, the design of the code is influenced by the programming person/device.

Most international legal instruments (such as the Declaration of Human Rights of the United Nations) as well as practically all national constitutions around the globe contain a provision prohibiting undue discrimination. Usually, without going into the details, the wording of the respective provisions is rather broad and encompasses direct and indirect discrimination.⁷¹ However, the problem of indirect discrimination consists in the fact that specific criteria seem to be “neutral” even if in their application some individuals are more affected than others.⁷² In such case, discrimination is solely based on the relation of a person to the protected group.⁷³

Another practically important aspect concerns the legal situation that – at least in principle – the non-discrimination rules may only be invoked against States as users of algorithmic systems, not directly against private enterprises and individuals, except if the so-called horizontal effect of human rights against private individuals is acknowledged in each legal order. Therefore, by further developing the fundamental rights, it should be generally accepted that the right of non-discrimination constitutes an obligation to analyze and mitigate, throughout AI systems’ life cycle, the risk of unjust biases.⁷⁴

Furthermore, the non-discrimination principle should also encompass aspects of “fairness” playing an important role in the AI context. Fairness is an aspect showing compliance with behavioral rules of a conscious human being. The constitutional provisions available usually do not literally refer to this term;

⁷⁰ CLAUDE CASTELLUCCIA/DANIEL LE MÉTAYER, *Understanding Algorithmic Decision-Making: Opportunities and Challenges*, Study for the European Parliament, Brussels 2019, 7 et seq.; NADJA BRAUN BINDER/THOMAS BURRI/MELINDA F. LOHMANN/MONIKA SIMMLER/FLORENT THOUVENIN/KERSTIN N. VOKINGER, *Künstliche Intelligenz, Handlungsbedarf im Schweizer Recht*, Jusletter 28 June 2021, nos. 21 et seq.

⁷¹ ROLF H. WEBER/SIMON HENSELER, *Regulierung von Algorithmen in der EU und in der Schweiz*, EuZ 2020, 28, 39 with further references.

⁷² BRAUN BINDER ET AL. (supra note 70), no. 17; see also SOLON BAROCAS/ANDREW D. SELBST, *Big Data’s Disparate Impact*, 104 Cal. L. Rev. 671, 680–687.

⁷³ SANDRA WACHTER, *Affinity Profiling and Discrimination by Association in Online Behavioural Advertising*, 30 BTLJ 2020, 367, 394–412.

⁷⁴ See also COUNCIL OF EUROPE, *Ad hoc Committee on Artificial Intelligence, Feasibility Study*, CAHAI(2020)23, Strasbourg, 17 December 2020, 23.

nevertheless, it appears to be adequate to apply a broad understanding of “discrimination”⁷⁵

In the meantime, some legislators have become active by implementing secondary legal instruments that contain specific provisions forbidding discrimination based on gender, race, or ethnic origin in connection with work, social support, or insurance coverage. Respective laws are particularly known in the Western hemisphere but not around the globe.⁷⁶ Certain justifications for “unequal” treatment are usually also foreseen but their scope is mostly restricted.

The non-discrimination principle can equally be based on personality rights being widely recognized by State legislators. The personality rights are expressions of the individual self-determination principle and of human dignity. The exact scope depends on the wording of the stated personality rights.⁷⁷ Irrespective of the concrete design, such rights’ position is always linked to the already generally mentioned concepts of trust and accountability.

The protection against discrimination is also foreseen in soft law guidelines developed and published by international and regional organizations; such guidelines (as for example drafted by an expert group of the EU) often state the principle that AI systems must be based on transparent analyses and enshrine equal treatment principles.⁷⁸ According to the Organisation for Economic Co-operation and Development (OECD) Principles, AI actors playing an active role in the lifecycle of AI systems have to see to it that human rights and equal treatment principles are observed in the technological design.⁷⁹ Similarly, private enterprises such as Google and Microsoft also state the principle of non-discrimination in their ethics guidelines.

The implementation of the non-discrimination principle can cause difficulties in practice since AI systems are not tailored to the specificities of the available legal instruments. Even if an actual legal vacuum does not exist, experience and academic research studies have shown that the prevailing fundamental

⁷⁵ BRAUN BINDER ET AL. (supra note 70), nos. 25–26 with further references.

⁷⁶ CASTELLUCCIA/LE MÉTAYER (supra note 70), 56.

⁷⁷ BRAUN BINDER ET AL. (supra note 70), no. 29; WEBER/HENSELER (supra note 71), 40.

⁷⁸ HIGH-LEVEL EXPERT GROUP of the European Union, Ethics Guidelines for Trustworthy AI, April 2019, nos. 80 et seq., <https://digital-strategy.ec.europa.eu/en/library/ethics-guide-lines-trustworthy-ai>.

⁷⁹ OECD, Principles on Artificial Intelligence, 22 May 2019, revised 3 May 2023, no.1.2, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.

rights principles as well as secondary normative provisions do have certain limits and cannot secure comprehensive non-discrimination protection.⁸⁰

Social systems cannot only be conducted by machines. The legal framework enshrines fundamental rights; consequently, the ongoing dialog regarding the ethics of AI should expand to consider the human rights implications of these technologies including the risks of discrimination. Thereby, ethics and law can act in a complementary way, being able to inspire each other; efforts must be strengthened to develop a comprehensive approach for the socio-ethical and the legal dimension of AI value conceptualizations moving into the direction of a potentially symbiotic relationship.

Furthermore, the commitment of cybercrime activities is an ethically and legally not acceptable behavior. Victims of cybercrime are particularly vulnerable individuals since defense instruments cannot easily be implemented.⁸¹ Therefore, the protection of victims by the States in combatting cybercrime appears to be a valuable governmental objective. Indeed, AI measures could support the normative system by realizing a procedural rights design that protects the attacked victims from the abuse of technologies (and unjustified power). AI and automation tools enable increased efficiency and fairness of defense actions. In particular, risk assessment tools do have the potential to exercise a significant positive impact on the rights of individuals.

A key parameter of such risk assessment tools consists in their predictive accuracy. The collection of past data and subsequently the appropriate processing and scoring is of utmost importance. In view of the fast technological advances, new variables appearing within short time intervals are an unavoidable factor. Therefore, risk elements need to be continually rebalanced in response to the changed inputs. If the tools are developed and applied by private enterprises (not by governmental agencies), precautionary measures are necessary to find an adequate balance between business secrecy needs, the exercise of private power and public interests. Furthermore, it must be avoided that risk assessment tools do have a negative impact on individual rights to equality and non-discrimination.

Artificial intelligence, at least partly based on big data analytics, is not mainly following the traditional proofs of causality, but puts the focus on correlations.

⁸⁰ See also COUNCIL OF EUROPE (supra note 74), 21–22.

⁸¹ The subsequent text is based on ROLF H. WEBER, Artificial Intelligence as Tool for Victims' Protection against Cybercrime, 31 May 2021, <https://medium.com/youth-igf-informs/artificial-intelligence-skills-new-cybersecurity-skills-to-protect-citizens-8c134c7e0ac2>.

Consequently, the protection of vulnerable individuals such as victims of cybercrime is dependent on further knowledge creation that can develop an ethical and legal environment in a non-discriminatory way. Apart from the principles of accountability and auditability as mentioned above, mechanisms of collective actions enlarging the required knowledge might have to be broadened. The respective tasks are not easy to fulfil but AI at least offers chances which should be taken at hand in order not to miss positive opportunities.

2. Safety and Robustness

Apart from the described regulatory principles, the safety and robustness of the AI-systems equally are of importance;⁸² these properties are the basis of trust.⁸³ The respective “infrastructure” and the related software programs must be safe and robust throughout their entire lifecycle so that the data-driven communications and transactions can overcome adverse conditions or foreseeable potential misuse. Consequently, the traceability of the datasets, processes and decisions must be secured.

International instruments already state that the likely impact of intended AI processing and its broader ethical and social implications must be adequately considered to safeguard human rights and fundamental freedoms. For example, the revised Recommendation of the OECD Council on Artificial Intelligence refers in part 1.4 to robustness, security and safety as follows:⁸⁴ “AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk. To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system’s outcomes and responses to inquiry, appropriate to the context and consistent with the state of art”.

Furthermore, the use of AI by modern data processing techniques and the trend towards implementation of data-intensive processes require a more advanced risks’ assessment understanding by individuals and businesses since possible adverse outcomes of data processes cannot be excluded.⁸⁵ In

⁸² The following wording is mainly taken over from TEXT 1, 7–8.

⁸³ See also ROLF H. WEBER, TEXT 1, 1–2 with references.

⁸⁴ OECD (supra note 79).

⁸⁵ COUNCIL OF EUROPE (supra note 74), 13.

particular, the challenge must be tackled that automated decision-making may have an impact on fundamental rights including discrimination as well as the collective social and ethical values. Compliance with socio-ethical values (as well as their assessment) is more complicated in case of AI deployment than in case of traditional data processing.⁸⁶

In this context it should not be overlooked that risks and compliance assessments are not only justified by the collective social and ethical values as well as by the nature of the fundamental rights and freedoms potentially affected by AI applications, but they also represent an opportunity (if undertaken in a participatory environment) to better fostering public trust as key objective of the information society.⁸⁷

3. Compliance

Enterprises generally are required to comply with the requirements of the applicable regulatory framework; thereby, an interdisciplinary approach appears to be appropriate.⁸⁸ In doing so, the following issues related to algorithms/artificial intelligence might have to be considered.⁸⁹

- *Property rights*: Algorithms could (also) be used to collect, aggregate, display, and share informational goods protected by intellectual property law.
- *Privacy rights*: Automated systems collecting personal data from users and processing them with algorithms cause concerns for data protection and privacy (taken now up by the new provision of Article 22 GDPR).⁹⁰
- *Censorship*: Algorithmic programs can introduce restrictions to control or block content that otherwise is accessible for users.

⁸⁶ See for example ONORA O'NEILL, *Justice, Trust and Accountability*, Cambridge 2005, *passim*.

⁸⁷ See ONORA O'NEILL, *A Question of Trust*, Cambridge 2002, 61 et seq.

⁸⁸ The following wording is mainly taken over from TEXT 1, 5–6.

⁸⁹ OECD, Directorate for Financial and Enterprise Affairs, Competition Committee, *Algorithms and Collusion – Background Note by the Secretariat*, Paris, 9 June 2017, 43.

⁹⁰ The specific challenges caused by automated systems in the data protection context are not a topic of this chapter; for further information see LILIAN EDWARDS/MICHAEL VEALE, *Slave to the Algorithm? Why a «Right to an Explanation» is probably not the Remedy You are Looking For*, 16/1 Duke L. & Tech. Rev. 2017, 18 et seq.; STEFANIE HÄNOLD, *Profiling and Automated Decision-Making: Legal Implications and Shortcomings*, in: Corrales et al. (eds.), *Robotics, AI and the Future of Law*, Singapore 2018, 123 et seq.

- *Discrimination*: Automated data-decision processes have the potential to lead to social discrimination based on the processed personal information.⁹¹
- *Abuse of market power*: Algorithms (artificial intelligence) could facilitate the application of exclusionary and exploitative measures.
- *Tacit collusion*: If algorithms coordinate competition parameters even without personal intent of market players, anti-competitive effects like the well-known concerted practices can occur.
- *Manipulation*: Manipulated algorithms collect and select information in view of given business and political interests instead of its relevance or quality.⁹²

The above list of substantive issues is neither comprehensive nor enumerative. Depending on the given environment, only some issues are relevant or additional issues merit attention. In case of AI systems' use, for example, property rights, privacy rights, discrimination, abuse of market power, and manipulation can play a role.

Irrespective of its design, a technical system such as an AI model should inspire trust; from an ethical perspective trust enhances cooperation and fosters reciprocal relations.⁹³ A proper compliance can support the building of trust enshrining more than legal standards but also socio-ethical values that have a broader scope than the traditional justice.⁹⁴ Further, trust is also linked to "reliance"; if an individual is relying on something or someone to display a certain behavior, than confidence in the respective activities will increase.⁹⁵ Reliance can be based on standards or contractual terms, observed by the offeror of a service, leading to transparency and accountability.⁹⁶

⁹¹ See above sub-chapter [IV.1](#).

⁹² To the democracy aspect of the manipulation see COUNCIL OF EUROPE, Declaration(13/02/2019)I of 13 February 2019.

⁹³ For further details see ROLF H. WEBER, Ethics in the Internet Environment, in: Global Commission on the Internet Governance, Paper Series no. 39. July 2026, 7.

⁹⁴ See also O'NEILL (supra note 87), 61 et seq.

⁹⁵ WEBER (supra note 93), 7.

⁹⁶ See above sub-sections [III.1](#) and [III.2](#).

V. Outlook

The new technological developments, particularly the AI-driven systems and processes, cause major challenges for the legal framework. Regulations that are capable to realize a human-centric environment must not only be implemented within short time intervals but also tackle new issues that have not been relevant in the traditional normative setting.

However, experience has shown that law, even as a relatively static discipline, can appropriately react to the regulatory needs of digital transformation. On the one hand, ethical guidelines that are based on general legal principles (such as good faith, human rights, anti-discrimination) are available and can be applied in a reasonable manner; on the other hand, known and newly developed regulatory principles (for example transparency, accountability, auditability, observability) merit to be better designed and focused in the AI environment.

This task can only be fulfilled if an interdisciplinary approach is realized. Legal, economic, ethical, philosophical and sociological sciences must meet technological developments and reconcile potential differently designed requirements. The exchange of opinions and ideas between the disciplines is imperative for the realization of a fruitful and forward-looking environment.

Rule of Law and Ethics-based Equality in the Age of Artificial Intelligence

Erstveröffentlichung in: Schweighofer et al. (Hrsg.), *Der Mensch im Zentrum – KI, Ethik und Recht* (IRIS 2025), Bern 2025, 23–30

The manifold opportunities provided by new technologies and algorithms also cause challenges for the inclusion of the whole society (deprivation, discrimination, distortion). As a counterbalance, the normative framework is called to implement measures securing access to and availability of AI applications. Respective concepts are on hand: the rule of law and the equality principle, if combined with AI literacy, can contribute to fairness and non-discrimination as well as combat tendencies leading to a digital or algorithmic divide.

Table of Contents

1.	Introduction	88
2.	Attributes and Challenges of Digital and Algorithmic Divides	89
2.1.	Attributes of Divides: 4A (Awareness, Access, Affordability, Availability)	89
2.2.	Challenges of Divides: 3D (Deprivation, Discrimination, Distortion)	90
3.	Normative Principles for the AI Age	91
3.1.	Rule of Law	91
3.1.1.	Notion	91
3.1.2.	Substantive Aspects	92
3.1.3.	Formal Aspects	93
3.2.	Ethics-based Equality	94
3.2.1.	Equality as Emanation of the Rule of Law and of Ethics	94
3.2.2.	Ethics: Contents and Standards	95
3.2.3.	Digital and Algorithmic Literacy	97
3.3.	Conclusion: Chances for Overcoming Digital and Algorithmic Divides	98
4.	Bibliography	98

1. Introduction

While Artificial Intelligence (AI) and machine learning technologies are still in their infancy, it is already obvious that their increased application has a transformative effect and influences the legal environment to a considerable extent. AI can change the decision-making processes in both the private and the public sector. Therefore, the viability of the basic legal principles, such as the rule of law and ethics-based equality as important normative concepts are at stake;¹ the legal framework may not stay aside if human decisions are concerned in the AI world.²

An important aspect of a human and civil society-oriented approach governing AI concepts concerns the “inclusion” issue; AI benefits can unfold if everybody has the chance to participate in them. For the last twenty years, the challenge of the digital divide has been intensively discussed. New technologies offer the potential to take advantage of economic benefits and societal participation, however, negative impacts cannot fully be excluded. The concerns about widening of the digital divide are extended to the risk of an algorithmic divide with the advent of AI and machine learning. For good reasons, in March 2024, the UN General Assembly agreed on a Resolution titled “Seizing the Opportunities of Safe, Secure, and Trustworthy Artificial Intelligence Systems for Sustainable Development”.³

AI and machine learning applications are built on patterns, rules, or decision trees that can lead to two situations:⁴ (i) If the AI applications are designed in a way that comprehension is advanced allowing for effective human oversight, appropriate decision-making and rights’ exercise based on enhanced transparency, civic engagement will be promoted leading to the prevailing situation of a “white box”; in such situation, the algorithms are explainable in relation to

¹ As the title of the IRIS 2025 Conference makes clear, AI must be assessed from a holistic perspective; therefore, law and ethics are addressed in view of developing human-oriented policies hereinafter.

² For good reasons, the Council of Europe as an institution pursuing objectives such as human rights, democracy, and rule of law developed a Framework Convention on Artificial Intelligence and Human Rights, 17 May 2024, CM[2024]52-addfinal, that particularly addresses these issues in the normative framework.

³ Joint Statement on the Proposed UN General Assembly Resolution on “Seizing the Opportunities of Safe, Secure, and Trustworthy Artificial Intelligence Systems for Sustainable Development” (14th March 2024).

⁴ For a general discussion of the white box and black box tensions see LOYOLA-GONZÁLEZ, 154097 and 154101.

the data used, their training procedure, gravity and relevance for the output, and the logic employed.⁵ (ii) In case of a high complexity of the AI systems consisting of multiple decision trees and partly unknown parameters, combined with a lack of transparency, the situation of a “black box” is given; in such a case, humans are unable to understand why and how automated decisions are reached leading to the paradox that the more complex and powerful an algorithm, the more opaque it is likely to be.⁶

Hereinafter, the attributes and the challenges of digital and algorithmic divides will be concisely discussed ([part 2](#)). Then, the discussion of normative solutions, particularly the rule of law and the ethics-based equality principle, is the core of the contribution, concluding that the potential divides can be overcome ([part 3](#)).

2. Attributes and Challenges of Digital and Algorithmic Divides

2.1. Attributes of Divides: 4A (Awareness, Access, Affordability, Availability)

AI and algorithmic developments are a multidimensional phenomenon and generate a wide variety of interdisciplinary challenges; a key issue consists in the attempt to ensure better access to machine learning and AI to avoid or at least minimize a digital divide or an inequitable gap between users and non-users. Academics point to four attributes of algorithmic divide,⁷ namely the 4A:

- *Awareness*: New technologies are only fruitful if the individuals and the enterprises are aware of the advantages and risks of algorithm-enhanced products and services; in addition, knowledge is needed to avoid harms (“biases and bugs”).
- *Access*: The most important attribute is access; the material individual and societal benefits of algorithm-enhanced products and services are only playing out if everybody being interested in taking advantage can get access to them.⁸

⁵ A detailed discussion of the relevance of legal standards can be found in SUSSKIND, part 2.

⁶ BERMAN, 1325 et seq.

⁷ For more details see YU, 336 et seq.

⁸ For a recent analysis of the inclusion aspect see WEBER, 2024, 231 et seq.

- *Affordability*: Next to the general requirement of access, affordability to use the new products and services must be secured; if individuals without the necessary resources are prevented from enjoying the algorithm-enhanced products and services, a digital or algorithmic divide occurs.
- *Availability*: If certain products or services needed by an individual are not available, drawbacks in society cannot be excluded.

Only if the four mentioned attributes⁹ are fulfilled, can the whole society (in developed and developing countries) take advantage of AI and machine learning; therefore, not only a digital divide, but also an algorithmic divide must be avoided.¹⁰

2.2. Challenges of Divides: 3D (Deprivation, Discrimination, Distortion)

Three fundamental challenges influence the possibility for individuals to benefit from machine learning and AI, namely the 3D:¹¹

- *Deprivation*: Those who have no access to algorithm-enhanced products and services are deprived of the benefits provided by machine learning and AI; this is particularly the case for civil society in less developed countries. Examples discussed are (i) disaster relief, (ii) public health, (iii) food production, (iv) education and knowledge broadening, and (v) policy analysis (predictive models with proactive advice).
- *Discrimination*: Errors, biases, dehumanization, etc. are regularly discussed in the context of discrimination; academics lament that we live in a society which does not take enough efforts for combatting inequalities among its members. Discrimination can occur in the use of algorithms through “masking” or through inappropriate data feed of either training or input data (intentionally or not intentionally).
- *Distortion*: If algorithms are improperly designed or if appropriate data are lacking, distortion can be created that affects all users relying on these algorithms for developing policies.

⁹ YU, 342, mentions the adaptability as a fifth attribute; however, the adaptation rather constitutes an element of access and availability.

¹⁰ See below ch. [3.3](#).

¹¹ For a very detailed discussion see YU, 343 et seq. with many further references.

From a general perspective, intricate algorithms additionally tend to cause opaque decision-making processes and can lead to problems in respect of their interpretability and of human comprehensibility; ultimately, such a situation constitutes an undesirable “black box society”.¹²

3. Normative Principles for the AI Age

A comprehensive approach looking at the challenges of the AI age must mainly consider the most basic normative principles governing human interactions; hereinafter, the fundamental rule of law concept and the interdisciplinary relevant ethics-based equality concept will be discussed.

3.1. Rule of Law

3.1.1. Notion

The rule of law is an implied philosophy of modern positive law. Already Aristotle mentioned the principle that the guardians of the laws should also be obliged to obey the laws; Plato referred to a commonly acknowledged state of affairs, in which the legal system is normatively good in shape.¹³ As already outlined by Dicey in 1885, the rule of law encompasses the ascendancy of law as such and of the institutions in a system of governance that realizes a “rule of reason” under the given circumstances.¹⁴ The objective is to answer the following question:¹⁵ What are the standards and values of the rule of law and how do they guide the formulation of a “legitimate” legal rule?

The rule of law must be seen as a cross-cutting issue and an open-ended concept, however, as pointed out by the Venice Commission, a consensus should encompass six main elements,¹⁶ namely (i) legality, including a transparent, accountable and democratic process for enacting law, (ii) legal certainty, (iii) prohibition of arbitrariness, (iv) access to justice before independent and impartial courts, including traditional review of administrative acts, (v) respect for human rights, and (vi) non-discrimination and equality before

¹² This term was used by PASQUALE as title of his seminal book, published ten years ago.

¹³ A detailed discussion of the views of Greek philosophers (Aristotle, Plato) can be found in TAMANAHA, 3 et seq. and 8 et seq.

¹⁴ DICEY, *passim*, with a broad range of arguments.

¹⁵ Legal philosophers having analysed these questions are Hart, Raz, Allan, Bingham, and others; details of their theories cannot be discussed in this contribution.

¹⁶ Council of Europe, Venice Commission, 10 et seq.

the law. Notwithstanding these principles, the rule of law is an evolving and dynamic concept, not inherently static, thereby adapting to the changing needs and values of society. Therefore, the key question is how the emergence of information technology and the widespread use of AI and machine learning applications can unfold within the limits of the rule of law concept.¹⁷

The rule of law enables modern societies to realize a stable and transparent environment that can resolve conflicts between citizens within a community; the term “rule” indicates the possibility to become a guide or norm for decisions in doubtful or unforeseen cases.¹⁸ The rule can also contribute to the realization of the equality principle: Everybody should have access to new AI technologies and AI applications are to be designed in a way that inequality is not exacerbated.¹⁹

3.1.2. Substantive Aspects

The rule of law must be distinguished from the rule by law. The second notion enables authoritarian governments to enforce the introduced regulations disregarding the intended meaning of the rule of law; indeed, the rule by law is often misinterpreted as a concept of legality being an important constituent of the rule of law.²⁰ Legality stipulates rules that are introduced to echo the substantive legitimacy of the norms and certain ideals such as proportionality and safeguards. The rule of law is also not identical to the competence of the legislator being a political actor and not a legal actor.²¹ The design standards for legality that aim to achieve “good law” have been concretized by Fuller in the form of eight in detail discussed “desiderata”.²²

- Norms should be general and follow a principles-based approach (“reasoned generality”).
- Norms should be promulgated; consequently, society must be educated to understand the full meaning of the regulations that might conceivably be imposed on them.

¹⁷ To the identification of rule of law values see BERMAN, 1309 et seq.

¹⁸ See RAZ, 210, 218, and also TASOULIAS, 4–5.

¹⁹ A “dramatic” description of the inequality risks in a world full of data can be found in O’NEIL, *passim*; to the equality concept as emanation of the rule of law see below ch. [3.2.1](#).

²⁰ See also HILDEBRANDT, 178 et seq.

²¹ WINTGENS, 5 et seq.

²² For a detailed description of the eight desiderata see FULLER, 46–91.

- Norms should be prospective and not retrospective; regulations must be designed in a future-oriented manner.
- Norms should not be unclear; legality is not attainable by obscure and inherent legislation.
- Norms should not be contradictory; if regulations fight each other, legal certainty is not achievable.
- Norms should not require the impossible; such regulations are confronted with the risk of doing serious injustice or diluting respect for the law.
- Norms should be relatively constant; frequent changes in the regulations can lead to injustice and legal uncertainty.
- Norms should achieve a congruence between official action and declared rule, i.e. correspond to the internal morality of law; thereby, the integrity of a legal system can be secured.

By implementing the eight mentioned desiderata (normative standards), a rational interpretation of the substantive legal norms can be achieved. In the AI context, standard 1 is of major importance; only a principles-based, not a (detailed) rules-based approach is suitable to cope with technological developments. Compliance with standard 3 is obviously needed since rules must be prospective; standards 4 and 5 formulate generally accepted wisdom; standard 8 refers to the relevance of an interdisciplinary approach encompassing morality and ethics as discussed hereinafter.

3.1.3. Formal Aspects

The formal aspects of the rule of law concern its operation and the way it is performed; clearly, procedure matters in making legal rules legitimate.²³ Therefore, normative regulations can also contain settings, procedures and formal constraints that govern the behavior of individuals to reach compliance with the substantive contents of the rule of law. This assessment corresponds to the mentioned eighth desideratum of Fuller that calls for a reasonable congruence between official action and declared rule (“internal morality of law”).²⁴ Another value of the formal rule of law encompasses the “right to due process”; specific aspects encompass the principle of equality before law and

²³ TASOULIAS, 13.

²⁴ See above ch. [3.1.2.](#) and the respective consideration expressed by TASOULIAS, 5.

the right to be heard. By building a bridge between the substantive and the formal aspects of the rule of law, the equal treatment principle and the non-discrimination requirement can realize a clear normative framework avoiding that a lack of interpretability of rules causes challenges. The respective values stem from the idea that all members of civil society are subject to the same set of rules being implemented in a way which allows ensuring justice to all.²⁵

Furthermore, transparency, accountability, and explainability requirements should and can be designed as procedural rules. Transparency must be realized in respect of the automation processes and the functioning of algorithms incl. the use of training data; a lack of transparency and an inability to understand the processes erodes the trust in the decision-making rules and undermines accountability.²⁶ Based on a proper understanding of the relevant procedural steps, individuals must have the means to challenge automated decisions in legal proceedings.

3.2. Ethics-based Equality

3.2.1. Equality as Emanation of the Rule of Law and of Ethics

Ideally, the rule of law contributes to the realization of equal opportunities; scholars in legal and political philosophy already developed many concepts.²⁷ In theory, the equality principle can be assessed from a formal and a substantive angle.²⁸ The formal dimension of equal opportunities is based on the assumption that – if all positions are open and attainable – a reasonable allocation based on merits will occur; however, this perception does not consider that the starting chances are not identical for all individuals. Therefore, the material dimension of equality must also be realized, for example by securing access to and availability of AI applications to all.

Equality is not only a legal principle derived from the rule of law concept, but equally an emanation of ethics. On the one hand, the substantive contents of equal opportunities assess guidance by fundamental norms of law, on the

²⁵ The principle of ensuring “justice for all” is a policy objective of many international (UN) organizations and groups and a particular objective of the World Justice Project issuing a yearly “Rule of Law Index” (see <https://worldjusticeproject.org>).

²⁶ For a detailed analysis of the transparency, explainability and accountability principles in the AI context see WEBER, 2023, 71 et seq.

²⁷ The ground-breaking research was mainly done by DWORKIN, *passim*.

²⁸ ELFORD, no. 3 and no. 5.

other hand, an ethical behavior understood as morality element is important for the realization of equality objectives.²⁹

Inequality can occur in an environment that does not allow everybody to have access to AI applications and use the algorithms; if a part of the society is excluded from the exploitation of the relevant technologies, the concerned individuals are also deprived of the potential advantages and benefits offered to them.³⁰ A reason for such inequalities can lie in a system bias, for example caused by the sampling of the data or by the design and implementation of the algorithms.³¹ Consequently, the problem of a digital divide or an algorithmic divide, respectively, is becoming critical.

The widening of the digital divide might cause severe impacts on the equality principle since the rule of law also encompasses the equality of opportunities, rights and obligations as a significant component.³² With the attempt of bridging the different divides, the rule of law, guided by its fundamental principle of equality, necessitates that legislators and regulators take appropriate actions to level the playing field in respect of technology access and AI literacy, i.e. civil society must be familiarized with the knowledge and skills to understand, use, evaluate and ethically apply the relevant technologies.³³

3.2.2. Ethics: Contents and Standards

Already the term “ethos” as used in Ancient Greece encompasses two different facets based on two different spellings of the “e”, namely, on the one hand, ethos means habit and custom, and, on the other hand, ethos enshrines character and morals. Aristotle distinguished three ethical categories, namely “phronesis” (practical skills, wisdom), “arete” (virtue, goodness), and “eunoia” (goodwill). Ethos mirrors guiding beliefs or ideals governing the community and is concerned with the pursuit of the good life as its teleological goal.

In the understanding of Locke, the identity of an individual in the light of self-sustaining attributes relates to natural freedom. Rousseau proclaimed an analysis of transformation from the natural man to the social man. Kant developed a moral dignity concept of the individual. As a result, it can be said

²⁹ For further details see DWORKIN, 185 et seq.

³⁰ The respective attributes are discussed above (ch. [2.1](#)).

³¹ FERRARA, *passim*.

³² YU, 385 et seq.

³³ NG et al., no. 3.12 (for a discussion of the tools that enable assessing the AI literacy, see NG no. 3.11.); see also below ch. [3.2.3](#).

that ethics refers to the “right” or “wrong” conduct as part of the philosophy of human behavior.³⁴

Several attributes relate to ethics and need to be thoroughly considered in the implementation of algorithms and AI applications.³⁵

- *Justice/equity*: Every human being has an inalienable dignity and is entitled to equal rights including justice, fair and equal access, and enabling information flows.
- *Freedom*: Human dignity calls for the development of various freedoms.
- *Care and compassion*: The ability for empathy and respect leads to solidarity and reciprocal support.
- *Participation*: The right and ability to participate in societal life and in decision-making processes about matters of concern are major values.
- *Sustainability and responsibility*: Sustainability is important for the protection of a life-worthy environment, and the assumption of accountability/responsibility related to one’s own actions is a core requirement in a societal setting.

The different ethical values are interlinked and can be balanced with each other. In the context of the development of new technologies such as AI and machine learning, such kind of balancing requires compliance with high ethical standards inside and outside the design process. Those involved in algorithmic design should devote special attention to developing best practices of codes of conduct concerning how to avoid or alleviate the problems of algorithmic discrimination and distortion.³⁶ As a result, policymakers must develop ethical standards that are built into the design, usage, and improvement of algorithms.

Already in 2017, the Institute of Electrical and Electronics Engineers (IEEE) released a (second) draft of its guiding document entitled “Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems”.³⁷ This policy paper recommends that software engineers

³⁴ For further details to the history and the philosophical foundation of the ethics principles see WEBER, 2015a, 100 with references.

³⁵ This list is based on WEBER, 2015a, 101–02.

³⁶ See also YU, 368–69.

³⁷ Institute of Electrical and Electronics Engineers (IEEE), Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, 58 (2017), 68,

document all their systems and related data flows, their performance, limitations, and risks with emphasis on auditability, accessibility, meaningfulness, and readability.

3.2.3. Digital and Algorithmic Literacy

Digital literacy in the increasingly e-permeated world refers to the ability to effectively use digital tools to support innovation and advantageously manage digital resources; digital literacy may also serve as a stimulator for digitalized business and economic growth.³⁸ The concept of digital literacy is directly linked to the emergence of new technologies that fundamentally change the approach towards information and knowledge and must be understood as a reflection of this development. The notion of being able to understand and use information in multiple formats from a wide variety of sources involves the ability to effectively apply current digital/algorithmic tools and sources as well as electronic infrastructures in dealing with information.³⁹ In addition, digital literacy does not only denote technical, but also cognitive competence in the sense of comprehension of information and critical thinking being of particular importance in the AI context.

Overcoming the individuals' lack of awareness of algorithm-related problems and their inability to adapt to AI and machine learning, increasing algorithmic literacy is crucial if civil society should be able to reap the benefits of the new technologies.⁴⁰ The individuals will need to know not only the impact of AI and machine learning on their daily activities, but also what it means to live in a society driven heavily by algorithms.⁴¹ Consequently, for good reasons the EU parliament insisted in February 2024 to include an AI literacy provision into the Artificial Intelligence Act (now Art. 4 AIA).

Improved algorithmic literacy will help individuals realize the full potential of the new technological means. A greater understanding of algorithmic operations will allow individuals to develop human-related responses to ensure a more successful engagement with algorithm-enhanced technological products and services. In the meantime, it has become obvious that AI appli-

https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf; see also YU, 369.

³⁸ GILSTER, 6.

³⁹ WEBER, 2015b, no. 5.

⁴⁰ YU, 362–63.

⁴¹ GILSTER, 6; WEBER, 2015b, no. 8.

cations do have a major impact on the legal environment and profession;⁴² research has also shown that individuals are capable of developing responses to technological challenges.⁴³

3.3. Conclusion: Chances for Overcoming Digital and Algorithmic Divides

The objective of developing a more inclusive environment in the digital and algorithmic context can hardly be contested as desirable purpose of normative efforts in a “healthy” civil society. Therefore, regulatory activities should be based on the following general question: what will work better in the given environment, human algorithms or digital ones.⁴⁴ Quite obviously, digital and algorithmic divides have a negative impact on the right to equality of all civil society members and lead to a situation in which “have-nots” are deprived of equal opportunities allowing them to benefit from new information technology advancements.⁴⁵

A comprehensive legal and ethical framework should contribute to mitigating the widening of digital and algorithmic divides. Ensuring equal access to modern technologies (incl. AI applications) by civil society members and promoting AI literacy that enables an improved understanding of these technologies can help to bridge the undesirable divides. Enhanced comprehension based on transparency, explainability, and accountability in an AI environment paves the way to an effective human oversight of automated processes;⁴⁶ in case of a corresponding regulatory framework, trust is built in AI systems allowing human interventions if desired and/or needed.

4. Bibliography

ALLAN, T.R.S., *Constitutional Justice: A Liberal Theory of the Rule of Law*, Oxford 2003

BERMAN, EMILY, *A Government of Laws and Not of Machines*, 98 *Boston University Law Review* 2018, 1277–1355

BINGHAM, TOM, *The Rule of Law*, London 2011

COGLIANESE, CARY/LAI, ALICIA, *Algorithm v. Algorithm*, 71 *Duke Law Journal* 2022, 1281–1340

⁴² See MIGLIORINI/MOREIRA, 5 et seq.

⁴³ See also YU, 363–64.

⁴⁴ COGLIANESE/LAI, 1281.

⁴⁵ YU, 334.

⁴⁶ See WEBER, 2023, 71 et seq. and 75 et seq.

- Council of Europe, European Commission for Democracy Through Law (Venice Commission), Report on the Rule of Law, Strasbourg 2011, CLD-AD(2011)003rev
- DICEY, ALBERT V., Introduction to the Study of the Law of the Constitution (1885), 8th ed. London 1915 (reprint with Foreword by Roger A. Michener, London 1982)
- DWORKIN, RONALD, What is equality? Part 1: Equality of welfare, 10 Philosophy and Public Affairs 1981, 185–246
- ELFORD, GIDEON, Equality of Opportunity, in: Stanford Encyclopedia of Philosophy, 2023, <https://plato.stanford.edu/entries/equal-opportunity>
- FERRARA, EMILIO, Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies, MDPI, Sci 2014, 6(1), 3, <https://doi.org/10.3990/sci6010003>
- FULLER, LON L., The Morality of Law, 2nd ed. New Haven/London 1969
- GILSTER, PAUL, Digital literacy, New York 1997
- HART, H.L.A., The Concept of Law, Oxford 1961
- HILDEBRANDT, MIREILLE, A Vision of Ambient Law, in: Roger Brownsword/Karin Yeung (eds.), Regulating Technologies, Oxford 2008, 175–191
- LOYOLA-GONZÁLES, OCTAVIO, Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View, 7 IEEE Access 2019, 154096–154113
- MIGLIORINI SARA/MOREIRA, JOAO ILHAO, The Case of Nurturing AI Literacy in Law Schools, Asian Journal of Legal Education 2024, 1–18
- NG, DAVY T.K. et al., Conceptualizing AI literacy: An exploratory review, 2 Computers and Education: Artificial Intelligence 2021, 100041, <https://doi.org/10.1016/j.caeai.2021.100041>
- O'NEIL, CATHY, Weapons of Mass Destruction, New York 2016.
- PASQUALE, FRANK, The Black Box Society: The Secret Algorithms That Control Money and Information, Cambridge MA 2015
- RAZ, JOSEPH, The Rule of Law and Its Virtues, in: Joseph Raz, The Authority of Law. Essays on Law and Morality, Oxford 1979, 210–229
- SUSSKIND, RICHARD, Online Courts and the Virtue of Justice, Oxford 2019
- TAMANAH, BRIAN Z., On the Rule of Law: History, Politics, Theory, Cambridge 2004
- TASOULIAS, JOHN, The Rule of Algorithm and the Rule of Law, in: Vienna Lectures on Legal Philosophy 2023, <https://dx.doi.org/10.2139/ssrn.4319969>
- WEBER, ROLF H., Financial Inclusion as Public Policy Issue in the Global Digital Economy, 19 Asian Journal of WTO & International Health Law and Policy 2024, 231–252 (cit. WEBER, 2024)
- WEBER, ROLF H., Artificial Intelligence: A Theoretical Analysis of Regulatory Models, in: Luca Belli/Walter B. Gaspar (eds.), The Quest for AI Sovereignty, Transparency and Accountability, IGF Kyoto, Rio de Janeiro 2023, 71–83 (cit. WEBER, 2023)
- WEBER, ROLF H., Ethics as Pillar of Internet Governance, 23 Annual Review of Law and Ethics 2015, 95–111 (cit. WEBER, 2015a)
- WEBER, ROLF H., Digital Literacy as Incentive for Knowledge Management in the Digital World, Jusletter IT of 26 February 2015 (cit. W , 2015b)
- WINTGENS, LUC J., Legisprudence as a New Theory of Legislation, 19 Ratio Juris 2006, 1–25

YU, PETER K., The Algorithmic Divide and Equality in the Age of Artificial Intelligence, 72 Florida Law Review 2020, 331–389

Künstliche Intelligenz: Einzelthemen

Regulierung von Algorithmen in der EU und in der Schweiz

Überlegungen zu ausgewählten Regulierungsthemen

Rolf H. Weber / Simon Henseler*

Erstveröffentlichung in: EuZ, Ausgabe 2/2020, 28–42

Sowohl in der Europäischen Union als auch in der Schweiz entwickelten sich Algorithmen in den letzten Jahren – nicht zuletzt für Regulatoren – zu einem Hot Topic. Insbesondere der technische Fortschritt im Digitalisierungsbereich rief eine Debatte über die bestehenden Regulierungen von Algorithmen hervor und provozierte zuweilen den Ruf nach weiteren Normierungen. Doch was sind eigentlich die wesentlichen Risiken, die mit dem Einsatz von Algorithmen verbunden sind? Welche Regulierungsansätze und -konzepte bestehen bereits? Welche regulatorischen Defizite sind auszumachen? Und wie vermögen die Regulatoren effizient und effektiv einzugreifen?

Inhaltsübersicht

I.	Einleitung	104
II.	Begriff «Algorithmus» und Erscheinungsformen	105
III.	Risiken von Algorithmen	106
1.	Intransparenz	107
2.	Unfairness	108
3.	Diskriminierung	109
IV.	Regulierungskonzepte und -ansätze	111
1.	Bisherige Entwicklungen	111
2.	Regulatorische Mischformen	112
V.	Ausgewählte Regulierungsthemen	113
1.	Datenschutz	114
a)	Verbot automatisierter Entscheidungen	115
b)	Informationspflicht und Auskunftsrecht	116
c)	Datenschutz-Folgenabschätzung	116
d)	Defizite bei algorithmischen Systemen	117

* Simon Henseler, Dr. iur, Rechtsanwalt, Zürich.

2.	Schaffung von Transparenz	118
a)	Primär- bzw. verfassungsrechtliche Transparenzerfordernisse	118
b)	Sekundärrechtliche bzw. einfachgesetzliche Transparenzerfordernisse	118
c)	Transparenzerfordernisse im Soft Law	120
d)	Defizite bei algorithmischen Systemen	121
3.	Schutz vor Diskriminierungen	122
a)	Primär- bzw. verfassungsrechtlicher Diskriminierungsschutz	122
b)	Sekundärrechtlicher bzw. einfachgesetzlicher Diskriminierungsschutz	123
c)	Persönlichkeitsrechtlicher Diskriminierungsschutz	124
d)	Diskriminierungsschutz im Soft Law	125
e)	Defizite bei algorithmischen Systemen	126
4.	Exkurs: Finanzmarktrechtliche Bestimmungen	126
VI.	Fazit und Ausblick	127

I. Einleitung

Als wichtiges Element der modernen Informationsgesellschaft beeinflussen die Algorithmen schon heute täglich erhebliche Teile der Arbeit und der Freizeit. Mit der fortschreitenden Digitalisierung dürften sie immer weitere Lebensbereiche prägen und damit vielfältige Fragen zum gesellschaftlichen, ethischen und rechtlichen Umgang mit ihnen aufwerfen¹. Deshalb überrascht wenig, dass nicht nur inter- und supranationale Organisationen (z.B. die EU), sondern auch nationale Gesetzgeber und Anwender sich der Algorithmen regulatorisch annehmen. Gerade in der Wintersession 2019 waren Algorithmen, deren Risiken durch das totalrevidierte Datenschutzgesetz eingeschränkt werden sollen², wieder Thema im Eidgenössischen Parlament³.

¹ Vgl. für eine allgemeine Einführung Karen Yeung/Martin Lodge, *Algorithmic Regulation – An Introduction*, in: Yeung/Lodge (Hrsg.), *Algorithmic Regulation*, Oxford 2019, 1 ff.

² Botschaft des Bundesrates (BR) vom 15. September 2017 zum Bundesgesetz über die Totalrevision des Bundesgesetzes über den Datenschutz und die Änderung weiterer Erlasse zum Datenschutz, BBl 2017, 6941 ff., 7067; Stellungnahme des BR zum Postulat Schwaab, 16.4007; Stellungnahme des BR zum Postulat Béglé, 16.3914.

³ Am 18. Dezember 2019 beriet der Ständerat den Entwurf des neuen Datenschutzgesetzes (E-DSG) und beschloss im Vergleich zum Nationalrat einige Verschärfungen. Die Differenzvereinbarung wird für die Frühlingssession 2020 erwartet; ein Inkrafttreten des revidierten DSG per 2021 ist damit nicht unwahrscheinlich (vgl. David Vasella, SR: Beratung des E-DSG abgeschlossen, 18. Dezember 2019, daen:recht, <<https://datenrecht.ch/sr-beratung-des-e-dsg-abgeschlossen/>>).

Der vorliegende Aufsatz analysiert die Regulierungen von Algorithmen in der EU und in der Schweiz in ausgewählten Diskussionsbereichen und bietet einen Überblick zu den laufenden rechtlichen Entwicklungen. Nach Erläuterungen zum Begriff «Algorithmus» und zu dessen Erscheinungsformen thematisieren die nachfolgenden Ausführungen vorerst wesentliche Risiken, die mit dem Einsatz von Algorithmen verbunden sind. Um diesen Risiken zu begegnen, stehen verschiedene Regulierungskonzepte und -ansätze zur Verfügung, die kurz erörtert werden, bevor einzelne praktisch bedeutende Regulierungsthemen wie Datenschutz, Schaffung von Transparenz und Schutz vor Diskriminierungen zur Sprache kommen. Abschliessend folgt ein zusammenfassender Ausblick.

II. Begriff «Algorithmus» und Erscheinungsformen

Ein Algorithmus ist eine in Einzelschritte zerlegte Handlungsanweisung zur Lösung einer bestimmten Aufgabenstellung⁴. Demnach handelt es sich auch etwa bei einem Kochrezept um einen Algorithmus⁵. Die vorliegend interessierenden (Computer-)Algorithmen bestehen aus einer in Programmiersprache gegossenen endlichen Anzahl klar definierter Abarbeitungsschritte, die anschliessend für die Computerhardware in Maschinensprache übersetzt werden⁶.

Phänomenologisch lassen sich verschiedene Erscheinungsformen von Algorithmen unterscheiden. Deterministische Algorithmen zeichnen sich dadurch aus, dass dieselben Inputs auch immer zu denselben Outputs oder Entscheidungen führen, da die Entscheidungsprozesse sich nicht ändern⁷. Demgegenüber modifizieren selbstlernende Algorithmen (sog. *Machine Learning*) ihre Entscheidungsprozesse anhand der in der Lernphase verwendeten Trainings-

⁴ Mario Martini, Algorithmen als Herausforderung für die Rechtsordnung, JZ 2017, 1017 ff., 1017 Fn. 2; Corinne Widmer Lüchinger, Apps, Algorithmen und Roboter in der Medizin: Haftungsrechtliche Herausforderungen, HAVE 2019, 3 ff., 5 m.w.H.; Martin Zobl/Michael Lysakowski, E-Persönlichkeit für Algorithmen?, digma 2019, 42 ff., 42.

⁵ Vgl. zu diesem Beispiel Zobl/Lysakowski (Fn. 4), 42; Stefan Hofer, Algorithmen, maschinelles Lernen und die Grenzen der KI, Jusletter, 26. November 2018, Rz. 1 f.

⁶ Expertengruppe zur Zukunft der Datenbearbeitung und Datensicherheit, Bericht, 17. August 2018, <<https://www.news.admin.ch/newsd/message/attachments/53591.pdf>>, 38 (zit. Expertengruppe Datenbearbeitung); vgl. auch Martini (Fn. 4), 1017 Fn. 2, wonach die rechtswissenschaftliche Diskussion stets den «Computeralgorithmus» vor Augen hat.

⁷ So Expertengruppe Datenbearbeitung (Fn. 6), 38; vgl. auch Mario Martini, Blackbox Algorithmen – Grundfragen einer Regulierung Künstlicher Intelligenz, Berlin 2019, 19.

daten; gleiche Inputs führen so je nachdem zu unterschiedlichen Outputs⁸. Selbstlernende Algorithmen, die zur Technologie der Künstlichen Intelligenz (KI) zählen⁹, sind darauf ausgerichtet, statistische Muster in Datensätzen zu erkennen und sich diese zunutze zu machen¹⁰.

Immer verbreiteter wird der Einsatz von Algorithmen, die menschliche Entscheidungen vorbereiten, unterstützen oder gar ersetzen. Das algorithmische System trifft bspw. eine Vorselektion im Bewerbungsverfahren, beurteilt das Krankheitsbild eines Patienten oder entscheidet über die Kreditwürdigkeit eines Verbrauchers¹¹. In diesem Zusammenhang wird bisweilen von Entscheidungsalgorithmen gesprochen¹², obschon ein Algorithmus («im ethisch gehaltvollen Sinne») selbst keine Entscheidung treffen kann¹³.

III. Risiken von Algorithmen

Der Einsatz von Algorithmen bietet zweifellos Chancen für Wirtschaft, Gesellschaft und Verwaltung. Es lässt sich denn auch nicht sagen, dass Algorithmen

⁸ Expertengruppe Datenbearbeitung (Fn. 6), 38; vgl. auch Nadja Braun Binder, Künstliche Intelligenz und automatisierte Entscheidungen in der öffentlichen Verwaltung, SJZ 2019, 467 ff., 473.

⁹ Hofer (Fn. 5), Rz. 2; Gesellschaft für Informatik, Technische und rechtliche Betrachtungen algorithmischer Entscheidungsverfahren, Studien und Gutachten im Auftrag des Sachverständigenrats für Verbraucherfragen, Berlin, <http://www.svr-verbraucherfragen.de/wp-content/uploads/GI_Studie_Algorithmenregulierung.pdf>, 30.

¹⁰ Claude Castelluccia/Daniel Le Métayer, Understanding algorithmic decision-making: Opportunities and challenges, Study for the European Parliament, Brüssel 2019, 41; Peter Georg Picht/Benedikt Freund, Wettbewerbsrecht auf algorithmischen Märkten, sic! 2018, 666 ff., 668.

¹¹ Vgl. die Beispiele aus unterschiedlichen Anwendungsfeldern bei Rolf H. Weber, Automatisierte Entscheidungen: Perspektive Grundrechte, SZW 2020, 18, 21 ff. (vgl. [Text 4, S. 47 ff.](#)); AlgorithmWatch, Atlas der Automatisierung, Automatisierte Entscheidungen und Teilhabe in Deutschland, Berlin 2019, <https://atlas.algorithmwatch.org/wp-content/uploads/2019/04/Atlas_of_Automation_by_AlgorithmWatch.pdf>, 26 ff.

¹² So Expertengruppe Datenbearbeitung (Fn. 6), 170.

¹³ Datenethikkommission der Bundesregierung, Gutachten, 23. Oktober 2019, <https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf?__blob=publicationFile&v=5>, 161 (zit. Datenethikkommission); vgl. auch Florent Thouvenin et al., Towards Principled Regulation of Automated Decision-Making (ADM) – A Workshop Report, 2019, <<https://www.itsl.uzh.ch/dam/jcr:ed-ba006c-8452-4ffc-bbb0-fd8fde62d114/Workshop%20Report%20Lavin.pdf>>, 1 f.

im Vergleich zu Menschen Entscheidungen von schlechterer Qualität treffen¹⁴. Im Gegenteil dürften algorithmische Systeme oftmals schneller, kosteneffizienter und weniger fehleranfällig sein, als wenn ein Mensch dieselbe Aufgabe erledigen müsste – und dies erst noch ermüdungsfrei¹⁵.

Wichtiger als die Chancen sind im Hinblick auf die Regulierung von Algorithmen aber die Risiken, die mit deren Einsatz verbunden sind. Nachfolgend wird deshalb auf drei zentrale Risiken eingegangen; eine umfassende Behandlung der Risiken kann der vorliegende Beitrag nicht leisten¹⁶.

1. Intransparenz

Intransparenz vermag zunächst hinsichtlich des Einsatzes von Algorithmen selbst zu bestehen; eine Person ist von der Verwendung eines algorithmischen Systems betroffen, ohne dass sie darüber Bescheid weiss. Dies kann beim Einsatz von Social- oder Chatbots dazu führen, dass aufgrund der Verwechslungsgefahr zwischen Mensch und Maschine der vertrauensvolle Umgang miteinander in der Gesellschaft leidet¹⁷. Ausserdem erweist sich dies als problematisch, weil der betroffenen Person so verwehrt bleibt, die ihr zustehenden – bspw. datenschutzrechtlichen – Rechte auszuüben¹⁸. Weiss z.B. eine Arbeitssuchende nicht, dass ein Algorithmus ihre Bewerbung aussortierte, ohne dass ein Mensch diese je beurteilte, bleibt es ihr verwehrt, ein (allfälliges) Auskunfts- oder Anfechtungsrecht geltend zu machen.

¹⁴ Florent Thouvenin/Alfred Früh, *Automatisierte Entscheidungen: Grundfragen aus der Perspektive des Privatrechts*, SZW 2020, 3, 9. Diese Annahme lag den ersten Regulierungen von algorithmischen Entscheidungen zugrunde (Thouvenin et al. [Fn. 13], 2; vgl. auch Florent Thouvenin/Alfred Früh/Damian George, *Datenschutz und automatisierte Entscheidungen*, Jusletter, 26. November 2018, Rz. 11).

¹⁵ *Expertengruppe Datenbearbeitung* (Fn. 6), 38; vgl. auch *Council of Europe, Committee of experts on internet intermediaries*, Study on the Human Rights Dimensions of Automated Data Processing Techniques (in particular Algorithms) and Possible Regulatory Implications, Strassburg 2018, 26; Picht/Freund (Fn. 10), 668; Widmer Lüchinger (Fn. 4), 4; Zobl/Lysakowski (Fn. 4), 43.

¹⁶ Vgl. für eine weiter- und tiefergehende Untersuchung der mit Algorithmen verbundenen Risiken statt vieler Castelluccia/Le Métayer (Fn. 10), 7 ff.; *Council of Europe* (Fn. 15), 11 ff.; Martini (Fn. 7), 27 ff.; Guido Noto la Diega, *Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information*, JIPITEC 1/2018, Rz. 10 ff.

¹⁷ *Datenethikkommission* (Fn. 13), 185; vgl. auch *Hochrangige Expertengruppe*, eingesetzt von der Europäischen Kommission, *Ethik-Leitlinien für eine vertrauenswürdige KI*, Brüssel 2019, <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60425>, Rz. 78.

¹⁸ Martini (Fn. 4), 1020; vgl. auch Martini (Fn. 7), 176; Thouvenin et al. (Fn. 13), 2.

Akzentuiert ist das Risiko der Intransparenz bei selbstlernenden Algorithmen, bei denen je nach angewendeter Methode des *Machine Learning* angesichts des heutigen Standes der Technik zuweilen selbst der Betreiber des Systems den Lösungsweg nicht mehr nachvollziehen kann¹⁹. Für die betroffene Person lässt sich entsprechend noch viel weniger rekonstruieren, welche Daten anhand welcher (sich anpassender) Variablen mit welcher (wechselnden) Gewichtung zum sie betreffenden Output geführt haben²⁰.

2. Unfairness

Algorithmische Systeme können in mannigfaltiger Hinsicht als unfair wahrgenommene Outputs erzeugen. Dies rührt nicht zuletzt aus der Vagheit der Begriffe «Unfairness» bzw. «Fairness», zu denen die Lehre verschiedene Begriffsverständnisse entwickelt hat²¹. Da sich einzelne als unfair erachtete Fallgruppen je nach den Umständen mit dem Risiko der Diskriminierung überschneiden, rechtfertigt es sich, diese vorab anzusprechen.

So gilt als unfair, wenn die Vorhersageleistung des Systems – etwa ausgedrückt durch den Anteil an *False Positives* und *False Negatives*²² – bei bestimmten Gruppen, insbesondere solchen mit vom Diskriminierungsverbot geschützten Merkmalen, erheblich schlechter ausfällt²³. Ebenso als unfair bezeichnet die Lehre bisweilen algorithmische Systeme, die auf solche geschützte Merkmale und/oder auf deren *Proxies*²⁴ abstellen²⁵.

¹⁹ Expertengruppe Datenbearbeitung (Fn. 6), 39; Datenethikkommission (Fn. 13), 169; Picht/Freund (Fn. 10), 668.

²⁰ Martini (Fn. 7), 28 f.; Expertengruppe Datenbearbeitung (Fn. 6), 39; vgl. auch Datenethikkommission (Fn. 13), 169.

²¹ Castelluccia/Le Métayer (Fn. 10), 43; Sam Corbett-Davies/Sharad Goel, The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning, 14. August 2018, <<https://arxiv.org/pdf/1808.00023.pdf>>, 5 f., je m.w.H.

²² Ein *False Positive* liegt z.B. vor, wenn das System eine nicht kreditwürdige Person als kreditwürdig einstuft; ein *False Negative* ist gegeben, wenn das System eine tatsächlich kreditwürdige Person als nicht kreditwürdig einstuft (vgl. zu diesem Beispiel Gesellschaft für Informatik [Fn. 9], 38).

²³ Castelluccia/Le Métayer (Fn. 10), 43; Corbett-Davies/Goel (Fn. 21), 5 f.; Gesellschaft für Informatik (Fn. 9), 39 ff., je m.w.H.

²⁴ Bei *Proxies* (oder Proxivariablen) handelt es sich um «harmlos erscheinende Attribute, die das geschützte Attribut ganz oder teilweise beinhalten», wie z.B. den Wohnort oder die Postleitzahl, die stark mit dem ethnischen Hintergrund korrelieren können (Gesellschaft für Informatik [Fn. 9], 36).

²⁵ Corbett-Davies/Goel (Fn. 21), 5; Castelluccia/Le Métayer (Fn. 10), 43; Thouvenin et al. (Fn. 13), 3. Solche geschützte Merkmale sind u.a. Herkunft, Rasse, Geschlecht, Alter etc. (vgl. Art. 8

Als unfair bezeichnet werden ausserdem algorithmische Systeme, die – statt auf der fundierten Untersuchung der relevanten Kausalitäten – auf sachfremden Korrelationen basieren²⁶. Ursprung dessen ist die Funktionsweise der Systeme, die darauf angelegt ist, mittels statistischer Analysen von Massendaten Korrelationen und Wechselwirkungen zu erkennen²⁷. Als Beispiel mag das Verfahren einer Bank dienen, einen Kreditantrag nicht aufgrund der Kreditwürdigkeit der Antragstellerin abzulehnen, sondern aufgrund ihrer Wohnadresse in einem Quartier mit einer vergleichsweise hohen Ausfallrate bei der Rückzahlung²⁸.

3. Diskriminierung

Diskriminierungen (in einem weiten Sinne verstanden²⁹) sind ein vieldiskutiertes Risiko algorithmischer Systeme³⁰. Ein bekanntes Beispiel dafür ist die in den USA eingesetzte Software COMPAS, welche die Rückfallgefahr von Straftätern ermittelt³¹. Gemäss einer Untersuchung ist die Abweichung zwischen der generierten Prognose und der tatsächlichen Rückfallhäufigkeit bei Dunkelhäutigen doppelt so hoch wie bei Weissen³². Zwar ist ohne Weiteres vorstellbar, dass auch Menschen, die mit der Entscheidungsfällung betraut sind, eine solche diskriminierende Haltung an den Tag legen. Allerdings haben sie

Abs. 2 Bundesverfassung der Schweizerischen Eidgenossenschaft [BV] vom 18. April 1999, SR 101).

²⁶ Thouvenin et al. (Fn. 13), 3; vgl. auch Council of Europe (Fn. 15), 37; Thouvenin/Früh (Fn. 14), 5 f.

²⁷ Frederik Zuiderveen Borgesius, Discrimination, artificial intelligence, and algorithmic decision-making, Strassburg 2018, 9 f.; vgl. auch Expertengruppe Datenbearbeitung (Fn. 6), 37.

²⁸ Thouvenin/Früh (Fn. 14), 5 f.; Thouvenin et al. (Fn. 13), 2 f.; Martini (Fn. 4), 1018.

²⁹ Der Begriff der Diskriminierung wird teilweise in der wissenschaftlichen Auseinandersetzung mit Algorithmen im Vergleich zum herkömmlichen Verständnis (von bspw. Art. 8 Abs. 2 BV) breiter verstanden und beschränkt sich nicht zwingend auf die Ungleichbehandlung aufgrund geschützter Merkmale.

³⁰ Statt vieler Castelluccia/Le Métayer (Fn. 10), 7 ff.; Lilian Edwards/Michael Veale, Slave to the Algorithm?, Why a Right to an Explanation Is Probably not the Remedy You Are Looking For, DLTR 1/2017, 1 ff., 27 ff.; Zuiderveen Borgesius (Fn. 27), 10 ff.; Weber (Fn. 11), 19 f., je m.w.H.

³¹ COMPAS steht für «Correctional Offender Management Profiling for Alternative Sanctions» (vgl. dazu Urteil des Supreme Court of Wisconsin, 881 N.W.2d 749 [Wis. 2016], Loomis/Wisconsin; Zuiderveen Borgesius [Fn. 27], 14 f.). Vgl. für weitere Beispiele Martini (Fn. 4), 1018; Weber (Fn. 11), 21 ff.

³² Julia Angwin et al., Machine Bias, There's software used across the country to predict future criminals. And it's biased against blacks., <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>>; Martini (Fn. 4), 1018.

eine ungleich kleinere Breitenwirkung als wenn algorithmische Systeme zu einer skalenmässig grossen Anwendung gelangen³³.

Aus technischer Sicht gibt es verschiedene Gründe, weshalb selbstlernende algorithmische Systeme zu Diskriminierungen führen können³⁴. Solche Gründe vermögen auf einer willentlichen Entscheidung der Systementwickler zu beruhen, indem etwa das verwendete Modell deren Vorurteile bewusst spiegelt; dies muss indessen nicht der Fall sein³⁵. Selbst korrekte und grundsätzlich geeignete Trainingsdaten können dazu führen, dass das System historisch gewachsene Vorurteile und entsprechende Diskriminierungen übernimmt³⁶. Dies ist besonders problematisch, weil das algorithmische System dadurch die diskriminierende Haltung perpetuiert und sich so gesellschaftlichen Veränderungen verschliesst³⁷. Ein bedenkliches Diskriminierungsrisiko besteht auch darin, dass das System bei der eigenständigen Mustererkennung – statt direkt an geschützten Merkmalen – an Proxies, wie etwa dem Wohnquartier, anknüpft³⁸ und so allenfalls gar unerkannt diskriminiert³⁹.

³³ Datenethikkommission (Fn. 13), 167.

³⁴ Vgl. dazu ausführlich Solon Barocas/Andrew D. Selbst, Big Data's Disparate Impact, Cal. L. Rev. 2016, 671 ff., 677 ff.; Natalia Criado/Jose M. Such, Digital Discrimination, in: Yeung/Lodge (Hrsg.), Algorithmic Regulation, Oxford 2019, 82 ff., 85 ff.; Gesellschaft für Informatik (Fn. 9), 34 ff.; Alexander Tischbirek, Artificial Intelligence and Discrimination: Discrimination Against Discriminatory Systems, in: Wischmeyer/Rademacher (Hrsg.), Regulating Artificial Intelligence, Cham 2020, 103 ff., Rz. 5 ff., je m.w.H.

³⁵ Castelluccia/Le Métayer (Fn. 10), 7; Datenethikkommission (Fn. 13), 167; vgl. auch Rolf H. Weber, Digitalisierung und der Kampf ums Recht, in: Dal Molin-Kränzlin/Schneuwly/Stojanovic (Hrsg.), Digitalisierung – Gesellschaft – Recht, Zürich/St. Gallen 2019, 3 ff., 17; Martini (Fn. 4), 1018.

³⁶ Braun Binder (Fn. 8), 374, mit dem Beispiel des österreichischen Arbeitsmarktservices, für den ein Algorithmus Personen nach ihren Chancen auf dem Arbeitsmarkt beurteilt und dabei die Arbeitsmarktchancen von Frauen pauschal als schlecht einstuft; Weber (Fn. 11), 21 ff. Vgl. für weitere Beispiele Expertengruppe Datenbearbeitung (Fn. 6), 38 f.; Datenethikkommission (Fn. 13), 167.

³⁷ Thouvenin et al. (Fn. 13), 3; Expertengruppe Datenbearbeitung (Fn. 6), 170; vgl. auch Datenethikkommission (Fn. 13), 167; Martini (Fn. 4), 1018.

³⁸ Datenethikkommission (Fn. 13), 194; Council of Europe (Fn. 15), 26 f.; vgl. auch Barocas/Selbst (Fn. 34), 691 f.; Joshua A. Kroll et al., Accountable Algorithms, U. Pa. L. Rev. 2017, 633 ff., 685.

³⁹ Edwards/Veale (Fn. 30), 28; Toon Calders/Žliobaitė Indrė, Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures, in: Custers et al. (Hrsg.), Discrimination and Privacy in the Information Society, Berlin/Heidelberg 2013, 43 ff., *passim*; vgl. auch Tischbirek (Fn. 34), Rz. 12 f.

IV. Regulierungskonzepte und -ansätze

1. Bisherige Entwicklungen

Bei der Thematik der Regulierung von Algorithmen stellt sich vorerst die Frage, «wer» reguliert. Gerade in Bereichen, die sich durch starke technologische Entwicklungen auszeichnen oder neue Themenbereiche eröffnen, lehrt die Erfahrung, dass die traditionellen Gesetzgebungsprozesse von Staaten nicht immer die einzig sinnvolle Vorgehensalternative darstellen. Einerseits dauert die Ausarbeitung und Verabschiedung von Gesetzen oft zu lange (sog. *Regulatory Lag*), andererseits sind grenzüberschreitende Regulierungen erwünscht.

Im Bereich der Algorithmen gibt es bisher kaum umfassende einzelstaatliche Gesetze; geregelt sind bestenfalls Einzelaspekte (z.B. automatisierte Entscheidungen im Datenschutzrecht). Internationale Verträge sind ebenfalls nicht vorhanden, weil deren Verhandlungen zu lange dauern würden. Indessen haben internationale Organisationen, insbesondere im Kontext der KI, verschiedene Richtlinien ausgearbeitet, die von eigentlichen Gesetzgebern, aber auch von öffentlichen Stellen und privaten Unternehmen, übernommen werden können.

Als Beispiele lassen sich nennen:

- Empfehlung des Rats der OECD zu künstlicher Intelligenz (Mai 2019)⁴⁰;
- Deklaration des Ministerkomitees des Europarats über die manipulativen Fähigkeiten algorithmischer Prozesse (Februar 2019)⁴¹;
- Ethik-Leitlinien für eine vertrauenswürdige KI der hochrangigen Expertengruppe, eingesetzt von der Europäischen Kommission (April 2019)⁴².

Neben den stärker staatlich orientierten Richtlinien haben auch private Unternehmen in letzter Zeit Leitprinzipien für den angemessenen Einsatz mit Algorithmen entwickelt. So liegen etwa Leitlinien von grossen US-amerikanischen

⁴⁰ Rat der OECD, Empfehlungen des Rats zu künstlicher Intelligenz, verabschiedet am 22. Mai 2019, <<http://www.oecd.org/berlin/presse/Empfehlung-des-Rats-zu-kuenstlicher-Intelligenz.pdf>>.

⁴¹ Ministerkomitee des Europarates, Deklaration über die manipulativen Fähigkeiten algorithmischer Prozesse, verabschiedet am 13. Februar 2019, <https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b>.

⁴² Hochrangige Expertengruppe (Fn. 17). Vgl. dazu Barbara Widmer, Quo vadis KI?, Neue OECD-Grundsätze, digma 2019, 226 ff.

Tech-Unternehmen wie Google⁴³ und Microsoft⁴⁴ mit eigenen Prinzipien für den Umgang mit KI vor.

2. Regulatorische Mischformen

Traditionell hatten rechtliche Normen grundsätzlich den Charakter von Gesetzen oder Ausführungserlassen staatlicher Organe. Dieses Konzept ist in den letzten Jahren erheblichen Veränderungen ausgesetzt gewesen. Neben dem *Hard Law* hat sich das *Soft Law* durchgesetzt, insbesondere in der Form von Selbstregulierungen. Die Lehre hat auch anerkannt, dass *Soft Law* ebenso wirkungsvoll sein kann wie *Hard Law*⁴⁵. Dies hat damit zu tun, dass die von Interessengruppen selbst aufgestellten Regeln auf wirkliche Bedürfnisse eingehen und die aktuellen Technologien besser abbilden. Zudem genießen die Regeln regelmässig relativ breite Akzeptanz, weil die Betroffenen sie selbst ausgehandelt haben⁴⁶.

Als neue Form ist in jüngerer Zeit die sog. *Co-Regulation*, die eine Vielzahl hybrider regulatorischer Phänomene umfasst, in Erscheinung getreten. Sie zeichnet sich dadurch aus, dass Private und öffentliche Behörden zusammenwirken, um eine private Aktivität (auch) im Sinne öffentlicher Interessen zu regulieren⁴⁷. Bspw. erarbeiten Unternehmen eigenständig Regulierungsmechanismen, die eine demokratisch legitimierte Regulierungsbehörde genehmigt und sodann auf ihre Effektivität hin überwacht⁴⁸. *Co-Regulation* ermöglicht demnach, die Vorzüge der Selbstregulierung wie insbesondere die einfache Adaptierbarkeit durch die staatliche Involvierung mit einer gewissen Robustheit und erhöhten demokratischen Legitimation zu kombinieren⁴⁹.

⁴³ Google, AI at Google: our principles, 7. Juni 2018, <<https://www.blog.google/technology/ai/ai-principles/>>.

⁴⁴ Microsoft, Microsoft AI Principles, <<https://www.microsoft.com/en-us/ai/our-approach-to-ai>>.

⁴⁵ Rolf H. Weber, Overcoming the Hard Law/Soft Law Dichotomy in Times of (Financial) Crises, *Journal of Governance and Regulation* 2012, 8 ff., 11 m.w.H.; Jacob E. Gersen/Eric A. Posner, Soft Law: Lessons from Congressional Practice, *Stan. L. Rev.* 2008, 573 ff., 601.

⁴⁶ Weber (Fn. 45), 12; vgl. auch Michèle Finck, Digital Co-Regulation: Designing a Supranational Legal Framework for the Platform Economy, *ELJ* 2018, 47 ff., 53.

⁴⁷ Finck (Fn. 46), 58; vgl. auch Chris Marsden/Trisha Meyer/Ian Brown, Platform values and democratic elections: How can the law regulate digital disinformation?, 36 *CLSR* 2020, 105373, 1, 8.

⁴⁸ Marsden/Meyer/Brown (Fn. 47), 9 ff. Vgl. für Beispiele der *Co-Regulation* im Bereich der Plattform-Wirtschaft Finck (Fn. 46), 59 f.

⁴⁹ Vgl. Marsden/Meyer/Brown (Fn. 47), 15; Finck (Fn. 46), 62.

Die Parallelität von traditionellen Rechtsnormen und weiteren Rechtsinstrumenten zeigt sich im Bereich der Algorithmen und der KI besonders deutlich. So besteht mit dem Datenschutzrecht⁵⁰ zunächst eine staatliche Anordnung, die einen sektor- und bereichs-unabhängigen Ansatz verfolgt. Überlagert wird diese horizontale Regulierung etwa durch punktuelle, nicht datenschutz-rechtsorientierte Transparenzernfordernisse sowie breit gefasste Transparenzvorgaben im *Soft Law*⁵¹. Hinzu kommen spezifische, auf bestimmte Lebensbereiche ausgerichtete gesetzliche Diskriminierungsverbote, die von Vorgaben zum Schutz vor Diskriminierungen des *Soft Law* flankiert sind⁵². In der Realität noch wenig durchgesetzt haben sich bisher indessen Standardisierungen, die auf einer *Co-Regulation* beruhen; Ansätze bestehen etwa im Bereich der Plattform-Wirtschaft und der Sozialen Medien⁵³. Von nicht zu unterschätzender Bedeutung sind hingegen die Deklaration des Europarates und die Empfehlungen der OECD sowie die Leitlinien der von der EU eingesetzten Expertengruppe, die ein Vakuum zu füllen vermögen, wenn es an *Hard Law* fehlt⁵⁴.

V. Ausgewählte Regulierungsthemen

Der Einsatz algorithmischer Systeme kann verschiedenste Rechtsgebiete betreffen, etwa das Arbeitsrecht, wenn Algorithmen im Rahmen von *Workforce Analytics* eingesetzt werden⁵⁵, das Haftungsrecht, wenn ein Algorithmus einen Schadensfall verursacht⁵⁶, oder das Kartellrecht, wenn Algorithmen zu einem wettbewerbswidrigen Verhalten führen⁵⁷. Auf diese Normenkomplexe, welche die Regulierung von Algorithmen nicht spezifisch vor Augen haben, soll nachfolgend – mit Ausnahme der Diskriminierungsverbote⁵⁸ – nicht eingegangen werden. Mit Blick auf die zuvor thematisierten Risiken, die sich beim Einsatz

⁵⁰ Dazu unten [V.1.](#)

⁵¹ Dazu unten [V.2.](#)

⁵² Dazu unten [V.3.](#)

⁵³ Vgl. dazu Marsden/Meyer/Brown (Fn. 47), *passim*; Finck (Fn. 46), *passim*.

⁵⁴ Dazu unten [V.2.c\)](#) und [V.3.d\)](#).

⁵⁵ Vgl. dazu etwa Raquel Pais/Noémie Ammann, Big Data am Arbeitsplatz, *AJP* 2019, 1095 ff., *passim*; Isabelle Wildhaber, Robotik am Arbeitsplatz: Robo-Kollegen und Robo-Bosse, *AJP* 2017, 213 ff., *passim*.

⁵⁶ Vgl. dazu etwa Zobl/Lysakowski (Fn. 4), 43 (Verkehrsunfall); Widmer Lüchinger (Fn. 4), 3 (Fehldiagnose).

⁵⁷ Vgl. dazu etwa *Expertengruppe Datenbearbeitung* (Fn. 6), 40 und 100; Picht/Freund (Fn. 10), 669.

⁵⁸ Diese Ausnahme rechtfertigt sich damit, dass Diskriminierungen ein zentrales Risiko algorithmischer Systeme darstellen; dazu oben [III.3.](#)

von Algorithmen ergeben können⁵⁹, konzentrieren sich die nachfolgenden Ausführungen vielmehr auf drei ausgewählte, praktisch bedeutende Regulierungsthemen. Als Exkurs werden schliesslich einige algorithmen-spezifische finanzmarktrechtliche Bestimmungen thematisiert.

1. Datenschutz

Sofern die algorithmischen Systeme Personendaten bearbeiten⁶⁰, ist das Datenschutzrecht anwendbar⁶¹; in einer solchen Situation muss der Verantwortliche z.B. die datenschutzrechtlichen Bearbeitungsgrundsätze⁶² einhalten. Neben den allgemeinen Grundsätzen enthält das Datenschutzrecht besondere Bestimmungen, die für algorithmische Systeme relevant werden können. Da diese – insbesondere im Zusammenhang mit der DSGVO⁶³ – in der Lehre schon ausführlich diskutiert worden sind⁶⁴, be-

⁵⁹ Dazu oben [III.](#)

⁶⁰ Art. 2 Abs. 1, Art. 4 Nr. 1 und 2 DSGVO; Art. 2 Abs. 1, Art. 4 lit. a und d E-DSG. Da erst das revidierte DSG Bestimmungen zu automatisierten Entscheidungen enthalten wird, ist vorliegend auf den Entwurf in der Fassung gemäss Beschluss des Ständerats vom 18. Dezember 2019, <<https://www.parlament.ch/centers/eparl/curia/2017/20170059/S3-22%20D.pdf>>, zu verweisen.

⁶¹ Es gibt durchaus problematische Anwendungen von Algorithmen, die ohne die Bearbeitung von Personendaten auskommen, wie z.B. gewisse Formen des *Predictive Policing* (vgl. *AlgorithmWatch* [Fn. 11], 38; *Zuiderveen Borgesius* [Fn. 27], 24 f.; *Commission nationale de l'informatique et des libertés* [CNIL], *Comment permettre à l'homme de garder la main?*, Dezember 2017, <https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf>, 46).

⁶² Art. 5–11 DSGVO; Art. 5 E-DSG. Mit Blick auf die Risiken (dazu oben [III.](#)) ist insb. der Transparenzgrundsatz (Art. 5 Abs. 1 lit. a DSGVO; Art. 5 Abs. 3 E-DSG) bei algorithmischen Systemen von Bedeutung (vgl. *Zuiderveen Borgesius* [Fn. 27], 21).

⁶³ Verordnung (EU) 2016/679 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung), ABl L 119 vom 4. Mai 2016, 1.

⁶⁴ Zur DSGVO statt vieler *Stephan Dreyer/Wolfgang Schulz*, Was bringt die Datenschutz-Grundverordnung für automatisierte Entscheidungssysteme?, *Potenziale und Grenzen der Absicherung individueller, gruppenbezogener und gesellschaftlicher Interessen*, Gütersloh 2018; *Edwards/Veale* (Fn. 30), 27 ff.; *Isak Mendoza/Lee A. Bygrave*, The Right Not to be Subject to Automated Decisions Based on Profiling, in: *Synodinou/Jougleux/Markou/Prastitou* (Hrsg.), *EU Internet Law – Regulation and Enforcement*, Cham 2017, 77 ff.; *Wachter et al.*, Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation, *IDPL* 2017, 76 ff. Zum E-DSG *David Rosenthal*, Der Vorentwurf für ein neues Datenschutzgesetz: Was er bedeutet, *Jusletter*, 20. Februar 2017, Rz. 58 ff.; *David Rosenthal*, Der Entwurf für ein neues Datenschutzgesetz, Was uns erwartet und was

schränken sich die nachfolgenden Ausführungen auf eine zusammenfassende Übersicht.

a) Verbot automatisierter Entscheidungen

Die DSGVO verwendet den Begriff «Algorithmus» nicht; sie knüpft hinsichtlich algorithmischer System vielmehr an den Begriff der «automatisierten Entscheidungen» an. Art. 22 DSGVO enthält gemäss überwiegender Lehre ein grundsätzliches Verbot solcher Entscheidungen⁶⁵, das indes nur greift, wenn die Entscheidung⁶⁶ ausschliesslich auf einer automatisierten Verarbeitung beruht⁶⁷. Das Erfordernis der Vollautomatisierung bedeutet – neben den Ausnahmen des Verbots in Art. 22 Abs. 2 DSGVO – eine erhebliche Einschränkung des Anwendungsbereichs, denn algorithmische System werden vielfach «nur» für die Unterstützung menschlicher Entscheidungen eingesetzt⁶⁸. Im totalrevidierten Datenschutzgesetz wird auch das Schweizer Recht automatisierte Entscheidungen normieren (insbesondere Art. 19 E-DSG), diese aber nicht grundsätzlich verbieten.

noch zu korrigieren ist, Jusletter, 27. November 2017, Rz. 100 ff.; Thowenin/Früh (Fn. 14), *passim*.

⁶⁵ Statt vieler Artikel-29-Datenschutzgruppe, Leitlinien zu automatisierten Entscheidungen im Einzelfall einschliesslich Profiling für die Zwecke der Verordnung 2016/679, 6. Februar 2018, 21 und 29 ff.; Dreyer/Schulz (Fn. 64), 9; Mario Martini, in: Paal/Pauly (Hrsg.), Beck'sche Kompakt-Kommentare, Datenschutz-Grundverordnung, Bundesdatenschutzgesetz, 2. Aufl., München 2018, Art. 22 N 1 und 29; Mendoza/Bygrave (Fn. 64), 87; a.M. (geltend zu machendes Individualrecht) etwa Wulf Kamlah, in: Plath (Hrsg.), DSGVO/BDSG Kommentar, 3. Aufl., Köln 2018, Art. 22 N 4; Anton Vedder/Laurens Naudts, Accountability for the use of algorithms in a big data environment, IRLCT 2017, 206 ff., 213 f.; offen gelassen bei Thowenin/Früh (Fn. 14), 11 f.

⁶⁶ Das Profiling an sich, wie bspw. die automatisierte Bewertung der Kreditwürdigkeit (sog. Kreditscoring) verbietet Art. 22 DSGVO mangels Entscheidung nicht (statt vieler Martini [Fn. 4], 1020; Sebastian Schulz, in: Gola [Hrsg.], Datenschutz-Grundverordnung, Kommentar, 2. Aufl., München 2018, Art. 22 N 4).

⁶⁷ Eine ähnliche Bestimmung für den Bereich der Strafverfolgung und die Strafvollstreckung enthält Art. 11 der Richtlinie (EU) 2016/680 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten durch die zuständigen Behörden zum Zwecke der Verhütung, Ermittlung, Aufdeckung oder Verfolgung von Straftaten oder der Strafvollstreckung sowie zum freien Datenverkehr und zur Aufhebung des Rahmenbeschlusses 2008/977/JI des Rates, ABl 2016 L 119 vom 4. Mai 2016, 89.

⁶⁸ Martini (Fn. 4), 1020; Edwards/Veale (Fn. 30), 74; Mendoza/Bygrave (Fn. 64), 87 f.

b) Informationspflicht und Auskunftsrecht

Sowohl die DSGVO als auch der E-DSG enthalten eine spezifische Informationspflicht und ein Auskunftsrecht bei automatisierten Entscheidungen. Der Verantwortliche hat die betroffene Person über das Bestehen einer automatisierten Entscheidungsfindung zu informieren⁶⁹ und überdies – im Anwendungsbereich der DSGVO – aussagekräftige Informationen über die involvierte Logik sowie die Tragweite und die angestrebten Auswirkungen einer derartigen Verarbeitung zur Verfügung zu stellen⁷⁰. Informationen über die der Entscheidung zugrunde liegende Logik kann unter dem E-DSG die betroffene Person mit dem Auskunftsrecht in Erfahrung bringen (Art. 23 Abs. 2 lit. f E-DSG); unter der DSGVO betrifft das Auskunftsrecht dem Wortlaut nach dieselben Informationen wie die Informationspflicht (Art. 15 Abs. 1 lit. h DSGVO). Inhalt und Tragweite dieser Informationspflichten bzw. dieser Auskunftsrechte sind indessen umstritten⁷¹. Nahe liegt, dass der Verantwortliche die Grundannahmen des Algorithmus offenlegen muss, nicht aber den Algorithmus selbst⁷².

c) Datenschutz-Folgenabschätzung

Der Zweck einer Datenschutz-Folgenabschätzung liegt darin, die Risiken einer Datenbearbeitung im Voraus zu beurteilen, einen informierten Entscheid über deren Umsetzung zu treffen und allenfalls notwendige Abhilfemassnahmen einzuplanen⁷³. Im Allgemeinen ist eine Folgenabschätzung durchzuführen, wenn die Bearbeitung voraussichtlich ein hohes Risiko für die Rechte und Freiheiten natürlicher Personen zur Folge hat⁷⁴. Dies dürfte nicht selten der Fall sein beim Einsatz von selbstlernenden Algorithmen⁷⁵. Ebenso hat der Verantwortliche eine Folgenabschätzung durchzuführen, wenn das algorithmische

⁶⁹ Art. 19 Abs. 1 E-DSG.

⁷⁰ Art. 13 Abs. 2 lit. f und Art. 14 Abs. 2 lit. g DSGVO.

⁷¹ Vgl. dazu unter der DSGVO *Edwards/Veale* (Fn. 30), 44 ff.; *Noto La Diega* (Fn. 16), Rz. 72 ff.; *Andrew D. Selbst/Julia Powles*, *Meaningful information and the right to explanation*, IDPL 2017, 233 ff., *passim*; *Wachter et al.* (Fn. 64), 76 ff.

⁷² Vgl. *Expertengruppe Datenbearbeitung* (Fn. 6), 40, mit zwei unterschiedlichen Interpretationsansätzen; *Botschaft Totalrevision DSG* (Fn. 2), 7067.

⁷³ Vgl. Art. 35 Abs. 7 DSGVO; *Zuiderveen Borgesius* (Fn. 27), 22; *Moritz Karg*, in: *Simitis/Hor-nung/Spiecker gen. Döhmman (Hrsg.)*, *Nomoskommentar, Datenschutzrecht, DSGVO mit BDSG*, Baden-Baden 2019, Art. 35 N 9 ff.

⁷⁴ Art. 35 Abs. 1 DSGVO; Art. 20 Abs. 1 E-DSG.

⁷⁵ Zur DSGVO *Edwards/Veale* (Fn. 30), 78; vgl. auch *Karg* (Fn. 73), Art. 35 N 31.

System umfangreich besonders schützenswerte Personendaten⁷⁶, wie Rassenzugehörigkeit oder politische Ansichten, bearbeitet⁷⁷ oder – im Anwendungsbereich der DSGVO – wenn es systematische und umfassende Bewertungen persönlicher Aspekte natürlicher Personen als Grundlage für automatisierte Entscheidungen vornimmt⁷⁸.

d) Defizite bei algorithmischen Systemen

Aufgrund der Einschränkung auf vollautomatisierte Systeme taugt das Verbot in Art. 22 DSGVO nur sehr bedingt als Instrument gegen die mit Algorithmen verbundenen Risiken; ohnehin ist fraglich, ob sich ein schlichtes Verbot vor dem Hintergrund der Risiken rechtfertigt⁷⁹. Die Informationsansprüche der betroffenen Person verfolgen mit Blick auf die Intransparenz algorithmischer Systeme⁸⁰ ein wichtiges Anliegen⁸¹. Ihre Ausgestaltung in der DSGVO und im E-DSG weist allerdings (ebenfalls) die Schwäche auf, dass sie an vollautomatisierten Entscheidungen ansetzen⁸². Einen weniger eingeschränkten Anwendungsbereich hat die Pflicht des Verantwortlichen, eine Datenschutz-Folgenabschätzung durchzuführen. Nicht zuletzt deshalb könnte sie sich als gutes Rezept zur Prävention der Risiken algorithmischer Systeme erweisen⁸³.

⁷⁶ Art. 9 Abs. 1 DSGVO; Art. 4 lit. c E-DSG.

⁷⁷ Art. 35 Abs. 3 lit. b DSGVO; Art. 20 Abs. 2 lit. a E-DSG. Vgl. dazu im Kontext der DSGVO *Edwards/Veale* (Fn. 30), 78; *Karg* (Fn. 73), Art. 35 N 41, und im Zusammenhang mit dem E-DSG *Botschaft Totalrevision DSG* (Fn. 2), 7060 f.

⁷⁸ Art. 35 Abs. 3 lit. a DSGVO. Vgl. dazu *Zuiderveen Borgesius* (Fn. 27), 22; *Karg* (Fn. 73), Art. 35 N 37 ff.

⁷⁹ Vgl. *Thouvenin/Früh* (Fn. 14), 16 f.; *Thouvenin et al.* (Fn. 13), 3.

⁸⁰ Dazu oben [III.1](#).

⁸¹ *Thouvenin et al.* (Fn. 13), 4; *Martini* (Fn. 4), 1020.

⁸² Zur DSGVO *Martini* (Fn. 4), 1020; *Zuiderveen Borgesius* (Fn. 27), 24. Zum E-DSG vgl. *Braun Binder* (Fn. 8), 475 f. Etwas anderes lässt sich auch aus dem Einschub «zumindest in diesen Fällen» in Art. 13 Abs. 2 lit. f, Art. 14 Abs. 2 lit. g sowie Art. 15 Abs. 1 lit. h DSGVO nicht ableiten (vgl. dazu ausführlich *Martini* [Fn. 7], 182 ff.).

⁸³ Zur DSGVO vgl. *Edwards/Veale* (Fn. 30), 78 f.

2. Schaffung von Transparenz

a) Primär- bzw. verfassungsrechtliche Transparenzerfordernisse

Das Primärrecht der EU⁸⁴ sowie – wenn auch in ungeschriebener Form – das Schweizer Verfassungsrecht⁸⁵ verbürgen die Transparenz als rechtstaatliches Grundprinzip. Während sich dem Primärrecht der EU über diese institutionelle Ebene hinaus ein Individualrecht eines jeden Unionsbürgers entnehmen lässt, unter Vorbehalt begründungsbedürftiger Ausnahmen einen Zugang zu Dokumenten der Union zu erhalten⁸⁶, findet sich in der Schweizer BV kein vergleichbares selbständiges verfassungsmässiges Recht⁸⁷. Selbst wenn diese primär- bzw. verfassungsrechtlichen Vorgaben bzw. Rechte auch hinsichtlich Algorithmen (zwischen Staat und Privaten) für Transparenz sorgen können, liegen Algorithmen nicht in deren regulatorischem Fokus, weshalb vorliegend eine vertiefte Betrachtung unterbleibt.

b) Sekundärrechtliche bzw. einfachgesetzliche Transparenzerfordernisse

Sekundärrechtliche bzw. einfachgesetzliche Transparenzerfordernisse ausserhalb des Datenschutzrechts⁸⁸ betreffen spezifische Algorithmen und finden sich daher verstreut in verschiedenen Spezialerlassen. Ein Beispiel jüngeren Datums betrifft Algorithmen, die Anbieter von Online-Vermittlungsdiensten wie Uber oder Online-Suchmaschinen wie Google für ihre Rankings einsetzen.

⁸⁴ Art. 1 Abs. 2 Vertrag über die Europäische Union, ABl C 326 vom 26. Oktober 2012, 13; Art. 15 Vertrag über die Arbeitsweise der Europäischen Union (AEUV), ABl C 326 vom 26. Oktober 2012, 47; Art. 42 Charta der Grundrechte der Europäischen Union (GRC), ABl C 326 vom 26. Oktober 2012, 391.

⁸⁵ Kerstin Noëlle Vokinger/Urs Jakob Mühlematter, Re-Identifikation von Gerichtsurteilen durch «Linkage» von Daten(banken). Eine empirische Analyse anhand von Bundesgerichtsbeschwerden gegen (Preisfestsetzungs-) Verfügungen von Arzneimitteln, Jusletter, 2. September 2019, Rz. 14; Franziska Sprecher, Transparenz – Ein Grundprinzip des Rechtsstaats und seine Bedeutung im Gesundheitsrecht, insbesondere im Heilmittel- und Humanforschungsrecht, ZSR 2016 II 139 ff., 182, je m.w.H.

⁸⁶ Art. 15 Abs. 3 AEUV und Art. 42 GRC. Vgl. dazu Martini (Fn. 7), 68 f.

⁸⁷ Vokinger/Mühlematter (Fn. 85), Rz. 16; vgl. auch Sprecher (Fn. 85), 187 f.

⁸⁸ Die auf Algorithmen zugeschnittenen Informationsansprüche im Datenschutzrecht wurden bereits thematisiert; vgl. dazu oben [V.1.b\).](#)

Art. 5 der Verordnung (EU) 2019/1150⁸⁹ verlangt von den Anbietern, die (wichtigsten) das Ranking⁹⁰ bestimmenden Hauptparameter, die Gründe für die relative Gewichtung derselben gegenüber anderen Parametern sowie die allfällige Möglichkeit, das Ranking mit einem Entgelt zu beeinflussen, transparent zu machen⁹¹; die Algorithmen selbst müssen die Anbieter hingegen nicht offenlegen⁹². Mit diesen Transparenzerfordernissen will die EU dem erheblichen Einfluss solcher Rankings auf die Wahlmöglichkeit der Verbraucher und auf den geschäftlichen Erfolg der gewerblichen Nutzer bzw. der Nutzer mit Unternehmenswebseiten gerecht werden⁹³.

Ein mit dem EU-Recht vergleichbares Transparenzerfordernis betreffend algorithmengestützte Rankings findet sich im Schweizer Recht nicht. Thematisch zumindest ansatzweise aufgenommen haben diese Problematik hingegen vereinzelte parlamentarische Vorstösse⁹⁴. Im Zusammenhang mit Internet-Vergleichsdiensten lehnt der Bundesrat⁹⁵ aber eine Pflicht zur Offenlegung von Provisionen, die auch für das Ranking relevant sein dürften, ausdrücklich ab und verweist im Übrigen auf die Schutzmechanismen des Bundesgesetzes gegen den unlauteren Wettbewerb⁹⁶.

⁸⁹ Verordnung (EU) 2019/1150 zur Förderung von Fairness und Transparenz für gewerbliche Nutzer von Online-Vermittlungsdiensten, ABl L 186 vom 11. Juli 2019, 57.

⁹⁰ Vgl. zum Begriff «Ranking» Erw.Gr. 24 der Verordnung (EU) 2019/1150.

⁹¹ Gleichartige Transparenzerfordernisse zugunsten der Verbraucher wurden per 7. Januar 2020 in die Richtlinie 2005/29/EG über unlautere Geschäftspraktiken, ABl L 149 vom 11. Juni 2005, 22, sowie in die Richtlinie 2011/83/EU über die Rechte der Verbraucher, zur Abänderung der Richtlinie 93/13/EWG des Rates und der Richtlinie 1999/44/EG des Europäischen Parlaments und des Rates sowie zur Aufhebung der Richtlinie 85/577/EWG des Rates und der Richtlinie 97/7/EG des Europäischen Parlaments und des Rates, ABl L 304 vom 22. November 2011, 64, aufgenommen (Richtlinie [EU] 2019/2161 zur Änderung der Richtlinie 93/13/EWG des Rates und der Richtlinien 98/6/EG, 2005/29/EG und 2011/83/EU des Europäischen Parlaments und des Rates zur besseren Durchsetzung und Modernisierung der Verbraucherschutzvorschriften der Union, ABl L 328 vom 18. Dezember 2019, 7).

⁹² Erw.Gr. 27 der Verordnung (EU) 2019/1150.

⁹³ Erw.Gr. 24 und 26 der Verordnung (EU) 2019/1150.

⁹⁴ Postulat Béglé, 16.3914 (algorithmengestützte Rankings, insb. im Zusammenhang mit Suchmaschinen und sozialen Netzwerken); Postulat Reimann, 18.4346 (Rankings von Internet-Vergleichsdiensten).

⁹⁵ Stellungnahme des BR zum Postulat Reimann, 18.4346.

⁹⁶ Bundesgesetz gegen den unlauteren Wettbewerb vom 19. Dezember 1986, SR 241.

Ein weiteres Beispiel eines spezifischen Transparenzerfordernisses findet sich in der Verordnung (EU) 2017/746⁹⁷, die sich mit *In-vitro*-Diagnostika⁹⁸ befasst. Vorgesehen ist, dass der Hersteller (zuhanden der zuständigen Behörden) eine technische Dokumentation über sein Diagnostikum zu verfassen hat⁹⁹. Als interessant erscheint, dass darin, wenn es sich beim Diagnostikum um eine Software handelt, «eine Beschreibung der Methodik zur Datenauswertung, d.h.» – im Vergleich zu anderen Transparenzerfordernissen unüblich – «der Algorithmus» selbst aufzunehmen ist¹⁰⁰.

Überdies lassen sich auf Algorithmen bezogene Transparenzerfordernisse insbesondere in meist sehr technischen Konkretisierungs- und Ergänzungserlassen wie Durchführungs- oder delegierten Verordnungen der Kommission¹⁰¹ bzw. in Bundesverordnungen¹⁰² finden.

c) Transparenzerfordernisse im Soft Law

Allgemeiner gefasst sind die im *Soft Law* postulierten Transparenzerfordernisse. Die hochrangige Expertengruppe der EU bezeichnet Transparenz als eine der sieben Grundvoraussetzungen für eine vertrauenswürdige KI. Als Element davon sollen gemäss ihren Richtlinien die Datensätze und die Entscheidungsprozesse von KI-Systemen samt der Datenerfassung und -kennzeichnung sowie den eingesetzten Algorithmen dokumentiert werden, um die Rückverfolgbarkeit sicherzustellen¹⁰³. Personen, auf deren Leben ein KI-System einen erheblichen Einfluss hat, sollen eine Erklärung über den Entscheidungsprozess verlangen können. Ausserdem sollen sich KI-Systeme nicht als Menschen

⁹⁷ Verordnung (EU) 2017/746 über *In-vitro*-Diagnostika und zur Aufhebung der Richtlinie 98/79/EG und des Beschlusses 2010/227/EU der Kommission, ABl L 117 vom 5. Mai 2017, 176.

⁹⁸ *In-vitro*-Diagnostika sind Medizinprodukte inkl. Software, die zur Untersuchung von aus dem menschlichen Körper stammenden Proben bestimmt sind und spezifische Gesundheitsinformationen liefern (vgl. Art. 2 Nr. 2 der Verordnung [EU] 2017/746).

⁹⁹ Art. 10 Abs. 4 und Abs. 7 Verordnung (EU) 2017/746.

¹⁰⁰ Anhang II, Ziff. 3.1.d Verordnung (EU) 2017/746.

¹⁰¹ Vgl. etwa Art. 22 Abs. 3 lit. h der Verordnung (EU) 2017/2195 der Kommission zur Festlegung einer Leitlinie über den Systemausgleich im Elektrizitätsversorgungssystem, ABl L 312 vom 28. November 2017, 6.

¹⁰² Vgl. etwa Art. 24 der Verordnung des EDI über den Strahlenschutz bei medizinischen Teilchenbeschleunigeranlagen vom 26. April 2017, SR 814.501.513.

¹⁰³ *Hochrangige Expertengruppe* (Fn. 17), Rz. 76, vgl. auch die Konkretisierungen zu regelbasierten einerseits und lernbasierten KI-Systemen andererseits auf S. 37.

ausgeben dürfen; Menschen sollen das Recht haben, im Falle einer Interaktion mit einem KI-System darüber informiert zu werden¹⁰⁴.

Auch die OECD empfiehlt, die KI-Akteure zu Transparenz und verantwortungsvoller Offenlegung zu verpflichten. Sie sollen sachdienliche, dem Kontext angemessene und dem neuesten Stand der Technik entsprechende Informationen bereitstellen, um u.a. die von einem KI-System Betroffenen zu befähigen, das Ergebnis zu verstehen und – sofern ihnen daraus Nachteile entstehen – das jeweilige Ergebnis auf der Grundlage von klaren und leicht verständlichen Informationen über die einzelnen Faktoren und die zugrunde liegende Logik zu hinterfragen¹⁰⁵.

Ähnlich verpflichtet sich Google in der Form einer Selbstregulierung darauf, KI-Systeme zu entwickeln, die einen angemessenen Feedbackmechanismus, relevante Erklärungen und eine allfällige Anfechtung gewährleisten¹⁰⁶. Microsoft hält ebenfalls in seinen KI-Prinzipien fest, dass KI-Systeme nachvollziehbar sein sollen¹⁰⁷.

d) Defizite bei algorithmischen Systemen

Selbst wenn Transparenzerfordernisse einen wichtigen Beitrag leisten können, betroffene Personen vor den mit Algorithmen verbundenen Risiken zu schützen, darf deren Wirkung nicht überschätzt werden. Empirische Untersuchungen zu daten- und verbraucherschutzrechtlichen Informationspflichten lassen einen jedenfalls zweifeln, ob Transparenzerfordernisse stets einen Mehrwert für die betroffene Person generieren und ihren angedachten Zweck tatsächlich erreichen¹⁰⁸. Grund dafür ist, dass die betroffenen Personen die erhaltenen Informationen regelmässig nicht verstehen oder gar nicht erst lesen

¹⁰⁴ Hochrangige Expertengruppe (Fn. 17), Rz. 77 f. und S. 37 f.

¹⁰⁵ Rat der OECD (Fn. 40), Ziff. 1.3.

¹⁰⁶ Google (Fn. 43), Ziff. 4.

¹⁰⁷ Microsoft (Fn. 44), Prinzip «Transparency».

¹⁰⁸ Vgl. zum Datenschutzrecht Sara Elisa Kettner/Christian Thorun/Max Vetter, Wege zur besseren Informiertheit, Verhaltenswissenschaftliche Ergebnisse zur Wirksamkeit des One-Pager-Ansatzes und weiterer Lösungsansätze im Datenschutz, Berlin 2018, 31 ff. Vgl. zum Verbraucherschutzrecht Oren Bar-Gill/Omri Ben-Shahar, Regulatory Techniques in Consumer Protection: A Critique of European Consumer Contract Law, CML Rev 2013, 109 ff., 117 ff. Vgl. zum Ganzen auch Rolf H. Weber, From Disclosure to Transparency in Consumer Law, in: Mathis/Tor (Hrsg.), Consumer Law and Economics, Cham 2020, 73, 77 f.

bzw. zur Kenntnis nehmen¹⁰⁹. Dies hängt nicht zuletzt damit zusammen, dass oft ein Überfluss an Informationen besteht, der die betroffene Person überfordert (sog. *Information Overload*)¹¹⁰. Im Zusammenhang mit Algorithmen dürfte sich diese Problematik ähnlich zeigen und sich aufgrund der hohen Komplexität insbesondere selbstlernender Systeme sogar verstärkt äussern. Ein zur Transparenz verpflichteter Anwender eines algorithmischen Systems hat deshalb darauf zu achten, dass er – wie im *Soft Law* zuweilen vorgesehen – sachdienlich und den Verhältnissen angemessen informiert. Eine eigentliche Erklärung von Entscheidungen selbstlernender algorithmischer Systeme ist indessen bisweilen technisch nur schwer zu bewerkstelligen und kann bald so kompliziert werden, dass sie keinen Gewinn mehr für die betroffene Person darstellt¹¹¹.

3. Schutz vor Diskriminierungen

a) Primär- bzw. verfassungsrechtlicher Diskriminierungsschutz

Auf der Stufe des Primär- bzw. Verfassungsrechts finden sich in Art. 21 GRC und in Art. 8 Abs. 2 BV unmittelbar anwendbare Diskriminierungsverbote. Wie die Wortlaute der Bestimmungen zum Ausdruck bringen, beschränkt sich ihr Schutz nicht auf abschliessend aufgezählte Merkmale, aufgrund derer eine Diskriminierung verboten ist¹¹². Ausserdem erfassen sie sowohl die unmittelbare als auch die mittelbare Diskriminierung¹¹³, die vorliegt, wenn dem Anschein nach neutrale Vorschriften Personen der geschützten Gruppe gegenüber anderen Personen benachteiligen. Dafür können die Diskriminierungsverbote nur gegenüber dem Staat als Anwender algorithmischer Systeme geltend gemacht werden: Adressaten der Grundrechtecharta sind nämlich nur die Institutionen der EU sowie – bei Durchführung des EU-Rechts – dieje-

¹⁰⁹ Vgl. Omri Ben-Shahar/Carl E. Schneider, The Failure of Mandated Disclosure, U Pa L Rev 2011, 647 ff., 709 ff.; Martini (Fn. 7), 188 f.; Eva Maria Tscherner, Can behavioral research advance mandatory law, information duties, standard terms and withdrawal rights?, ALJ 2014, 144 ff., 148 f.; Weber (Fn. 108), 79 f.

¹¹⁰ Vgl. dazu Ben-Shahar/Schneider (Fn. 109), 686 ff.; Tscherner (Fn. 109), 148 f.; Weber (Fn. 108), 79 f.

¹¹¹ Martini (Fn. 7), 194; vgl. auch Edwards/Veale (Fn. 30), 59 ff.

¹¹² Art. 21 Abs. 1 GRC («insbesondere»); Art. 8 Abs. 2 BV («namentlich»).

¹¹³ Zu Art. 21 GRC vgl. Sven Hölscheidt, in: Meyer/Hölscheidt (Hrsg.), Nomoskommentar, Charta der Grundrechte der Europäischen Union, 5. Aufl., Baden-Baden 2019, Art. 21 GRC N 36 m.w.H. Zu Art. 8 Abs. 2 BV vgl. BGE 132 I 49 E. 8.1 S. 66; BGE 129 I 217 E. 2.1 S. 223.

nigen der Mitgliedstaaten (Art. 51 GRC). Ebenso wenig entfaltet das Diskriminierungsverbot der Bundesverfassung direkte Drittwirkung unter Privaten¹¹⁴.

b) Sekundärrechtlicher bzw. einfachgesetzlicher Diskriminierungsschutz

Diskriminierungsverbote, die sowohl gegenüber dem Staat als auch Privaten geltend gemacht werden können, sind im Sekundärrecht der EU¹¹⁵ bzw. in der Schweiz einfachgesetzlich vorgesehen¹¹⁶. Die Richtlinie 2000/43/EG¹¹⁷ verbietet Diskriminierungen aufgrund der Rasse oder ethnischen Herkunft im Bereich der Erwerbstätigkeit, des Sozialschutzes und Wohlfahrtsystems, der Bildung sowie des Zugangs zu und Versorgung mit Gütern und Dienstleistungen. Diskriminierungen aufgrund des Geschlechts verbieten die Richtlinie 2006/54/EG¹¹⁸ in Arbeits- und Beschäftigungsfragen sowie die Richtlinie 2004/113/EG¹¹⁹ beim Zugang zu und bei der Versorgung mit Gütern und Dienstleistungen. Ausschliesslich im Bereich der Beschäftigung und des Berufs schützt die Richtlinie 2000/78/EG¹²⁰ vor Diskriminierungen wegen der Religion oder der Weltanschauung, einer Behinderung, des Alters oder der sexuellen Ausrichtung.

Das Schweizer Recht kennt zwei Gesetze, die dem Diskriminierungsverbot auch gegenüber Privaten Geltung verleihen: das Behindertengleichstellungs-

¹¹⁴ BGE 137 III 59 E. 4.1 S. 61; Giovanni Biaggini, Kommentar, Bundesverfassung der Schweizerischen Eidgenossenschaft, 2. Aufl., Zürich 2017, Art. 8 N 18; Rainer J. Schweizer, in: Ehrenzeller et al. (Hrsg.), Die schweizerische Bundesverfassung, St. Galler Kommentar, Zürich 2014, Art. 8 N 56.

¹¹⁵ Sandra Wachter, Affinity Profiling and Discrimination by Association in Online Behavioural Advertising, BTLJ 2020, 367, 388 ff.; Castelluccia/Le Métayer (Fn. 10), 56.

¹¹⁶ Vgl. Biaggini (Fn. 114), Art. 8 N 18; Schweizer (Fn. 114), Art. 8 N 56.

¹¹⁷ Richtlinie 2000/43/EG zur Anwendung des Gleichbehandlungsgrundsatzes ohne Unterschied der Rasse oder der ethnischen Herkunft, ABl L 180 vom 19. Juli 2000, 22.

¹¹⁸ Richtlinie 2006/54/EG zur Verwirklichung des Grundsatzes der Chancengleichheit und Gleichbehandlung von Männern und Frauen in Arbeits- und Beschäftigungsfragen, ABl L 204 vom 26. Juli 2006, 23.

¹¹⁹ Richtlinie 2004/113/EG zur Verwirklichung des Grundsatzes der Gleichbehandlung von Männern und Frauen beim Zugang zu und bei der Versorgung mit Gütern und Dienstleistungen, ABl L 373 vom 21. Dezember 2004, 37.

¹²⁰ Richtlinie 2000/78/EG zur Festlegung eines allgemeinen Rahmens für die Verwirklichung der Gleichbehandlung in Beschäftigung und Beruf, ABl L 303 vom 2. Dezember 2000, 16.

gesetz¹²¹, das Behinderte u.a. vor Diskriminierungen durch Private schützt, die Dienstleistungen öffentlich anbieten, sowie das Gleichstellungsgesetz¹²², das die direkte oder indirekte Benachteiligung von Arbeitnehmenden aufgrund ihres Geschlechts verbietet.

Alle aufgeführten EU-Richtlinien untersagen sowohl die unmittelbare als auch – bei fehlender Rechtfertigung – die mittelbare Diskriminierung¹²³. Dasselbe gilt für das Schweizer Gleichstellungsgesetz (vgl. Art. 3 Abs. 1 GlG). Demgegenüber ist umstritten, ob Art. 6 BehiG auch die mittelbare Diskriminierung verbietet¹²⁴. Der sekundärrechtliche bzw. einfachgesetzliche Diskriminierungsschutz ist somit durch die angesprochenen Lebensbereiche sowie durch das Abstellen auf bestimmte geschützte Merkmale beschränkt¹²⁵.

c) Persönlichkeitsrechtlicher Diskriminierungsschutz

Weder die EU noch die Schweiz kennt ein über die genannten Diskriminierungsverbote hinausgehendes allgemeines Anti-Diskriminierungsgesetz, das auch im privatrechtlichen Verhältnis, bspw. zwischen Unternehmen und Verbrauchern, unzulässige Diskriminierungen verbieten würde. Die Schweizer Lehre leitet indessen vereinzelt aus dem allgemeinen Persönlichkeitsrecht,

¹²¹ Art. 6 Bundesgesetz über die Beseitigung von Benachteiligungen von Menschen mit Behinderungen vom 13. Dezember 2002, SR 151.3. Eine Diskriminierung liegt vor, wenn Behinderte mit dem Ziel oder der Folge, sie herabzuwürdigen oder auszugrenzen, besonders krass unterschiedlich und benachteiligend behandelt werden (Art. 2 lit. d Verordnung über die Beseitigung von Benachteiligungen von Menschen mit Behinderungen vom 19. November 2003, SR 151.31). Dies stellt eine hohe Hürde dar, zumal der Gesetzgeber damit bloss «besonders stossendes Verhalten» von Privaten verhindern wollte (Botschaft des Bundesrates vom 11. Dezember 2000 zur Volksinitiative «Gleiche Rechte für Behinderte» und zum Entwurf eines Bundesgesetzes über die Beseitigung von Benachteiligungen behinderter Menschen, BBl 2001, 1715 ff., 1780).

¹²² Art. 3 Bundesgesetz über die Gleichstellung von Frau und Mann vom 24. März 1995, SR 151.1. Anders als das BehiG (vgl. Fn. 121) setzt das GlG nicht voraus, dass die ungleiche Behandlung besonders krass wiegt und es deren Ziel oder Folge ist, die betroffene Person herabzuwürdigen oder auszugrenzen.

¹²³ Auch direkte bzw. indirekte Diskriminierung genannt: Art. 2 Abs. 1 lit. b der Richtlinie 2000/43/EG; Art. 2 Abs. 1 lit. b der Richtlinie 2006/54/EG; Art. 2 lit. b der Richtlinie 2004/113/EG; Art. 2 Abs. 2 lit. b der Richtlinie 2000/78/EG.

¹²⁴ Gemäss *Manuel Jaun*, Diskriminierungsschutz durch Privatrecht, Die Schweiz im Schatten der europäischen Rechtsentwicklung, ZBJV 2007, 457 ff., 464, wird nur die unmittelbare, gemäss *Kurt Pärli*, Urteil des Bundesgerichts 5P.97/2006 vom 1. Juli 2006, HAVE 2007, 46 ff., 51, auch die mittelbare Diskriminierung erfasst.

¹²⁵ *Wachter* (Fn. 115), 388 ff.; vgl. auch *Zuiderveen Borgesius* (Fn. 27), 19 f.

insbesondere Art. 28 ZGB¹²⁶, einen solchen Diskriminierungsschutz ab¹²⁷. Dieser dürfte zwar in Anlehnung an Art. 8 Abs. 2 BV auf die dort geschützten Merkmale beschränkt sein¹²⁸, dafür aber auch mittelbare Diskriminierungen, die etwa beim Abstellen algorithmischer Systeme auf Proxies auftreten, als Persönlichkeitsverletzung erfassen¹²⁹. Anders als im EU-Recht wird mit dieser Lehrmeinung der Diskriminierungsschutz über bestimmte Lebensbereiche hinaus auf das gesamte Privatrecht erweitert.

d) Diskriminierungsschutz im Soft Law

Auch das *Soft Law* verpönt Diskriminierungen und enthält entsprechend Vorgaben, die sich dagegen richten. Die von der EU-Kommission eingesetzte Hochrangige Expertengruppe plädiert zur Verhinderung von Diskriminierungen in KI-Systemen insbesondere für technische Massnahmen wie das Bereinigen von Verzerrungen in Datensätzen oder das Einführen von Aufsichtsverfahren, mit deren Hilfe der Zweck, die Einschränkungen, Anforderungen und Entscheidungen des Systems klar und transparent analysiert und angegangen werden können. Zudem sei eine barrierefreie Zugänglichkeit der Technologie sicherzustellen und ein möglichst breiter Nutzerkreis anzusprechen¹³⁰.

Etwas allgemeiner drücken sich die Empfehlungen der OECD aus, wonach die KI-Akteure, d.h. Akteure, die im Lebenszyklus des KI-Systems eine aktive Rolle spielen, die Rechtstaatlichkeit, die Menschenrechte und die demokratischen Rechte achten sollen. Dazu zählt auch die Nichtdiskriminierung und die Gleichbehandlung, zu deren Gunsten die OECD als Schutzmassnahme die Möglichkeit menschlicher Entscheidungen beliebt macht¹³¹. Google hält in sei-

¹²⁶ Schweizerisches Zivilgesetzbuch vom 10. Dezember 1907, SR 210.

¹²⁷ Tarkan Göksu, Rassendiskriminierung beim Vertragsabschluss als Persönlichkeitsverletzung, Freiburg i.Ü. 2003, Rz. 214 ff.; Florent Thouvenin, Privatversicherungen: Datenschutzrecht als Grenze der Individualisierung?, in: Epiney/Sangsue (Hrsg.), Datenschutz und Gesundheitsrecht/Protection des données et droit de la santé, Zürich 2019, 15 ff., 26, je m.w.H.; vgl. auch Ruth Arnet, Freiheit und Zwang beim Vertragsabschluss, Zürich 2008, Rz. 356. Im Arbeitsrecht wird aus Art. 328 Bundesgesetz betreffend die Ergänzung des Schweizerischen Zivilgesetzbuches vom 30. März 1911, SR 220, ein allgemeines Diskriminierungsverbot, das sich ebenfalls auf den Persönlichkeitsschutz stützt, abgeleitet (Pais/Amann [Fn. 55], 1095).

¹²⁸ Thouvenin (Fn. 127), 27; vgl. auch Göksu (Fn. 127), Rz. 215 i.V.m. Rz. 12.

¹²⁹ Vgl. Göksu (Fn. 127), Rz. 215 i.V.m. Rz. 13.

¹³⁰ Hochrangige Expertengruppe (Fn. 17), Rz. 80 ff., vgl. auch die Konkretisierungen zur praktischen Umsetzung auf S. 38 f.

¹³¹ Rat der OECD (Fn. 40), Ziff. 1.2.

nen Prinzipien ebenfalls fest, das Unternehmen strebe an, ungerechte Auswirkungen ihrer KI-Systeme zu vermeiden, insbesondere solche, die an sensiblen Merkmalen anknüpfen¹³².

e) Defizite bei algorithmischen Systemen

Zwar werden durch die Figur der mittelbaren Diskriminierung grundsätzlich auch Fälle erfasst, in denen ein algorithmisches System trotz eines neutralen Kriteriums einzelne Personen mit einem geschützten Merkmal schlechter stellt¹³³. In der Praxis kann es den Betroffenen indessen schwer fallen, den Beweis für eine mittelbare Diskriminierung zu erbringen¹³⁴: Versagt bspw. eine Bank gestützt auf einen selbstlernenden Algorithmus, der Frauen ungewollt und für den Anwender unerkannt diskriminiert, einer Antragstellerin den Kredit, weiss sie nicht ohne Weiteres, dass ein unverhältnismässig hoher Anteil von Frauen kein Darlehen erhält und kann dies auch nur schwer in Erfahrung bringen.

Überdies können selbstlernende Algorithmen durch ihre Mustererkennung neue (bisher ungeschützte) Diskriminierungsmerkmale schaffen, die ebenso problematisch sind, nicht aber im Anwendungsbereich der geltenden Verbote liegen¹³⁵. Erfolgsversprechender als Diskriminierungsverbote wären daher im Zusammenhang mit algorithmischen Systemen die im *Soft Law* postulierten technischen Massnahmen wie z.B. das Bereinigen von Verzerrungen in Datensätzen oder das Einführen von Aufsichtsverfahren.

4. Exkurs: Finanzmarktrechtliche Bestimmungen

Weitere Normierungen algorithmischer Systeme finden sich insbesondere im Bereich des algorithmischen Börsenhandels und des Hochfrequenzhandels, der sich durch eine hohe Anzahl von Auftragseingaben, -änderungen oder -lösungen innerhalb von Mikrosekunden auszeichnet¹³⁶. In der EU reguliert

¹³² Google (Fn. 43), Ziff. 2.

¹³³ Datenethikkommission (Fn. 13), 194; Zuiderveen Borgesius (Fn. 27), 20.

¹³⁴ Zuiderveen Borgesius (Fn. 27), 19 f.; vgl. auch Criado/Such (Fn. 34), 89; Tischbirek (Fn. 34), Rz. 40.

¹³⁵ Datenethikkommission (Fn. 13), 194; Zuiderveen Borgesius (Fn. 27), 20; Criado/Such (Fn. 34), 89.

¹³⁶ Dazu Interpellation Chevalley, 16.3029.

die Richtlinie 2014/65/EU¹³⁷ diese Handelsmethoden, in der Schweiz das FinfraG¹³⁸ und die FinfraV¹³⁹. Beide Rechtsordnungen auferlegen den Händlern, die sich solcher algorithmischer Systeme bedienen, verschiedene Pflichten, darunter die Kennzeichnungspflicht der durch einen Algorithmus erzeugten Aufträge¹⁴⁰.

Weitere finanzmarktrechtliche Regulierungen von Algorithmen in der EU finden sich in den Verordnungen (EU) 575/2013¹⁴¹ und (EU) 600/2014¹⁴². Art. 174 der Verordnung (EU) 575/2013 stellt etwa verschiedene Anforderungen, wenn ein Kreditinstitut oder eine Wertpapierfirma für die Zuordnung von Risikopositionen zu Schuldner- bzw. Fazilitäts-Ratingstufen oder Risikopools statistische Modelle oder andere algorithmische Verfahren verwendet. Art. 26 Abs. 1–3 der Verordnung (EU) 600/2014 verpflichtet Wertpapierfirmen, die Geschäfte mit Finanzinstrumenten tätigen, zur Meldung bestimmter Geschäfte, wobei die Meldung u.a. den für den Anlageentscheid verantwortlichen Computeralgorithmus umfassen muss.

VI. Fazit und Ausblick

Mit dem Einsatz algorithmischer Systeme – insbesondere der selbstlernenden Art – sind Risiken wie Intransparenz, Unfairness und Diskriminierung verbunden. Dieser nehmen sich zwar schon heute Regulierungen unterschiedli-

¹³⁷ Richtlinie 2014/65/EU über Märkte für Finanzinstrumente sowie zur Änderung der Richtlinien 2002/92/EG und 2011/61/EU, ABl L 173 vom 12. Juni 2014, 349. Diese wird durch die Delegierte Verordnung (EU) 2017/589 zur Ergänzung der Richtlinie 2014/65/EU des Europäischen Parlaments und des Rates durch technische Regulierungsstandards zur Festlegung der organisatorischen Anforderungen an Wertpapierfirmen, die algorithmischen Handel betreiben, ABl L 87 vom 31. März 2017, 417, konkretisiert.

¹³⁸ Bundesgesetz über die Finanzmarktinfrastrukturen und das Marktverhalten im Effekten- und Derivatehandel (Finanzmarktinfrastukturgesetz, FinfraG) vom 19. Juni 2015, SR 958.1.

¹³⁹ Verordnung über die Finanzmarktinfrastrukturen und das Marktverhalten im Effekten- und Derivatehandel (Finanzmarktinfrastukturverordnung, FinfraV) vom 25. November 2015, SR 958.11.

¹⁴⁰ Art. 30 FinfraG; Art. 31 FinfraV; Art. 48 Richtlinie 2014/65/EU.

¹⁴¹ Verordnung (EU) 575/2013 über Aufsichtsanforderungen an Kreditinstitute und Wertpapierfirmen und zur Änderung der Verordnung (EU) 646/2012, ABl L 176 vom 27. Juni 2013, 1.

¹⁴² Verordnung (EU) 600/2014 über Märkte für Finanzinstrumente und zur Änderung der Verordnung (EU) 648/2012, ABl L 173 vom 12. Juni 2014, 84. Diese wird durch die Delegierte Verordnung (EU) 2017/590 zur Ergänzung der Verordnung (EU) Nr. 600/2014 des Europäischen Parlaments und des Rates durch technische Regulierungsstandards für die Meldung von Geschäften an die zuständigen Behörden, ABl L 87 vom 31. Juli 2017, 449, konkretisiert.

cher Ansätze und verschiedener Urheber, allerdings nicht immer mit einer erfolgsversprechenden Herangehensweise, an. So knüpft etwa der Datenschutz wenig überzeugend an der Vollautomatisierung an und erfasst damit einen Grossteil der in der Praxis eingesetzten Entscheidungsalgorithmen nicht.

Ebenso wenig einen effektiven Schutz zu gewährleisten vermögen bisweilen die Diskriminierungsverbote, weil sie nicht auf die Eigenheiten algorithmischer Systeme ausgerichtet sind. Mehr hingegen verspricht der Ansatz, wonach, wie im *Soft Law* vereinzelt vorgesehen, KI-Akteure technische Massnahmen ergreifen (müssen), um bspw. Diskriminierungen durch die Bereinigung von Verzerrungen zu verhindern und die Voraussetzungen für das algorithmische System zu schaffen, damit angemessene Erklärungen abgegeben werden können.

Als Gefäss für solche und weitere Regeln würde sich in der EU, wie die deutsche Datenethikkommission empfiehlt, eine Verordnung mit einem horizontal angelegten Anwendungsbereich anbieten. Diese könnte zentrale Grundprinzipien algorithmischer Systeme, Regelungen zur Zulässigkeit und Gestaltung algorithmischer Systeme, Anforderungen an die Transparenz sowie organisatorische und technische Absicherungen vorsehen. Nicht zuletzt mit Blick auf die Vielschichtigkeit der heutigen Regulierung algorithmischer Systeme spräche einiges für eine solche Verordnung. Dass allerdings, wie vorgeschlagen, eine Gleichbehandlung von öffentlichen Behörden und Privaten als Anwender von algorithmischen Systemen angezeigt ist, bleibt zu bezweifeln¹⁴³. Vorerst ist abzuwarten, ob sich die EU tatsächlich des Erlasses einer solchen Verordnung annimmt und ob die Schweiz gegebenenfalls nachzieht.

¹⁴³ Vgl. zum Ganzen *Datenethikkommission* (Fn. 13), 180 f., die eine EU-Verordnung für Algorithmische Systeme (EUVAS) beliebt macht.

Künstliche Intelligenz und Datenschutz

Erstveröffentlichung in: Weblaw, 4. Juli 2024

In den letzten Jahren haben verfeinerte Gesetze in den meisten Ländern die persönlichkeitsbezogenen Grundprinzipien des Datenschutzes zu stärken versucht. Die neuesten technologischen Entwicklungen, v.a. die «Künstliche Intelligenz», führen nun aber zu weiteren (zusätzlichen) Herausforderungen. Das vorhandene Regelungsumfeld, sachgerecht ausgelegt und zur Anwendung gebracht, vermag indessen die meisten durch KI-Systeme hervorgerufenen Datenschutzprobleme zu bewältigen.

Inhaltsübersicht

1.	Einleitung	130
2.	DSG-Anknüpfungspunkte	132
2.1.	Schutz von Personendaten	132
2.2.	Datensicherheit	133
3.	Spannungsfelder zwischen künstlicher Intelligenz und Datenschutz	134
3.1.	Datenbearbeitungsgrundsätze	134
3.2.	Rechtfertigungsgründe	135
3.2.1.	Datenbearbeitung zu nicht-personenbezogenen Zwecken	136
3.2.2.	Weitere Rechtfertigungsgründe	137
3.3.	Automatisierte Einzelentscheidungen	139
3.4.	Verwendung von Personendaten	140
3.5.	Informations- und Sorgfaltspflichten	141
3.6.	Verantwortlichkeit	142
4.	Sonderfragen	143
4.1.	Trainings-KI	143
4.2.	Transparenz	143
4.3.	Betroffenenrechte	144
4.3.1.	Auskunftsrecht	145
4.3.2.	Ansprüche auf Widerspruch und Korrektur	145
5.	Ausblick	146

1. Einleitung

[1] Der weite Begriff «Künstliche Intelligenz» (KI), der z.B. Erscheinungen wie automatisierte Entscheidungsabläufe, algorithmische Programmierungen und auch juristische Sprachmodelle bzw. allgemein Large Language Models (LLM) umfasst, hat in den letzten Monaten erheblich an Bedeutung gewonnen. Nachfolgend wird ein umfassendes Begriffsverständnis der KI in Betracht gezogen, welches nicht nur die generative KI, sondern auch das Machine Learning und das Deep Learning mitumfasst. Die KI vermag grosse Mengen an Daten zu verarbeiten und daraus Muster und Zusammenhänge zu eruieren, die für menschliche Analysen nicht leicht erkennbar sind. Solche Fähigkeiten hat die KI zu einem wertvollen Werkzeug in einer Vielzahl von Bereichen gemacht, aber auch neue und komplexe Fragen in Bezug auf den Datenschutz aufgeworfen.

[2] Entgegen verbreiteter Meinungsäusserungen ist die KI aber nicht nur eine Herausforderung für die Einhaltung der Datenschutzvorgaben, sondern algorithmische Programme können durchaus zur Verbesserung der Datensicherheit beitragen.¹ Beispiele sind (i) die Unterstützung bei der Detektion unerwünschter Inhalte, (ii) die Unterstützung bei der Berichterstattung zu Sicherheitsvorfällen, (iii) die Unterstützung bei der Erstellung und Analyse von Programmcodes (Untersuchungen mit Blick auf Sicherheitslücken) sowie (iv) die Unterstützung bei der Analyse des Datenverkehrs.² Datenschutz und KI stehen indessen in einem komplexen Wechselspiel zueinander; datenschutzrechtliche Vorgaben zur Anonymisierung von Daten oder der Grundsatz der Datensparsamkeit können etwa dazu beitragen, dass bestimmte KI-Techniken (z.B. Deep Learning), die eine Bearbeitung grosser Mengen persönlicher Daten bezwecken, nur sehr eingeschränkt Verwendung finden dürfen.³

[3] Das Datenschutzrecht spielt in mehreren Phasen des KI-Einsatzes eine Rolle, nämlich (i) bei der Datenbeschaffung für den Aufbau eines KI-Systems, (ii) bei dessen Training, (iii) bei dessen Besitz und Betrieb sowie (iv) bei der

¹ ROLF H. WEBER, Juristische Sprachmodelle zwischen Transparenz und Datensicherheit, in: Erich Schweighofer et al. (Hrsg.), Sprachmodelle: Juristische Papageien oder mehr?, Bern 2024, S. 19, S. 22 (vgl. [Text 9, S. 149 ff.](#)).

² Eine konzise und übersichtliche Darstellung möglicher technischer Datenschutzmassnahmen findet sich bei DAVID ROSENTHAL/VEREIN UNTERNEHMENS-DATENSCHUTZ, Datenschutz-Folgeabschätzung, Dezember 2023, Ziff. 3.12-3.22, https://www.rosenthal.ch/downloads/vud_dsfa.xlsm.

³ WEBER (Fn. 1), S. 23.

praktischen Verwendung durch den Nutzer (Eingabe eines «Prompt»);⁴ Die «Bedrohung» des Datenschutzes durch den Einsatz von KI ist in den letzten Monaten bereits Gegenstand verschiedener wissenschaftlicher Untersuchungen und praktischer Handlungsempfehlungen gewesen;⁵ dass die Datenschutzregeln auf KI-Systeme grundsätzlich, wie der EDÖB zutreffend bemerkt,⁶ anwendbar sind, dürfte unbestritten sein. Angesichts der vorhandenen Literatur rechtfertigt es sich, den vorliegenden Beitrag auf spezifische Themen zu konzentrieren und nicht alle Problembereiche anzusprechen;⁷ die nachfolgenden Ausführungen beziehen sich demgemäss auf im Vordergrund stehende Herausforderungen, nämlich die DSGVO-Anknüpfungspunkte, die spezifischen Spannungsfelder zwischen KI und Datenschutz sowie einzelne Sonderfragen.

[4] Der Datenverkehr läuft angesichts des nicht-physischen Charakters von Daten sehr oft grenzüberschreitend ab. Aus diesem Grunde lehnt sich das Schweizer Datenschutzgesetz (DSG) über weite Strecken an die Datenschutz-Grundverordnung der EU (DSGVO) von 2016 an; die Äquivalenz der Normenwerke ist denn auch anerkannt.⁸ Ungeachtet dieser Ausgangslage wird nachfolgend aber lediglich das DSG erörtert, und zwar deshalb, weil sich vom Regelungsansatz her betrachtet das DSG und die DSGVO grundlegend unterscheiden, denn das DSG beruht auf dem Prinzip der Zulässigkeit der Datenbearbeitung mit Verbotsvorbehalt, die DSGVO hingegen auf dem Prinzip des Verbots mit Erlaubnisvorbehalt.⁹ Zwar löst sich dieser prinzipielle Unterschied

⁴ DAVID ROSENTHAL, Datenschutz beim Einsatz generativer künstlicher Intelligenz, in: Jusletter 6. November 2023, Rz. 2 und Rz. 7.

⁵ Aus der neueren Literatur vgl. FLORENT THOUVENIN/STEPHANIE VOLZ, «Datenschutz», White Paper, Juni 2024, <https://www.itsl.uzh.ch/de/Wissenstransfer/Publicationen.html#Positionspapiere>; ROSENTHAL (Fn. 4); DANIEL W. SEILER/MARCEL GRIESINGER, Spannungsfeld Künstliche Intelligenz (KI) und Datenschutzrecht, in: Jusletter 6. September 2023.

⁶ EDÖB, Geltendes Datenschutzgesetz ist auf KI direkt anwendbar, Version 11. Dezember 2023, https://www.edoeb.admin.ch/content/edoeb/de/home/kurzmeldungen/2023/20231109_ki_dsg.html.

⁷ Umfassende Überblicke bei ROSENTHAL (Fn. 4) und SEILER/GRIESINGER (Fn. 5); diese beiden Aufsätze enthalten auch viele weitere Literaturnachweise, die nachfolgend grundsätzlich nicht wiederholt werden. Thematisch nicht angesprochen werden hernach die besonders schützenswerten Personendaten, das Profiling, der grenzüberschreitende Datenverkehr, die Auftragsdatenbearbeitung, die Daten-Portabilität und die Datenschutz-Folgeabschätzung.

⁸ Art. 45 DSGVO; Entscheid vom 15. Januar 2024, Medienmitteilung des Bundesamtes für Justiz, <https://www.bj.admin.ch/bj/de/home/aktuell/mm.msg-id-99695.html>.

⁹ Vgl. Art. 6 i.V.m. Art. 31 DSG; dazu auch SEILER/GRIESINGER (Fn. 5) Rz. 10 mit Fn. 22.

in vielen praktischen Aspekten auf, doch gerade beim Einsatz von KI ergeben sich nicht selten unterschiedliche Rechtseinschätzungen, was die Konzentration dieses Aufsatzes auf das DSG legitimiert.¹⁰

2. DSG-Anknüpfungspunkte

[5] Die beiden grundlegenden Regelungsziele des DSG sind der Schutz von Personendaten und die Gewährleistung der Datensicherheit. Das Ziel der Regulierungen besteht darin, normative Vorgaben auf der Basis des Grundsatzes der Technologieneutralität zu realisieren.¹¹

2.1. Schutz von Personendaten

[6] Der Begriff der Personendaten ist im DSG sehr weit gefasst; nicht nur identifizierte Personen, sondern ebenso bestimmbare Personen genießen den Schutz des DSG. Der Begriff der Bearbeitung umfasst jeden Umgang mit Personendaten (z.B. das Beschaffen, Speichern, Aufbewahren, Verwenden, Verändern, Bekanntgeben, Offenlegen, Archivieren, Löschen und Vernichten der Daten, Art. 5 lit. d [DSG](#)). Ausserhalb des Anwendungsbereichs des DSG liegen deshalb nur die Sachdaten und die anonymisierten bzw. pseudonymisierten Daten¹² sowie die Datenbearbeitung zu nicht-personenbezogenen Zwecken.¹³

[7] Im Kontext der KI ist zu differenzieren zwischen den Personendaten als «Input» und den Personendaten als «Output»; die Differenzierung ist relevant, weil diejenigen Personen, welche die DSG-Vorschriften beachten müssen (Verantwortlicher oder Auftragsbearbeiter), nicht identisch sind.¹⁴ Im Lichte der vorerwähnten Phasen des KI-Einsatzes fallen der Aufbau eines KI-Systems, dessen Training und dessen Besitz/Betrieb in den «Input»-Bereich, während die möglichen Verwendungen zum «Output»-Bereich gehören.¹⁵

¹⁰ Relevante Unterschiede betreffen insbesondere die Rechtfertigungsgründe und die automatisierten Einzelentscheidungen.

¹¹ THOUVENIN/VOLZ (Fn. 5), S. 3.

¹² Eingehender zur Verschlüsselung von Trainingsdaten DANY PINA, Die Offenbarung von KI-Patenten, Zürich 2024, S. 46 ff. m.w.H.

¹³ Zu diesem im KI-Kontext wichtigen Bereich vgl. nachfolgend [Rz. 19 ff.](#)

¹⁴ Zu den verschiedenen «Input»- und «Output»-Faktoren eingehend ROSENTHAL (Fn. 4), Rz. 46 ff.; zur unterschiedlichen Verantwortlichkeit für diese Faktoren vgl. DERS., Rz. 18.

¹⁵ Im Einzelnen zu den Elementen eines KI-Modells vgl. DAVID ROSENTHAL, Was in einem KI-Modell steckt und wie es funktioniert, Blog Vischer, Teil 17, Mai 2024, <https://www.vischer.com/kuenstliche-intelligenz/>.

[8] Mit «Input» ist konkret gemeint, dass die Eingaben zur Verwendung eines KI-Systems keine geschützten Personendaten enthalten dürfen, sofern nicht sichergestellt ist, dass der Anbieter des KI-Systems sich bei der weiteren Arbeit mit dem KI-System an die Datenschutzvorgaben hält. Personendaten können in den von den Benutzern eingegebenen «Prompts» enthalten sein, aber sich auch aus dem Sammeln von Daten aus öffentlichen Quellen (durch ein sog. «Crawlen») ergeben.¹⁶ Seitens des KI-Anbieters dürfen die allenfalls übermittelten Personendaten wegen des Grundsatzes der Zweckbindung insbesondere nicht für andere Zwecke verwendet werden (z.B. das Training eigener Modelle)¹⁷ und es sind technische Vorsichtsmassnahmen geboten, wenn ein Outsourcing der Datenbearbeitung erfolgt oder wenn sich der Auftragsbearbeiter im Ausland befindet.

[9] Datenschutzrechtliche Problembereiche, welche die durch KI-Systeme verfügbar gemachten «Outputs» betreffen, sofern Personendaten enthalten sind, bestehen insbesondere mit Blick auf drei Erscheinungen:¹⁸ (i) KI-Systeme können «halluzinieren», d.h. Inhalte erfinden, die dem Benutzer als wahre Informationen erscheinen. (ii) Outputs von KI enthalten ggf. ungewollt Personendaten. (iii) Der Output der KI verzerrt allenfalls die Daten, d.h. es entsteht ein «Bias», der mit Blick auf die datenschutzrechtlichen Grundsätze der Richtigkeit, Verhältnismässigkeit sowie von Treu und Glauben zu Problemen führt. In allen drei Situationen sind die datenschutzrechtlichen Herausforderungen anhand der konkreten Gegebenheiten zu analysieren.

2.2. Datensicherheit

[10] Ein wichtiger Aspekt des Datenschutzrechts ist die Datensicherheit (Art. 8 [DSG](#)). Wie erwähnt¹⁹ dürfen beim Stichwort «Datensicherheit» nicht nur mögliche Risiken der KI in Betracht gezogen werden, sondern die Verwendung von KI-Systemen vermag auch zu einer Erhöhung der Datensicherheit beizutragen, weil solche Systeme strukturierte Programmabläufe ermöglichen.

[11] Als Risiken von KI-Programmen und algorithmischen Verfahren lassen sich die fehlende Vertraulichkeit der eingegebenen Daten, die mangelnde Faktizität und Reproduzierbarkeit, die ungenügende Aktualität, die fehlerhafte Reaktion auf spezifische Eingaben und die Anfälligkeit für «versteckte» Dateninfiltratio-

¹⁶ SEILER/GRIESINGER (Fn. 5), Rz. 8 und Rz. 31.

¹⁷ Zum Sonderproblem des Fine-Tuning eines KI-Systems vgl. ROSENTHAL (Fn. 4), Rz. 44 f.

¹⁸ ROSENTHAL (Fn. 4), Rz. 50 ff.

¹⁹ Vgl. vorne [Rz. 2](#).

nen mit manipulativer Absicht identifizieren. Konkrete Risikosituationen betreffen (i) das Social Engineering (z.B. durch Preisgabe von Daten, Umgehung von Schutzmassnahmen, Installierung eines Schadcodes), (ii) den Einsatz von Malware (z.B. Erzeugung eines Schadcode ohne viel Hintergrundwissen dank ausgereifterer Code-Generierungsfähigkeiten, selbst bei Vorhandensein polymorpher Schadsoftware) und (iii) das Auftreten einer sog. «Hoax» (z.B. Falschmeldung wegen Desinformation oder Propaganda).²⁰

[12] Die Datensicherheit ist mindestens so sehr eine technische wie eine rechtliche Herausforderung; aus diesem Grunde konzentrieren sich die nachfolgenden Ausführungen auf den vorgenannten Schutz von Personendaten.

3. Spannungsfelder zwischen künstlicher Intelligenz und Datenschutz

[13] Die Zahl der Spannungsfelder zwischen KI und Datenschutz ist gross; neben einer kurzen Erläuterung von einzelnen weniger zentralen Themenfeldern und Sonderfragen kommen schwergewichtig die Datenbearbeitungsgrundsätze, die Rechtfertigungsgründe, die automatisierten Einzelentscheidungen und die Verwendung von «Output»-Daten zur Sprache.

3.1. Datenbearbeitungsgrundsätze

[14] Das DSG geht (wie erwähnt im Gegensatz zur DSGVO) vom Grundsatz aus, dass eine Datenbearbeitung zulässig ist, wenn der Verantwortliche die gesetzlich statuierten Datenbearbeitungsprinzipien einhält, es gilt also kein Erlaubnisvorbehalt. Die Beschaffung und Bearbeitung von Personendaten setzen im Kontext der Entwicklung, des Trainings, der Validierung und des Testens von KI-Systemen aber insbesondere die Beachtung der folgenden Datenschutzprinzipien voraus:²¹

- *Grundsatz der Erkennbarkeit:* Oft ist die Beschaffung und Bearbeitung von Personendaten für die Betroffenen nicht erkennbar, weil die Daten vom KI-System aus verschiedenen Quellen bezogen werden, ohne dass die betroffenen Personen eine Information erhalten.

²⁰ WEBER (Fn. 1), S. 24.

²¹ Allgemein zu den Datenbearbeitungsgrundsätzen RETO FANGER, OFK-DSG, Kommentar, Zürich 2023, Art. 6 Rz. 4 ff.; zu den KI-spezifischen Themenstellungen vgl. THOUVENIN/VOLZ (Fn. 5), S. 4; SEILER/GRIESINGER (Fn. 5), Rz. 10.

- *Grundsatz der Zweckbindung*: KI-Systeme bearbeiten Daten oft zu einem anderen Zweck als demjenigen, der bei der ursprünglichen Beschaffung relevant gewesen ist, insbesondere im Falle eines «General Purpose AI-Systems».²²
- *Grundsatz der Datenminimierung*: Die Verwendung von mehr Daten durch KI-Systeme verbessert meist die Ergebnisse, d.h. den «Output», was aber dem Ziel der Datenminimierung widerspricht.
- *Grundsatz der Speicherbegrenzung*: Zwar werden die verwendeten Personendaten regelmässig nicht im trainierten KI-System gespeichert, doch entsteht ein DSGVO-Problem, sofern nach Abschluss der Entwicklung eine Speicherung in einer gesonderten Datenbank erfolgt.

[15] Ein weiterer zentraler Grundsatz der Datenbearbeitung ist die Datenrichtigkeit; der Verantwortliche muss sich vergewissern, dass die gesammelten Daten während der ganzen Aufbewahrungsdauer ihre Richtigkeit behalten.²³ In Frage steht also die inhaltliche Korrektheit der Daten und die Vermeidung von Fehlinformationen, was bei grossen KI-Systemen die Einrichtung umfangreicher technischer Sicherheitsvorkehrungen notwendig macht.

[16] Der Einsatz von KI führt nicht in jeder Situation zu einer ungenügenden Beachtung der Datenbearbeitungsgrundsätze, vielmehr sind Einzelfall-Analysen durchzuführen; für gewisse Zwecke von KI-Trainings mag die Verwendung möglichst vieler Informationen ungeachtet des Datenminimierungsgebots sinnvoll sein, um dem Grundsatz der Datenrichtigkeit Genüge zu tun.

[17] Überdies lässt sich die Datenbearbeitung durch Rechtfertigungsgründe legitimieren, die hernach im Einzelnen zu beurteilen sind.

3.2. Rechtfertigungsgründe

[18] Weil das DSGVO vom Grundsatz der Zulässigkeit der Bearbeitung von Personendaten im Falle der Einhaltung der Datenbearbeitungsprinzipien ausgeht bzw. im Falle der Nichteinhaltung den Nachweis einer Rechtfertigung der Datenbearbeitung einschränkt, haben die Rechtfertigungsgründe (Art. 31 [DSG](#))

²² Insoweit stellt sich das besondere Problem der Zweitnutzung bzw. der Zweitverwertung; vgl. dazu ROSENTHAL (Fn. 4), Rz. 28, mit weiteren Ausführungen.

²³ Allgemein dazu vgl. FANGER (Fn. 21), Rz. 11 m.w.H.; zu den KI-spezifischen Themenstellungen vgl. PAULINA J. PESCH/RAINER BÖHME, Verarbeitung personenbezogener Daten und Datenrichtigkeit bei grossen Sprachmodellen, MMR 12/2023, S. 917 ff. m.w.H.

eine grosse praktische Bedeutung. Nachfolgend wird der für KI-Systeme zentrale Aspekt der Datenbearbeitung zu nicht-personenbezogenen Zwecken gesondert, vor den allgemeinen Rechtfertigungsgründen, erläutert.²⁴

3.2.1. Datenbearbeitung zu nicht-personenbezogenen Zwecken

[19] Zu den überwiegenden Interessen des Verantwortlichen für eine Datenbearbeitung gehört der Rechtfertigungsgrund der Bearbeitung zu nicht-bezogenen Zwecken (Art. 31 Abs. 2 lit. e [DSG](#)),²⁵ im Vordergrund stehen dabei Forschung, Planung oder Statistik. KI-Systeme zielen in der Praxis oft darauf ab, nicht Erkenntnisse mit Bezug auf spezifische Personen zu gewinnen, sondern Programme zu entwickeln, welche bestimmte Aufgaben möglichst gut zu erfüllen in der Lage sind. Eine solche Bearbeitung ist nicht vornehmlich personenbezogen.

[20] Immerhin setzt dieser Rechtfertigungsgrund auch die Erfüllung weiterer Voraussetzungen voraus.²⁶ (i) Sobald der Bearbeitungszweck es erlaubt, sind die Daten zu anonymisieren; technische Massnahmen ermöglichen in der Praxis oft eine solche Anonymisierung von KI-Systemen. (ii) Im Falle der Bearbeitung von besonders schützenswerten Personendaten darf eine Bekanntgabe an Dritte nur so erfolgen, dass die betroffene Person nicht bestimmbar ist; meist lässt sich diese Voraussetzung wohl erfüllen. (iii) Die Veröffentlichung von Ergebnissen muss in einer Art erfolgen, dass die betroffenen Personen nicht bestimmbar sind. Ob diese Anforderung bei KI-Systemen erfüllt ist, hängt von der konkreten «Veröffentlichung» der entsprechenden Daten ab.

[21] Im Einzelnen ergeben sich aber schwierige Abgrenzungsfragen bei der Beurteilung der Datenbearbeitung zu «nicht-personenbezogenen Zwecken» oder auch beim «überwiegenden Interesse» an der Datenbearbeitung gemäss Art. 31 Abs. 2 [DSG](#); zutreffend nimmt die Lehre in neuester Zeit diese Rechtfertigungsgründe etwas genauer unter die Lupe.²⁷

[22] Weil der Auslegungsspielraum für die Gerichte relativ gross ist, stellt sich die Frage, ob der Gesetzgeber die gegebene Liste für das Vorliegen eines überwiegenden Interesses (Art. 31 Abs. 2 [DSG](#)) um ein weiteres Beispiel (Datenbe-

²⁴ ROSENTHAL (Fn. 4), Rz. 34 ff. erläutert diesen Rechtfertigungsgrund unter dem Titel «knifflige Fragen».

²⁵ Eingehender dazu THOUVENIN/VOLZ (Fn. 5), S. 4 f. und ROSENTHAL (Fn. 4), Rz. 35 ff., insbesondere Rz. 38.

²⁶ Vgl. auch THOUVENIN/VOLZ (Fn. 5), S. 4 f.

²⁷ Vgl. THOUVENIN/VOLZ (Fn. 5), S. 5; ROSENTHAL (Fn. 4), Rz. 36.

arbeitung für das Trainieren, Validieren und Testen von KI-Systemen) ergänzen oder ob der EDÖB in einem Leitfaden bzw. Merkblatt klarstellen sollte, dass die Inbetriebnahme eines KI-Systems nicht als Veröffentlichung von Ergebnissen gilt.²⁸ Einzelne Stimmen gehen demgegenüber von einer normativen Kraft des Faktischen bzw. einer «sozialen Realität» aus, die wohl darauf hinauslaufe, dass gewisse «Datenschutz-Beeinträchtigungen» durch KI-Systeme von der Gesellschaft mit der Zeit hingenommen würden.²⁹

3.2.2. Weitere Rechtfertigungsgründe

[23] (i) Der aus der Sicht des Gesetzgebers naheliegendste Grund für eine Rechtfertigung der Datenbearbeitung ist die Einwilligung der betroffenen Person. Im KI-Kontext hat dieser Grund indessen praktisch nur eine geringe Bedeutung: Die Einholung der Einwilligung des Betroffenen ist beim «Input» von Daten in KI-Systeme sehr schwierig und mit hohen Hürden verbunden;³⁰ deshalb ist es wenig erstaunlich, dass eine Umfrage des Vereins Unternehmens-Datenschutz bei seinen Mitgliedern im Jahre 2023 zum Ergebnis kam, die Einwilligung stelle bei den meisten Unternehmen nur in 10%-20% der Fälle einen Rechtfertigungsgrund dar.³¹

[24] Zudem führen selbstlernende Algorithmen über die Zeit zu Veränderungen in der Datenbearbeitung, die zu Beginn ggf. nicht vorausgesehen werden konnten; dementsprechend deckt sogar eine erfolgte Einwilligung neue, ursprünglich nicht bekannte Vorgänge in KI-Systemen nicht ab. Weil es somit an der Einwilligung im Rahmen von KI-Systemen regelmässig fehlt, haben andere Rechtfertigungsgründe eine grössere praktische Bedeutung.³²

[25] (ii) Die Berufung auf ein berechtigtes Interesse des Verantwortlichen ist – ausser in bestimmten Konstellationen mit besonders schützenswerten Personendaten – dann in Betracht zu ziehen, wenn das KI-System höherwertigen Zielen dient. Das Spannungsfeld zwischen DSG- und KI-Interessen, das sich nicht gestützt auf allgemeine Überlegungen, sondern vielmehr nur mit Blick

²⁸ THOUVENIN/VOLZ (Fn. 5), S. 5.

²⁹ Im Einzelnen ROSENTHAL (Fn. 4), Rz. 40–42; die Einwände gegen diese Betrachtungsweise entsprechen den Überlegungen zur sog. normativen Kraft des Faktischen.

³⁰ Vgl. SEILER/GRIESINGER (Fn. 5), Rz. 14 (wenn zwar in problematischer Weise relativierend DIES., Rz. 26).

³¹ Vgl. VEREIN UNTERNEHMENS-DATENSCHUTZ, Benchmarking 2023: Informationelle Selbstbestimmung, https://www.vud.ch/customer/files/180/VUD_Benchmarking-2023_Informationelle-Selbstbestimmung.pdf.

³² Vgl. dazu auch THOUVENIN/VOLZ (Fn. 5), S. 5; SEILER/GRIESINGER (Fn. 5), Rz. 16 ff.

auf die konkreten Umstände auflösen lässt, hat im Artificial Intelligence Act (AIA) der EU eine besondere Regelung erfahren: So findet sich in Art. 10 Ziff. 5 AIA eine Anordnung zur Bearbeitung von besonderen Personendaten-Kategorien für das Entdecken und Korrigieren von Verzerrungen («Biases»): Anbieter von KI-Systemen sind verpflichtet, angemessene Garantien für die Grundrechte und die Grundfreiheiten der Betroffenen einzurichten.³³

[26] (iii) Überdies ist in diesem Kontext auch das Spannungsfeld zwischen der öffentlichen Information und dem Datenschutz zu beachten (Art. 31 Abs. 2 lit. f [DSG](#)). KI-Systeme sammeln nicht nur Personendaten, sondern durch ein oft umfassendes «Crawlen» auch sog. öffentliche Informationen. Datenschutzrechtlich scheint das «Crawlen» auf den ersten Blick nicht problematisch zu sein, denn grundsätzlich verfügen Informationen aus öffentlich verfügbaren Quellen nicht über einen besonderen Schutz. Dennoch ist nicht zu übersehen, dass nach Auffassung des EDÖB auch veröffentlichte Personendaten weiterhin vom Datenschutz erfasst sind und nicht frei verwendet werden dürfen.³⁴

[27] Die Spezialregelung von Art. 30 Abs. 3 [DSG](#) privilegiert den Verwender öffentlicher Inhalte lediglich unter dem Vorbehalt, dass der Inhalt mit Zustimmung der betroffenen Person allgemein zugänglich gemacht und der Nutzung im Einzelfall nicht widersprochen worden ist.³⁵ Das Öffentlichkeitsinteresse an KI-Systemen allgemein ist überdies ein Thema des nachfolgend detaillierter zu erörternden Transparenzprinzips.³⁶

[28] In neuester Zeit stellen grosse KI-Anbieter zudem in Aussicht, in ihren (versteckten) Datenschutzbestimmungen eine Einwilligung der Vertragspartner vorzusehen, die besagt, dass auch «Informationen, die über die Produkte und Dienstleistungen von Meta geteilt werden», für die Bearbeitung in KI-Sys-

³³ Dazu International Association of Privacy Professionals (IAPP), The AI Act's debiasing exception to the GDPR, 21 February 2024, <https://iapp.org/news/a/the-ai-acts-debiasing-exception-to-the-gdpr>; zur Diskriminierungsproblematik vgl. allgemein FLORENT THOUVENIN/STEPHANIE VOLZ/SORAYA WEINER/CHRISTOPH HEITZ, Diskriminierung beim Einsatz Künstlicher Intelligenz (KI): Technische Grundlagen für Rechtsanwendung und Rechtsentwicklung, Rz. 4 ff., in diesem Jusletter.

³⁴ Vgl. EDÖB in Zusammenarbeit mit ausländischen Datenschutzbehörden: Gemeinsame Erklärung «Data Scraping» und Datenschutz, 24. August 2023, <https://datenrecht.ch/edoeb-et-al-gemeinsame-erklaerung-zu-data-scrapingund-datenschutz/>; vgl. auch SEILER/GRIESINGER (Fn. 5), Rz. 19.

³⁵ ROSENTHAL (Fn. 4), Rz. 30; vgl. auch DERS., Rz. 32 f.

³⁶ Vgl. hinten [Rz. 43 ff.](#)

temen herangezogen werden können;³⁷ erfasst sind damit Daten, Fotos usw., die von den Nutzern täglich auf den Plattformen hochgeladen werden; ob eine solche Einwilligung der Betroffenen den Klarheitsvorgaben des DSG Genüge tut, erscheint immerhin als sehr zweifelhaft und hat denn auch die EU-Datenschutzbehörden bereits auf den Plan gerufen.

3.3. Automatisierte Einzelentscheidungen

[29] Vorausschauend haben Art. 22 DSGVO und Art. 21 [DSG](#) schon vor einem breiten Einsatz von KI-Systemen eine Regelung zu den automatisierten Einzelentscheidungen eingeführt; inhaltlich geht der kurz zu erläuternde Art. 21 [DSG](#) weniger weit als der umfassendere Art. 22 DSGVO.³⁸ Bei der automatisierten Bearbeitung von Personendaten unterliegt der Verantwortliche einer Informationspflicht an die Betroffenen, falls die Bearbeitung eine rechtliche Wirkung für die betroffene Person entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt.³⁹ Automatisierte Bearbeitungen kommen insbesondere in KI-Systemen vor, die auf maschinellem Lernen basieren und die in der Lage sind, komplexe Vorhersagen oder Entscheidungen ohne menschliches Zutun zu treffen.⁴⁰

[30] Die betroffene Person hat das Recht, Widerspruch gegen eine automatisierte Einzelentscheidung einzulegen, ausser wenn diese Entscheidung für den Abschluss oder die Erfüllung eines Vertrags zwischen der betroffenen Person und dem Verantwortlichen erforderlich ist oder ein öffentlich-rechtlicher Rechtfertigungsgrund vorliegt.⁴¹ Die Auslegung dieser Gründe kann nach den allgemeinen Grundsätzen erfolgen, weil sie nicht KI-spezifisch sind.

[31] Ein für KI-Systeme besonders relevantes Problem von Art. 21 [DSG](#) liegt darin, dass der Gesetzeswortlaut verlangt, es müsse sich um «ausschliesslich» automatisierte Datenbearbeitungen (ohne Dazutun einer natürlichen Person) handeln (Abs. 1). Dieser Tatbestand mag bei ADM-Systemen («algorithmic de-

³⁷ Vgl. Tages Anzeiger vom 13. Juni 2024, S. 17; zur Intervention der EU-Datenschutzbehörden vgl. NZZ vom 25. Juni 2024, S. 22. Demgegenüber behauptet Apple in seinen öffentlichen Verlautbarungen, dem Datenschutz bzw. der Privatsphäre eine besondere Aufmerksamkeit zu schenken (vgl. NZZ vom 12. Juni 2024, S. 23).

³⁸ Für einen detaillierten Überblick vgl. FABIENNE SUTER, Automatisierte Einzelentscheidungen im (Schweizer) Datenschutzrecht, Diss. Zürich 2024, S. 245 ff.

³⁹ Vgl. dazu ADRIAN BIERI/JULIAN POWELL, OFK-DSG, Kommentar, Zürich 2023, Art. 21 Rz. 7 f.

⁴⁰ SEILER/GRIESINGER (Fn. 5), Rz. 56.

⁴¹ ROSENTHAL (Fn. 4), Rz. 6.

cision-making») und beim Profiling von Personen in der Regel vorliegen.⁴² Im Kontext des Einsatzes grosser KI-Systeme findet aber nicht selten zusätzlich noch ein nicht automatisiertes Monitoring im Sinne einer Plausibilitätsüberprüfung statt, was an sich die Erfüllung des Kriteriums «ausschliesslich» in Frage stellt. Die Lehre geht davon aus, dass nicht das «Können» einer Beeinflussung der Entscheidung durch eine natürliche Person massgebend ist, sondern die Frage, «ob eine Überprüfung [...] durch eine natürliche Person tatsächlich erfolgt, welche zudem rechtlich und faktisch in der Lage wäre, den maschinell getroffenen Entscheid umzustossen».⁴³

3.4. Verwendung von Personendaten

[32] Wie erwähnt, ist bei KI-Systemen zwischen dem «Input» und dem «Output» zu differenzieren. Auch bei der Verwendung von Personendaten («Output») sind in vergleichbarer Weise die zuvor erläuterten Grundsätze der Datenbearbeitung zu beachten. Zusätzlich gewinnt aber der Grundsatz der Richtigkeit der Daten an Bedeutung;⁴⁴ gerade im Kontext von KI-Systemen ist von besonderer Bedeutung, dass sich die Richtigkeit mit Blick auf den Zweck der Datenbearbeitung beurteilt und dieser Zweck dem Risiko einer «Halluzination» (z.B. durch frei erfundene Angaben über Personen) unterliegt.⁴⁵ Werden «Prompts» (mit Personendaten) von KI-Nutzern eingegeben und hernach verarbeitet, hat der Betreiber von KI-Systemen beim «Output»-Angebot oft die Rolle eines Auftragsbearbeiters.⁴⁶

[33] Gelingt es einem Dritten, durch gezielte Abfragen bzw. Attacken einzelne Personendaten von Betroffenen aus KI-Systemen zu entnehmen, die z.B. zu Trainingszwecken verwendet worden sind, liegt ggf. ein (unfreiwilliges) Bekanntgeben von Personendaten (Art. 5 lit. e [DSG](#)) vor. Das Bekanntgeben muss den Datenbearbeitungsgrundsätzen genügen; der Betreiber des KI-Systems unterliegt deshalb zudem dann einem Haftungsrisiko, wenn er den anwendbaren Datensicherheitsvorkehrungen nicht genügt.⁴⁷ Im Einzelnen können sich überlagernde Datenbearbeitungen aber schwierige Abgrenzungsprobleme hervorrufen: Der KI-Nutzer trägt die Verantwortung, wenn er die «Prompts» mit den Personendaten nicht sachgerecht gemäss den technischen Vorgaben

⁴² Vgl. SUTER (Fn. 38), S. 294 ff.

⁴³ BIERI/POWELL (Fn. 39), Rz. 6.

⁴⁴ Vgl. vorne [Rz. 15](#).

⁴⁵ THOUVENIN/VOLZ (Fn. 5), S. 7.

⁴⁶ Eingehender dazu ROSENTHAL (Fn. 4), Rz. 16 f.

⁴⁷ Vgl. vorne [Rz. 19 ff.](#)

des KI-Anbieters implementiert; der Betreiber des KI-Systems hat für die DSGVO-kompatible Bearbeitung der Personendaten einzustehen.⁴⁸

3.5. Informations- und Sorgfaltspflichten

[34] Die Verantwortlichen haben die Pflicht, die betroffenen Personen über die Verarbeitung ihrer Daten zu informieren (Art. 19 ff. [DSG](#)). Diese über die erwähnte Aufklärung zu den automatisierten Einzelentscheidungen⁴⁹ hinausgehenden Informationspflichten stellen einen zentralen Pfeiler des Datenschutzrechts dar. Einerseits wird durch die Informationen die Transparenz der Datenbearbeitung erhöht und andererseits ermöglichen die Informationen den betroffenen Personen die Ausübung ihrer Auskunfts- und Korrektur- bzw. Widerrufsrechte.

[35] Der Anbieter von KI-Systemen hat demgemäss die KI-Nutzer darüber zu informieren, wie ihre Daten konkret bearbeitet werden, unter Hinweis auf die Zwecke der Bearbeitung und die verwendeten Algorithmen.⁵⁰ Im Falle der Beurteilung von Kreditrisiken ist der Nutzer eines KI-Systems zum Beispiel darüber zu informieren, welche Daten (z.B. Kreditverlauf, Einkommen, Alter) konkret Verwendung finden, wie die Datenbearbeitung erfolgt und wie das System seine Vorhersagen trifft. Die Verfügbarkeit der Information ist im Übrigen auch deshalb relevant, weil ggf. Auskunftsansprüche vom Betroffenen geltend gemacht werden, die zur Herausgabe von Informationen zwingen.⁵¹

[36] Die Informationspflichten verursachen aber grosse Herausforderungen im Kontext der KI-Systeme; Algorithmen sind oft komplex und für Nicht-Experten schwer zu verstehen.⁵² Transparente Information bedeutet, dass gemäss den Prinzipien der Erklärbarkeit und Nachvollziehbarkeit den KI-Nutzern verständlich zu machen ist, wie und mit welchem Zweck die Datenbearbeitung abläuft.⁵³ Abgesehen von der in vielen Situationen vorliegenden Schwierigkeit, eine sachgerechte «Aufklärung» vorzunehmen, erweist sich auch eine Abwägung zwischen den Interessen der KI-Nutzer an der Kenntnis über die datenbezogenen Vorgänge und den Interessen der Anbieter

⁴⁸ Vgl. zur Sanktionierung von DSGVO-Verstössen auch DAVID ROSENTHAL, Datenschutz und KI: Worauf in der Praxis zu achten ist, in: Jusletter IT 26. April 2022, Rz. 55.

⁴⁹ Vgl. vorne [Rz. 29 ff.](#)

⁵⁰ SEILER/GRIESINGER (Fn. 5), Rz. 52 ff., v.a. Rz. 53.

⁵¹ Vgl. hinten [Rz. 48 f.](#)

⁵² Vgl. dazu auch SEILER/GRIESINGER (Fn. 5), Rz. 54.

⁵³ Eingehender dazu WEBER (Fn. 1), S. 22.

von KI-Systemen an der Vertraulichkeit individueller Programmierungen, die auf dem Konzept des Unternehmensgeheimnisses basiert, als notwendig.⁵⁴

[37] Überdies ermöglicht es die Technik, dass sich Algorithmen im Laufe der Zeit «selbstständig» anpassen bzw. verfeinern und damit das KI-System weiterentwickeln, was die Übermittlung von Informationen an die Nutzer des KI-Systems erschwert. In solchen Fällen wäre an sich eine repetitive Information sachgerecht; in welcher Art und Weise in dieser Konstellation die Information in klarer und nachvollziehbarer Weise bereitgestellt werden kann, um das Recht der Nutzer auf Information zu gewährleisten, stellt die Anbieter von KI-Systemen aber vor nicht zu unterschätzende Probleme und muss Anlass für eine verbesserte «Medienkompetenz» der Mitarbeitenden sein.⁵⁵

3.6. Verantwortlichkeit

[38] Fehlt es an der Beachtung der Datenbearbeitungs- oder Datenverwendungsgrundsätze im Kontext eines KI-Systems, stellt sich die Frage, wer die Verantwortung für eine etwaige DSGVO-Verletzung zu übernehmen hat. Im Vordergrund steht der Ersteller bzw. Besitzer und damit typischerweise der Anbieter, nicht der Kunde des KI-Systems; datenschutzrechtlich liegt der Anknüpfungspunkt nämlich bei der Festlegung der wesentlichen Parameter (Zweck und Mittel) der Datenbearbeitung (Art. 5 lit. j [DSG](#)).⁵⁶

[39] Die Haftungszuordnung ist deshalb mit Schwierigkeiten behaftet, weil oft nicht klar ist, wer in welchem Stadium ggf. Daten bearbeitet.⁵⁷ Auch der Auftragsbearbeiter kommt bei Datenbearbeitungen in der Cloud als verantwortliche Person in Betracht. Im Falle von reinen Hosting-Tätigkeiten vermag unter Umständen eine Haftungsbeschränkung gestützt auf die Grundsätze der Hosting-Provider-Haftung einzutreten.⁵⁸

⁵⁴ Vgl. dazu LUC DESAUNETTES, Von den Zünften zur künstlichen Intelligenz: Die Renaissance von Geschäftsgeheimnissen, Max-Planck-Gesellschaft, München 2018, https://www.mpg.de/12529808/ip_jb_2018.

⁵⁵ Zur Transparenz gegenüber den Betroffenen vgl. hinten [Rz. 43 ff.](#); zur «Medienkompetenz» vgl. auch ROSENTHAL (Fn. 4), Rz. 56.

⁵⁶ ROSENTHAL (Fn. 4), Rz. 20 und Rz. 22.

⁵⁷ Zu den Haftungsfragen im Besonderen vgl. ISABELLE WILDHABER, KI und Haftung: Lösungsansätze für die Schweiz, Rz. 20 ff., in diesem Jusletter IT.

⁵⁸ ROSENTHAL (Fn. 4), Rz. 23

4. Sonderfragen

4.1. Trainings-KI

[40] Trainingsdaten lassen oft keinen direkten Schluss auf bestimmbare Personen zu; soweit es am Kriterium der Bestimmbarkeit fehlt, entfallen die datenschutzrechtlichen Herausforderungen. In der Regel erfolgt die Verschlüsselung der Personendaten im Laufe bzw. am Ende des Lernprozesses; in neuerer Zeit wird aber vermehrt diskutiert und technologisch versuchsweise umgesetzt, Lösungen für eine Verschlüsselung bereits vor dem Lernprozess zu implementieren.⁵⁹ In einem solchen Fall wäre der Personenbezug nicht nachvollziehbar, was die DSGVO-Relevanz beseitigen würde.

[41] Wird der Personenbezug von Trainingsdaten nicht ausgeschlossen, ergibt sich bei ihrem Einsatz das Problem, dass der finale Bearbeitungszweck oft vom ursprünglichen Datenbearbeitungszweck abweicht. In einer solchen Konstellation stellt sich die Frage, ob eine Zweitverwertung vorliegt; verschiedene Formen der Zweitverwertung von Daten haben eine immer grössere Bedeutung und rufen nach einer datenschutzrechtlichen Beurteilung.⁶⁰

[42] Nach dem revidierten DSGVO (Art. 6 Abs. 3) muss der Zweck einer Zweitverwertung für die betroffene Person bereits bei der Beschaffung von Daten mindestens erkennbar sein; der «Sekundärzweck» (wie z.B. das Training der KI) ist deshalb dann vom Zweckbindungsgrundsatz gedeckt, wenn er mit dem ursprünglichen «Primärzweck» vereinbar und für den Betroffenen in der Entwicklung auch nachvollziehbar ist.⁶¹

4.2. Transparenz

[43] Ein grundlegendes Problem beim Einsatz von KI-Systemen besteht – wie erwähnt – in der mangelnden Transparenz mit Bezug auf deren Funktionsweise (z.B. Einsatz von Algorithmen, angewendete Methoden); akzentuiert wird das Risiko der Intransparenz bei den kaum steuerbaren selbstlernenden Algorithmen.⁶² Die

⁵⁹ Vgl. PINA (Fn. 12), S. 46 f.

⁶⁰ ROSENTHAL (Fn. 4), Rz. 28.

⁶¹ Vgl. auch THOUVENIN/VOLZ (Fn. 5), S. 4.

⁶² WEBER (Fn. 1), S. 21.

Transparenz ist denn auch im dichten Regulierungswerk des AIA der EU ein zentrales Thema.⁶³

[44] Selbst soweit es datenschutzrechtlich gesetzliche Transparenzanforderungen gibt, wie z.B. mit Blick auf automatisierte Entscheidungen (Art. 21 [DSG](#)), ist deren Wirkung nicht zu überschätzen.⁶⁴ Empirische Untersuchungen zu den vorhandenen daten- und verbraucherschutzrechtlichen Informationspflichten lassen nämlich daran zweifeln, ob die gesetzlich geforderte Transparenz stets einen Mehrwert für die betroffenen Personen generiert und ihren angedachten Zweck erfüllt.⁶⁵ Die zur Verfügung gestellten Informationen sind oft nicht ausreichend verständlich und überfordern angesichts der Informationsintensität (Länge, Detaillierungsgrad) die betroffenen Personen.⁶⁶

[45] Weiter müssen KI-Systeme grundsätzlich erklärbar und interpretierbar sein, um den betroffenen Personen die Nachvollziehbarkeit von algorithmischen Vorgängen zu ermöglichen. Notwendig ist somit, abhängig vom Wissenshorizont des Betroffenen und der Bedeutung der Vorgänge, eine ausreichende Verständlichkeit, d.h. die zugrunde liegende Logik hat auch begreiflich zu sein, was bei komplexen technischen Vorgängen nur schwer zu erreichen ist.⁶⁷

[46] Das DSG schiebt beim Einsatz von KI-Systemen im Rahmen der Bearbeitung von Personendaten nicht ausdrücklich eine Information der Betroffenen vor. Bearbeiten indessen KI-Anbieter die Daten von betroffenen Personen in einer Art und Weise, mit welcher nicht hat gerechnet werden müssen, und die sich in relevanter Weise negativ auswirken kann, macht der Grundsatz der Transparenz eine Information erforderlich, selbst wenn Art. 19 [DSG](#) nicht direkt als Rechtsgrundlage herangezogen werden kann.⁶⁸

4.3. Betroffenenrechte

[47] Bei den Betroffenenrechte stehen die Auskunftsrechte sowie die Ansprüche auf Widerspruch und Korrektur im Vordergrund.

⁶³ Im Einzelnen dazu DARIA BOHATCHUK/ALFRED FRÜH, Transparenz im Fokus der Europäischen KI-Verordnung, in: Jusletter 12 Februar 2024, Rz. 6 ff.

⁶⁴ Vgl. vorne [Rz. 36](#).

⁶⁵ Dazu im Einzelnen WEBER (Fn. 1), S. 21.

⁶⁶ Vgl. auch SEILER/GRIESINGER (Fn. 5), Rz. 33–35 und ROSENTHAL (Fn. 4), Rz. 55 ff.

⁶⁷ WEBER (Fn. 1), S. 22.

⁶⁸ Vgl. ROSENTHAL (Fn. 4), Rz. 57.

4.3.1. Auskunftsrecht

[48] Das Auskunftsrecht setzt voraus, dass der Betreiber eines KI-Systems tatsächlich Personendaten verarbeitet; die betroffene Person hat diesen Nachweis zu erbringen, was in der Realität oft nicht einfach sein dürfte, weil meist unklar ist, wer oder was in einem KI-System zu finden ist bzw. wer welche Datenbearbeitung vorgenommen hat.⁶⁹

[49] Wie der Verantwortliche im Einzelfall seiner Auskunftspflicht nachzukommen vermag, ist technisch noch wenig geklärt;⁷⁰ vorgenommene Tests haben nicht wirklich befriedigende Resultate bzw. konkrete Antworten z.B. von ChatGPT gebracht.⁷¹ Oft dürfte der Anbieter eines KI-Systems die Auskunft geben, dass keine Personendaten gesammelt worden seien, mangels Transparenz fehlt dem KI-Nutzer indessen der Einblick in das KI-System und das Auskunftsbegehren droht in die Leere zu laufen. Würde in Betracht gezogen, die Auskunft auf alle bisher generierten Antworten zur betroffenen Person in einem bestimmten KI-System zu erstrecken, wären, jedenfalls wenn es sich um ein grosses System handelt, die entsprechenden Antworten nicht oder nur beschränkt verfügbar, weil die Aufbewahrungszeit typischerweise relativ kurz ist.⁷²

4.3.2. Ansprüche auf Widerspruch und Korrektur

[50] In der Praxis einfacher durchzusetzen sind die Widerspruchs- und Korrekturansprüche (Recht auf Vergessen).⁷³ Gewisse Anbieter, z.B. OpenAI für ChatGPT, stellen ein entsprechendes Formular auf der Webseite zur Verfügung.⁷⁴ Hernach wäre über das Mittel des Auskunftsrechts zu prüfen, ob die Löschung auch wirklich durchgeführt worden ist.⁷⁵

[51] Alternativ zur Löschung erscheint es als denkbar, dass der Anbieter des KI-Systems mittels technischer Massnahmen sicherstellt, dass die betroffenen

⁶⁹ ROSENTHAL (Fn. 4), Rz. 59 f.; leicht relativierend THOUVENIN/VOLZ (Fn. 5), S. 6.

⁷⁰ Vgl. auch CHANTAL LUTZ/BENJAMIN DOMENIG/ANJA FLÜKIGER, Datenschutzkonformer Einsatz von ChatGPT an Schulen, sic! 2024, S. 94, S. 96.

⁷¹ ROSENTHAL (Fn. 4), Rz. 63.

⁷² Eingehender dazu ROSENTHAL (Fn. 4), Rz. 64 f.

⁷³ Vgl. auch THOUVENIN/VOLZ (Fn. 5), S. 6.

⁷⁴ Vgl. https://share.hsforms.com/1UPy6xqxZSEq.TrGDh4ywo_g4sk30.

⁷⁵ Vgl. auch LUTZ/DOMENIG/FLÜKIGER (Fn. 70), S. 96.

Personendaten im «Output» des Systems nicht mehr vorkommen; damit wird derselbe Zweck erreicht wie mit der tatsächlichen Löschung.⁷⁶

5. Ausblick

[52] Angesichts der vorerwähnten, nicht zu vernachlässigenden Rechtsunsicherheiten im Verhältnis zwischen Datenschutz und KI erscheint es als sinnvoll, wenn sich Anbieter und Betreiber von KI-Systemen, die Personendaten verarbeiten, eine Checkliste bereitlegen, um die DSGVO-Herausforderungen möglichst weitgehend erfüllen zu können. Folgende Fragen sind insbesondere für Betreiber mittlerer und grosser KI-Systeme relevant:⁷⁷

- Verarbeitet das KI-System auch Personendaten?
- Ist die Funktionsweise des KI-Systems nachvollziehbar?
- Ersetzt das KI-System ein anderes System?
- Kann das KI-System direkt oder indirekt negative Auswirkungen auf besonders schützenswerte Personendaten haben?
- Wird der Grundsatz der Datenminimierung eingehalten?
- Ist die datenschutzrechtliche Rollenverteilung klar festgelegt?
- Nimmt das KI-System automatisierte Einzelentscheidungen vor?
- Wird ein Large Language Model in der Datenbearbeitung verwendet?
- Erhält der Nutzer adäquate Informationen über den Einsatz von KI-Programmen?
- Auf welcher Grundlage erfolgt der Einsatz des KI-Systems?
- Ist der Einsatz des KI-Systems überhaupt grundrechtskonform möglich?
- Liegen transparente und nachvollziehbare Informationen über die Algorithmen vor?
- Sind angemessene Datensicherheitsvorkehrungen eingerichtet?
- Werden die Grundsätze der «Privacy by Design» und der «Privacy by Default» eingehalten?

⁷⁶ ROSENTHAL (Fn. 4), Rz. 64.

⁷⁷ Vgl. auch HANS-JÜRGEN POLLIRER, Checkliste Künstliche Intelligenz und Datenschutz, Datenschutz konkret 4/2023, S. 87, S. 89 f.

- Ist die Vornahme einer Datenschutz-Folgeabschätzung nötig? Wenn ja, wird sie durchgeführt?
- Erfolgt die Erfüllung der Informationspflichten sachgerecht?
- Können die Auskunfts- und Widerspruchsrechte der Betroffenen eingehalten werden?

[53] Die Beachtung einer solchen Checkliste bedeutet nicht, dass die DSG-Kompatibilität eines KI-Systems zweifelsfrei sichergestellt ist, aber zumindest wird ein sinnvoller Beitrag zur DSG-Vereinbarkeit der Bearbeitung von Personendaten geleistet. Wichtig ist in jedem Fall, dass der Betreiber eines KI-Systems intern Vorsorge trifft, dass sich die Mitarbeitenden der Datenschutz-Problematik bewusst sind und entsprechend handeln. Das Schweizer DSG statuiert nicht unüberwindbare Hürden (jedenfalls weniger als die DSGVO im grenzüberschreitenden Datenverkehr); mit angemessenen Datenschutzvorkehrungen lassen sich KI-Systeme durchaus erfolgreich nutzen.

*

Eine verwandte Diskussion der Thematik findet sich im vorliegend nicht erneut abgedruckten Beitrag von Rolf H. Weber, Datenschutz-Compliance beim Einsatz von KI-Systemen, in: Janisch/Krempelmeier/Mader/Staudegger/Thiele (Hrsg.), Am Glasfaserpuls, Festschrift für Dietmar Jähnel, Wien 2025, 749–762.

Juristische Sprachmodelle zwischen Transparenz und Datensicherheit

Erstveröffentlichung in: Schweighofer et al. (Hrsg.), *Sprachmodelle: Juristische Papageien oder mehr?* (IRIS 2024), Bern 2024, 19–26

Juristische Sprachmodelle kommen vermehrt und in umfangreicherem Ausmass bei der Erbringung von Rechtsdienstleistungen zum Einsatz. Oft erfahren die Transparenz und die Datensicherheit aber nur eine ungenügende Beachtung. Diese beiden Ziele sind zudem nicht deckungsgleich. Der Beitrag analysiert die eintretenden Spannungsfelder und schlägt Vorkehren zur Risikominimierung beim Einsatz von rechtlichen Sprachmodellen vor, etwa zur Gewährleistung der Datenqualität oder zum Schutz vor Angriffen (z.B. Datenschutzverletzungen)

Inhaltsübersicht

1.	Einleitung	150
2.	Wesen, Charakteristiken und Risiken juristischer Sprachmodelle	150
2.1.	Wesen und Charakteristiken	150
2.2.	Risiken	151
3.	Transparenz und Nachvollziehbarkeit	153
3.1.	Transparenz	153
3.2.	Nachvollziehbarkeit	154
4.	Datensicherheit	155
4.1.	Chancen und Nutzen für die IT-Sicherheit	155
4.2.	Risiken und Missbrauch bei der Nutzung	156
5.	Entwicklung transparenter und sicherer Sprachmodelle im Rechtsbereich	156
5.1.	Überbrückung unterschiedlicher Interessenlagen	157
5.2.	Konkrete Vorgaben für transparente und sichere Sprachmodelle	158
6.	Ausblick	160
7.	Literatur	160

1. Einleitung

Die juristischen Sprachmodelle bzw. allgemein die Large Language Models (LLM) haben jüngst erheblich an Bedeutung gewonnen. Solche Modelle sind in der Lage, die natürliche Sprache durch Künstliche Intelligenz (KI) zu verarbeiten, und gestützt darauf sachbezogene Texte zu entwickeln.¹ Dank des Trainings mit grossen Datenmengen lassen sich wiederkehrende Muster erkennen, z.B. im Falle oft aufeinanderfolgender Worte in einem bestimmten Kontext.²

Die technischen Innovationen bzw. gar «Umwälzungen» führen auch zu neuen rechtlichen Herausforderungen. Abgesehen vom zentralen Thema des Urheberrechtsschutzes, das einer gesonderten Behandlung bedürfte,³ wird nachfolgend aus der Vielzahl der rechtlichen Fragestellungen insbesondere auf die Transparenz und die Datensicherheit sowie deren Spannungsverhältnis eingegangen. Bewusst sind die Überlegungen allgemein formuliert und nicht auf die heute am meisten eingesetzten Programme (z.B. ChatGPT,⁴ Google Bard, usw.) konzentriert.

2. Wesen, Charakteristiken und Risiken juristischer Sprachmodelle

2.1. Wesen und Charakteristiken

Juristische Sprachmodelle zeichnen sich dadurch aus, dass sie die natürliche Sprache in schriftlicher Form auf der Basis von maschinellem Lernen verarbeiten und auch als Text präsentieren.⁵ Weil juristische Arbeit oft sehr sprachorientiert ist, sind rechtliche LLM ein geeignetes Unterstützungsinstrument.⁶ Entsprechende Sprachmodelle lassen sich vielfältig in allen Bereichen einsetzen, in denen eine (teil-)automatisierte Textverarbeitung und/oder Textproduktion stattfindet. Beispiele sind:⁷

¹ GRUETZEMACHER, 2022, *passim*.

² VOKINGER et al., 2022, 2.

³ Dazu grundlegend STRAUB, 2023, Rz 2 ff.

⁴ Eine umfangreiche Sammlung von Materialien (inkl. Podcasts) zur Funktionsweise von und zu rechtlichen Fragestellungen mit Bezug auf ChatGPT lassen sich unter <https://www2.weblaw.ch/home/academy/chatgpt-archiv> (accessed on 17 October 2023) finden.

⁵ Bundesamt für Sicherheit in der Informationstechnik, 2023, 5; BARENKAMP, 2023, 1.

⁶ SELMAN et al., 2023, 289 ff.

⁷ Bundesamt für Sicherheit in der Informationstechnik, 2023, 7.

- *Textgenerierung*: Verfassen von Entwürfen für Dokumente, Verfassen von Texten in einem bestimmten Schreibstil, Werkzeuge zur Textfortführung oder -vervollständigung;
- *Textbearbeitung*: Rechtschreib- und Grammatikprüfung;
- *Textverarbeitung*: Wort- und Textklassifikation, Markierung von Begriffen im Text, Textzusammenfassung, Frage-Antwort-Systeme, Übersetzung;
- *Programmcode*: Werkzeuge zur Unterstützung beim Programmieren, Erzeugen von Programmcode, Umprogrammierung und Übersetzung eines Programmes in andere Programmiersprachen.

Die konkrete Ausgestaltung der einzelnen juristischen Sprachmodelle hängt von den beabsichtigten Einsatzformen und den angestrebten Innovationen ab. Der Einsatz der Sprachmodelle im Rechtswesen vermag insbesondere die Erledigung von standardisierten Aufgaben zu erleichtern.⁸ Denkbar sind vollautomatisierte Prozesse, die direkt zur Rechtsprognose bzw. zur Entscheidung führen, oder teilautomatisierte Prozesse, die nur Entscheidungsgrundlagen bzw. Empfehlungen generieren.

2.2. Risiken

Juristische Sprachmodelle sind nicht zu unterschätzenden Risiken ausgesetzt. Large Language Models generieren in Regel einen sprachlich fehlerfreien und meist auch inhaltlich überzeugenden Text. Diese Tatsache darf aber nicht den Eindruck eines menschenähnlichen Leistungsvermögens erwecken und ein unbegrenztes Vertrauen in die Aussagen begründen. Als allgemeine Risiken sind, ungeachtet der durch mangelnde Transparenz oder ungenügende Datensicherheit hervorgerufenen Herausforderungen, folgende Problembereiche zu beachten:⁹

- *Fehlende Faktizität und Reproduzierbarkeit*: Die Generierung von Text beruht auf stochastischen Korrelationen, d.h. das «Wissen» wird aus (bereits gesehenen) Texten abgeleitet. Bezüge zur realen Welt existieren für das Modell nicht; deshalb kann es bei für Menschen verständlichen Sachverhalten zu inkorrekten Aussagen kommen oder ggf. fallen Aussagen ge-

⁸ Vgl. auch BARENKAMP, 2023, 2; zu den Besonderheiten der Automatisierung von Rechtsprechung vgl. GLESS, 2023, 434 ff.

⁹ Bundesamt für Sicherheit in der Informationstechnik, 2023, 10 f.

stützt auf dieselbe Eingabe aufgrund des wahrscheinlichkeitsbasierten Ansatzes unterschiedlich aus.

- *Fehlende Aktualität*: Fehlt ein Zugriff auf Live-Internetdaten, d.h. auf Informationen über aktuelle Ereignisse, produziert das Sprachmodell aufgrund von in der Vergangenheit geprägten Texten; die Ergebnisse sind diesfalls nicht aktuell.
- *Fehlerhafte Reaktion auf spezifische Eingaben*: Fehler vermögen aufzutreten, wenn die Sprachmodelle gewisse Eingaben verarbeiten, die so stark von den Texten in den Trainingsdaten abweichen, dass sie sich nicht mehr korrekt als Text bzw. Word-Dokument verarbeiten lassen (im Schrifttum zum Teil «Hallucination»¹⁰ genannt). Abweichungen können zufällig entstehen oder absichtlich als Täuschungsakt integriert werden.
- *Anfälligkeit für «versteckte» Eingaben mit manipulativer Absicht*: Gefahrgeneigt sind Modelle, die als Input ungeprüfte Dokumente Dritter enthalten, oder Situationen, in denen Angreifende für Nutzende un bemerkt gewisse Eingaben in das Sprachmodell einzubringen vermögen, insbesondere wenn der Betrieb den Zugriff auf Live-Daten aus dem Internet erlaubt; ein solcher Angriff kann z.B. ein Chat-Tool betreffen, das eine Person beim Surfen im Internet unterstützt.
- *Vertraulichkeit der eingegebenen Daten*: Bei der Nutzung von externen Schnittstellen fließen alle Eingaben, die im Sprachmodell verwendet werden, zunächst an den Betreiber des Modells ab. Die Nutzung externer Schnittstellen sollte deshalb bei der Verarbeitung von sensiblen und vertraulichen Informationen mit besonderen Schutzvorkehrungen verknüpft sein.

Eine grosse Herausforderung stellt die oft fehlende Datenhoheit bei der Verwendung von juristischen Sprachmodellen dar; überdies ist eine meist grosse Abhängigkeit vom Hersteller und Betreiber des Modells (insbesondere in technischer Hinsicht) problematisch.

¹⁰ Für Einzelheiten vgl. ZHANG et al., 2023, 2 ff.

3. Transparenz und Nachvollziehbarkeit

3.1. Transparenz

Ein grundlegendes Problem ist die mangelnde Transparenz in der Funktionsweise von juristischen Sprachmodellen. Intransparenz vermag hinsichtlich des Einsatzes von Algorithmen selbst zu bestehen, aber auch mit Blick auf die Verwendung von Daten betroffener Personen. Akzentuiert wird das Risiko der Intransparenz bei selbstlernenden Algorithmen, bei denen je nach angewandter Methode des Machine Learning angesichts des gegenwärtigen Standes der Technik u.U. selbst der Systembetreiber den Lösungsweg nicht mehr nachvollziehen kann.¹¹

Die primär- und verfassungsrechtlichen Regeln und Vorgaben zum Zugang zu Dokumenten können für mehr Transparenz sorgen, doch nicht einzelne Vorgänge erklären. Im Falle der Verwendung von Personendaten erlangen die Datenschutzbestimmungen Relevanz. «Automatisierte Entscheidungen» sind, wenn das Erfordernis der Vollautomatisierung erfüllt ist, grundsätzlich verboten (Art. 22 DSGVO);¹² ob bzw. inwieweit tatsächlich Personendaten verarbeitet werden, hängt von den gegebenen Umständen ab.

Konkrete Vorgaben sind im EU-Sekundärrecht vorgesehen; so verlangt Art. 5 der Verordnung (EU) 2019/1150 von den Anbietern, die wichtigsten ein Ranking bestimmenden Hauptparameter, die Gründe für deren relative Gewichtung gegenüber anderen Parametern und die potentielle Beeinflussung des Rankings durch Entgeltsleistungen transparent zu machen.¹³

Selbst wenn Transparenzerfordernisse von Bedeutung sind, um die betroffenen Personen von den mit Sprachmodellen verbundenen Risiken zu schützen, ist deren Wirkung nicht zu überschätzen. Empirische Untersuchungen zu den vorhandenen daten- und verbraucherschutzrechtlichen Informationspflichten lassen daran zweifeln, ob Transparenzerfordernisse stets einen Mehrwert für die betroffenen Personen generieren und ihren angedachten Zweck tatsächlich erreichen.¹⁴ Die erhaltenen Informationen sind oft nicht ausreichend verständlich und überfordern angesichts des Überflusses an Informationen die

¹¹ WEBER/HENSELER, 2020, 30; BARENKAMP, 2023, 3.

¹² In der Schweiz geht Art. 19 DSG weniger weit als Art. 22 DSGVO (vgl. dazu WEBER/HENSELER, 2020, 35).

¹³ Verordnung (EU) 2019/1150 zur Förderung von Fairness und Transparenz für gewerbliche Nutzer von Online-Vermittlungsdiensten, ABl 2019 L 186/57 vom 11. Juli 2019.

¹⁴ WEBER/HENSELER, 2020, 38.

Empfänger (sog. «Information Overload»),¹⁵ Diese allgemeine Problematik verstärkt sich angesichts der hohen Komplexität selbstlernender System- und Sprachmodelle, nicht zuletzt auch angesichts der Tatsache, dass eine eigentliche Erklärung der Verarbeitungsschritte bei Machine Learning-Verfahren bisweilen technisch nur schwer zu bewerkstelligen ist.¹⁶

Schliesslich ist nicht zu übersehen, dass selbst eine ausreichende Transparenz bei automatisierten Entscheidungen die menschliche Nachkontrolle nicht vollständig zu ersetzen vermag.

3.2. Nachvollziehbarkeit

Die Arbeitsweise von juristischen Sprachmodellen muss erklärbar und interpretierbar sein;¹⁷ die von der Datenbearbeitung betroffenen Personen sind in die Lage zu versetzen, die algorithmischen Vorgänge nachvollziehen zu können. Von grosser Bedeutung ist die Erklärbarkeit insbesondere in den Bereichen der Gesundheit, der Arbeitswelt, der Finanzen und der Justiz.

Erklärbarkeit bedingt, abhängig vom Wissenshorizont des Betroffenen und der Bedeutung der Vorgänge, ausreichende Verständlichkeit; die zugrundeliegende Logik muss begreiflich sein und die notwendigen Informationen enthalten, um einen ausreichenden Grad an Nachvollziehbarkeit zu erreichen.¹⁸ Dieses Ziel ist in einer hoch technologisierten Umgebung nur schwer zu erreichen, doch sind die Rahmenbedingungen durch die entsprechend konkrete Ausgestaltung der Sprachmodelle zu schaffen, damit die Betroffenen mit vernünftigen Aufwand die Vorgänge zu verstehen in der Lage sind.¹⁹

Erklärbarkeit und Nachvollziehbarkeit haben einen Bezug zur Schaffung von Transparenz innerhalb des Kreises der Betroffenen von eingesetzten LLM; die Vertraulichkeit bzw. der Geheimnisschutz betrifft «nur» die Drittbeziehungen.

¹⁵ WEBER, 2023, 67 ff. m.w. H.

¹⁶ WEBER/HENSELER, 2020, 39.

¹⁷ DANILEVSKY et al., 2022, *passim*.

¹⁸ WEBER, 2022, B 12.

¹⁹ Bundesamt für Sicherheit in der Informationstechnik, 2023, 8.

4. Datensicherheit

4.1. Chancen und Nutzen für die IT-Sicherheit

Juristische Sprachmodelle können zur Verbesserung der IT-Sicherheit beitragen. Als Chancen der Nutzung von juristischen Sprachmodellen lassen sich nennen:²⁰

- *Unterstützung bei der Detektion unerwünschter Inhalte:* Sind juristische Sprachmodelle für Textklassifikationsaufgaben geeignet, ergeben sich auch Anwendungsmöglichkeiten im Bereich der Detektion von Spam-Mails oder unerwünschten Inhalten.
- *Unterstützung bei der Textverarbeitung:* Sprachmodelle im Bereich der Textanalyse und -strukturierung, sind geeignet, grössere Mengen von Text zu verarbeiten und bei der Berichterstellung zu Sicherheitsvorfällen unterstützend mitzuwirken.
- *Unterstützung bei der Erstellung und Analyse von Programmcode:*²¹ Sprachmodelle lassen sich dazu einsetzen, die vorhandenen Programme auf Sicherheitslücken zu untersuchen sowie Wege zur Ausnutzung dieser Schwächen für Angriffe oder zur Codeverbesserung vorzuschlagen.
- *Unterstützung bei der Analyse von Datenverkehr:*²² Nach zusätzlichen Trainings vermögen Sprachmodelle auch bei Aufgaben zu unterstützen, die nicht natürlich-sprachigem Text im engeren Sinne entsprechen, aber doch die Detektion von böartigem Netzwerk-Verkehr erleichtern.

Chancen und Nutzen hängen aber von der Benutzerfreundlichkeit der Anwendungen der verschiedenen Modelle ab; regelbasierte Systeme (z.B. Experten-Entscheidungsbäume) können dabei hilfreich sein.²³ Die Schnittstellen der einzelnen Modelle müssen so ausgestaltet sein, dass sich Brücken zwischen den einzelnen Modellen bilden lassen.²⁴

²⁰ Bundesamt für Sicherheit in der Informationstechnik, 2023, 9; vgl. auch ZELLERS et al., 2020, *passim*.

²¹ BUBECK et al., 2023, 1 ff.

²² ALMODOVAR et al., 2022, 1 ff.

²³ Vgl. GLESS, 2023, 440 f.

²⁴ Das Kriterium der Nachvollziehbarkeit ist in vielen Bereichen, die von neuen Technologien geprägt sind, zu einem Problem geworden. Wer versteht die Code-basierte Ausgestaltung eines Smart Contract im Zeitpunkt seines Abschlusses? Wer vermag die mit Künstlicher Intelligenz durchgeführten Datenanalysen nachzuvollziehen? Diesem Problem wird die Rechtswissenschaft vermehrt Beachtung schenken müssen.

4.2. Risiken und Missbrauch bei der Nutzung

Wie erwähnt sind juristische Sprachmodelle verschiedenen Risiken und Missbrauchsgefahren ausgesetzt. Abgesehen von den bereits erläuterten Risiken, die von juristischen Sprachmodellen verursacht werden, ist in der Realität nicht auszuschliessen, dass die Produktion von Text für böswillige Zwecke missbraucht wird. Als Beispiele lassen sich nennen:²⁵

- *Social Engineering*: Cyber-Angriffe, bei denen Kriminelle versuchen, ihre Opfer zu veranlassen, Daten preiszugeben, Schutzmassnahmen zu umgehen oder selbstständig einen Schadcode zu installieren, fallen unter den Begriff des Social Engineering.
- *Erhöhte Risiken wegen Malware*: Neuere Sprachmodelle besitzen immer ausgereifere Code-Generierungsfähigkeiten, die es den Kriminellen mit geringen technischen Fähigkeiten ermöglichen, einen Schadcode ohne viel Hintergrundwissen zu erzeugen; polymorphe Schadsoftware gibt es zwar bereits seit einiger Zeit, doch ist heute die Generierung von Malware «leichter».
- *Hoax (Falschmeldung)*: Angesichts der grossen Zahl an Daten, die sich nicht vollständig überprüfen lassen, verbleiben regelmässig Texte mit fragwürdigem Inhalt (z.B. Desinformation, Propaganda) in der Trainingsmenge; sie tragen zu einer unerwünschten Struktur des Modells bei, die eine Neigung zu potentiell kritischen Inhalten zeigt.²⁶

Um die Sicherheit juristischer Sprachmodelle zu gewährleisten, sind Massnahmen gegen mögliche Risiko- oder Missbrauchsfälle rechtzeitig einzurichten.²⁷ Gegenmassnahmen können sowohl technischer als auch organisatorischer Art sein; Ziel ist die Absicherung der Authentizität von Texten und Nachrichten sowie der Nachweis, dass bestimmte Texte oder Informationen tatsächlich von einer bestimmten Person, Personengruppe oder Institution stammen.

5. Entwicklung transparenter und sicherer Sprachmodelle im Rechtsbereich

Der immer umfangreichere Einsatz von juristischen Sprachmodellen zwingt zu Anstrengungen, den beiden Anliegen der Transparenz und der Datensi-

²⁵ Bundesamt für Sicherheit in der Informationstechnik, 2023, 12 f.

²⁶ Vgl. auch WEIDINGER et al., 2022, 1 ff.

²⁷ Bundesamt für Sicherheit in der Informationstechnik, 2023, 13 f.

cherheit eine grössere Beachtung zu schenken. Nur wenn die Betroffenen davon ausgehen können, dass die algorithmischen Abläufe und Vorgehensweisen nachvollziehbar sind und die eingegebenen Daten ein hohes Mass an Datensicherheit aufweisen, ist es vertretbar, sich im beruflichen Alltag auf Large Language Models abzustützen.

5.1. Überbrückung unterschiedlicher Interessenlagen

(i) *Transparenz* und *Nachvollziehbarkeit* bedeuten, dass klare Verfahrensgrundsätze mit Blick auf den Einsatz von rechtlichen Sprachmodellen implementiert werden. Transparent ist ein Modell, wenn es Klarheit schafft.²⁸ Zu umschreiben sind die typischen Tätigkeitsbereiche von LLM im Rechtswesen (z.B. Verfassen von Dokumenten, Beratungstätigkeit, Literaturrecherche, Administration);²⁹ innerhalb der einzelnen Tätigkeitsgebiete lassen sich Ausmass und Art des Einsatzes von Sprachmodellen festlegen. Über die anwaltlichen Qualitätskriterien an die Leistungserbringung hinaus bedürfen insbesondere die Vorgaben des Datenschutzes und des Berufsgeheimnisses einer besonderen Beachtung.³⁰

Im Rechtsbereich stellen sich spezifische Herausforderungen mit Blick auf die Validierung von Personendaten (insbes. Datenrichtigkeit) sowie auf die Rückverfolgung der Gewichtung von Parametern innerhalb des Modells; zudem sind Methoden zu nutzen, die eine Angabe von Rechtsquellen ermöglichen. Ausgeschlossen sein sollte eine Nutzung von rechtlichen Sprachmodellen, welche auf die Prognose von Informationen abzielen, die sich nicht auf anderem Weg verifizieren lassen.

Der Transparenz zu dienen vermögen auch neue Verfahren zur Authentisierung durch technische Prozesse,³¹ welche die Urheberschaft einer Informationsübermittlung kryptografisch nachweisen. Wenn solche technische Prozesse in nachvollziehbarer Weise die Identifikation der personellen Quellen ermöglichen, lässt sich verhindern, dass Akteure, die ggf. einen negativen Einfluss auf die Dienstleistungserbringung ausüben, in der Intransparenz «verschwinden».

(ii) Aus der Perspektive der *Datensicherheit* sind zwei Schutzstrategien auf der Modellebene denkbar: Einerseits lassen sich die Nutzungsmöglichkeiten allge-

²⁸ Grundlegend zu den Transparenzanforderungen WEBER, 2009, 121 ff.

²⁹ LIPS et al., 2023, 323 ff.; vgl. auch SELMAN et al. 2023, 291.

³⁰ Eingehender dazu SEILER/GRIESINGER, 2023, Rz 7 ff.

³¹ Für eine Problemanalyse zum Finanzmarktbereich vgl. BOŠKIĆ/HEPP, 2023, 214 ff.

mein einschränken, andererseits können Massnahmen zur Unterbindung potentiell schädlicher Auswirkungen auf die Sprachmodelle in Betracht gezogen werden. Zudem sind Massnahmen zur Verringerung des Angriffsrisikos durch Verbesserung der Sensibilisierung und Aufklärung der Nutzenden zu treffen.

Zwischenzeitlich bestehen auch verschiedene komplementäre Ansätze zur Detektion automatisch generierter Texte; durch die Detektion erhalten Nutzende die Fähigkeit, Texte als maschinengeschrieben zu erkennen und somit ggf. ihre Authentizität und die Richtigkeit der enthaltenen Daten anzuzweifeln. Automatisierte Detektionswerkzeuge mit Bezug auf maschinengenerierte Texte nutzen in der Regel statische Eigenschaften der Texte aus und verwenden Parameter eines Modells, um einen Score zu berechnen, der als Indiz für maschinengenerierte Texte dient.

5.2. Konkrete Vorgaben für transparente und sichere Sprachmodelle

Mit Blick auf die Entwicklung transparenter und sicherer juristischer Sprachmodelle können verschiedene Risikovermeidungs- und Risikominderungs-massnahmen relevant sein:³²

Datenqualität bei der Auswahl von Trainingsdaten: Die konkrete und transparente Auswahl der Trainingsdaten ist wichtig, um ein hohes Qualitätsniveau beim juristischen Sprachmodell zu erreichen. Die verwendeten Trainingsdaten sind Grundlage für die späteren Textgenerierungs-Modelle.³³ Zur Qualität gehört auch die Beachtung der rechtlichen Vorgaben; im Vordergrund steht dabei die Berücksichtigung der urheberrechtlichen Rahmenbedingungen.³⁴

Zur Datenqualität gehört weiter die Vermeidung von (versteckten) Diskriminierungen, die sich durch die Verwendung der Algorithmen perpetuieren könnten.³⁵ Zwar lassen sich Diskriminierungen in der Praxis nie ganz ausschliessen, z.B. wenn geographische oder personenbezogene Merkmale verwendet werden; zumindest ist aber zu versuchen, entsprechende negative Effekte zu minimieren.

Privacy Attacks: Angesichts der grossen Menge an Trainingsdaten ist es oft schwierig, durch transparente technische Massnahmen sicherzustellen, dass

³² Bundesamt für Sicherheit in der Informationstechnik, 2023, 15 ff.

³³ Vgl. auch ROSENTHAL, 2023, Rz 26 ff.

³⁴ STRAUB, 2023, Rz 2 ff.; aus US-amerikanischer Sicht vgl. LEUJEUNE, 2023, 142 ff.

³⁵ WEBER/HENSELER, 2020, 31 f.

die Personendaten lediglich für die datenschutzrechtlich unbedenklichen eingeschränkten Zwecke verwendet werden.³⁶ Möglichkeiten zur Verminderung der Anfälligkeit für Privacy Attacks sind beispielsweise³⁷ die manuelle Auswahl oder automatische Filterung bzw. Anonymisierung von Daten, die Entfernung von Dopplungen aus den Trainingsdaten zwecks Herabsetzung der Wahrscheinlichkeit einer möglichen Rekonstruktion,³⁸ die Anwendung von Mechanismen, welche die sog. Differential Privacy garantieren,³⁹ die Einschränkung des Zugriffs auf das Modell (enge Zahl an Nutzenden) sowie die Vornahme von zusätzlichen Trainings bzw. eines besonderen Training für sensible Daten.⁴⁰ *Adversarial Attacks*: Angriffe können darauf abzielen, Texte absichtlich nur leicht zu verändern bzw. intransparent zu gestalten, um die Wahrnehmung der Veränderung unwahrscheinlich zu machen. Als besonders anfällig dafür erweisen sich Klassifikationen. Gegenmassnahmen sind das Training des Modells mit realen oder möglichst realistischen Daten, die Vorverarbeitungen des möglicherweise adversarialen Textes,⁴¹ automatische Rechtschreibkorrekturen, Modellverbesserungen durch Trainings mit manipulierten/veränderten Texten, die Einbindung einer externen Wissensbasis oder das Clustering von Word-Embeddings.

Poisoning Attacks: Sofern die Daten aus öffentlichen Quellen stammen, sind sie einer Vielzahl von Angriffsmöglichkeiten ausgesetzt; Institutionen weisen oft einen offenen Zugang und einen sicherheitstechnisch nicht immer ausreichenden Schutz auf.⁴² Möglichkeiten zur Verminderung der Anfälligkeit für Poisoning-Angriffe sind die Verwendung vertrauenswürdiger Quellen als Trainingsdaten, der Einsatz von geschultem und vertrauenswürdigen Personal, die intensive Analyse von Bewertungen (Voreintritt von Rückwirkungen auf das Modell) und die Beschränkung der Auswirkungen des Einsatzes auf ein kontrollierbares Feld. Alle getroffenen Massnahmen sollten weiter zu einem hohen Grad an Transparenz führen;⁴³ damit lässt sich eine Annäherung an die angesprochene Kohärenz zwischen Transparenz und Datensicherheit erreichen.

³⁶ Zur Herstellung eines datenschutzrechtlich sicheren Modells vgl. ROSENTHAL, 2023, Rz 10 ff.

³⁷ Für einen allgemeinen Überblick vgl. Bundesamt für Sicherheit in der Informationstechnik, 2023, 16.

³⁸ CARLINI et al., 2021, 1 ff.

³⁹ DESFONTAINES/PEJÓ, 2020, *passim*.

⁴⁰ Aus informationstechnologischer Sicht dazu STAAB et al., 2023, *passim*.

⁴¹ WANG et al., 2019, *passim*.

⁴² Bundesamt für Sicherheit in der Informationstechnik, 2023, 18.

⁴³ Vgl. dazu auch ROSENTHAL, 2023, Rz 55 ff.

6. Ausblick

Gesamthaft betrachtet steht somit ein beachtliches Arsenal an Vorgaben zur Verfügung, das sich einsetzen lässt, um geeignete Rahmenbedingungen für transparente und sichere juristische Sprachmodelle im Rechtsbereich zu schaffen. Ausgangspunkt muss die transparente Festlegung der Voraussetzungen und Anwendungskriterien für die Nutzung von Sprachmodellen sein. Gestützt darauf hat eine Implementierung der Programme unter Beachtung hoher Datensicherheitsstandards zu erfolgen.

Der Transfer von Wissen auf neuartige algorithmische Anwendungen beim Einsatz von LLM bringt regelmässig gewisse Risiken mit sich. Deren Bewältigung ist indessen möglich, wenn die sachgerechten Schutzvorkehrungen getroffen werden, d.h. ein transparentes Vorgehen ist mit ausreichenden Datensicherheitsgarantien vereinbar. Unausweichlich ist aber im Kontext der juristischen Sprachmodelle eine verstärkte interdisziplinäre Zusammenarbeit; nur wenn Recht und Informationstechnologie in der Theorie und in der Praxis kooperieren, ist die gewünschte Harmonie von Transparenz und Datensicherheit zu erreichen.

7. Literatur

- ALMODOVAR, CRISPIN/SABRINA, FARIZA/KARIMI, SARVNAZ/AZAD, SALAHUDDIN, Can Language Models Help in System Security? Investigating Log Anomaly Detection using BERT, Proceedings of the 20th Annual Workshop of the Australasian Language Technology Association, Adelaide, December 2022, <https://aclanthology.org/2022.alta-1.19/> (accessed on 17 October 2023).
- BARENKAMP, MARCO, Einflussnahme grosser Sprachmodelle auf die moderne Arbeitswelt, Informatik Spektrum, 22. August 2023, <https://doi.org/10.1007/s00287-023-01546-8> (accessed on 17 October 2023).
- BOŠKIĆ, MAGDALENA/HEPP, SEBASTIAN, zkKYC in Decentralized Finance (DeFi), GesKR 2023, 209–225.
- BUBECK, SÉBASTIEN/CHANDRASEKARAN, VARUN/ELGAN, RONEN/GEHRKE, JOHANNES/HORVITZ, ERIC/KAMAR, ECE/LEE, PETER/LEE, YIN TAT/LI, YUANZHI/LUNDBERG, SCOTT/NORI, HARSHA/PALANGI, HAMID/RIBEIRO, MARCO TULLIO/ZHAN, YI, Sparks of Artificial General Intelligence: Early Experiments with GPT-4, April 2023, <https://arxiv.org/abs/2303.12712> (accessed on 17 October 2023).
- Bundesamt für Sicherheit in der Informationstechnik, Grosse KI-Sprachmodelle, Chancen und Risiken für Industrie und Behörden, Bonn 2021.

- CARLINI, NICHOLAS/TRAMÈR, FLORIAN/WALLACE, ERIC/JAGIELSKI, MATTHEW/HERBERT-VOSS, ARIEL/LEE, KATHERINE/ROBERTS, ADAM/BROWN, TOM/SONG, DAWN/ERLINGSSON, ÚLFAR/OPREA, ALINA/RAFFEL, COLIN, Extracting Training Data from Large Language Models, 30th Usenix Security Symposium, 2021, <https://www.usenix.org/conference/usenixsecurity21/presentation/carlini-extracting> (accessed on 17 October 2023).
- DANILEVSKY, MARINA/QIAN, KUN/AHARONOV, RANIT/KATSIS, YANNIS/KAWAS, BAN/SEN, PRITHVIRAJ, A Survey of the State of Explainable AI for Natural Language Processing, December 2020, <https://aclanthology.org/2020.aacl-main.46/> (accessed on 17 October 2023).
- DESFONTAINES, DAMIEN/PEJÓ, BALÁSZ, Differential Privacies, Proceedings on Privacy Enhancing Technologies Symposium, December 2020, <https://arxiv.org/abs/1906.01337> (accessed on 17 October 2023).
- GLESS, SABINE, Künstliche Intelligenz in der Gerichtsbarkeit, Zeitschrift für Schweizerisches Recht 2023 I 429–462.
- GRUETZEMACHER, ROSS, The Power of Natural Language Processing, Harvard Business Review, April 2022, <https://hbr.org/2022/04/the-power-of-natural-language-processing> (accessed on 17 October 2023).
- LEUJEUNE, MATHIAS, AI Systems and their Output under U.S. Copyright Law, CRI 5/2023, 141–148.
- LIPS, MARKUS/ŞAHİN, FATİH/ROSENAUER, PHILIPP, Effizienzsteigerungsmöglichkeiten durch generative AI für Rechtsdienstleistungen, Anwaltsrevue 2023, 323–327.
- ROSENTHAL, DAVID, Datenschutz beim Einsatz generativer künstlicher Intelligenz, Weblaw Jusletter, 6. November 2023.
- SEILER, DANIEL W./GRIESINGER, MARCEL, Spannungsfeld Künstliche Intelligenz (KI) und Datenschutzrecht, Weblaw Jusletter, 25. September 2023.
- SELMAN, SINE/BURRICHTER, ANNIKA/HUBLI, PASCAL, Die Verwendung von KI und ChatGPT im Anwaltsberuf, Anwaltsrevue 2023, 289–294.
- STAAB, ROBIN/VERO, MARK/BALUNOVIC, MISLAV/VECHEV, MARTIN, Beyond Memorization: Violating Privacy via Interference with Large Language Models, October 2023, <https://arxiv.org/pdf/2310.07298v1.pdf> (accessed on 17 October 2023).
- STRAUB, WOLFGANG, Immaterialgüterrechtlicher Schutz mit KI geschaffener Werke und Erfindungen, Weblaw Jusletter, 7. August 2023.
- VOKINGER, KERSTIN NOËLLE/SCHNEIDER, DAVID/LOCHER, LUCA/HERRLE, CASCAL/MÜHLEMATTER, URS JAKOB, Analyse von Bundesgerichtsurteilen mit ChatGPT-3.5, Weblaw Jusletter, 7. August 2023.
- WANG, WENQI/TANG, BENXIAO/WANG, RUN/WANG, LINA/YE, AOSHUANG, A Survey on Adversarial Attacks and Differences in Text, February 2019, https://www.researchgate.net/publication/331246520_A_survey_on_Adversarial_Attacks_and_Defenses_in_Text (accessed on 17 October 2023).
- WEBER, ROLF H., The Disclosure Dream – Towards a New Transparency Concept in EU Consumer Law, EuCML 2023, 67–70 (WEBER, 2023).
- WEBER, ROLF H., Künstliche Intelligenz: Regulatorische Überlegungen zum «Wie» und «Was», EuZ 01/2022, B1–B18 (WEBER, 2022) (Text 2, S. 17 ff.).
- WEBER, ROLF H., Shaping Internet Governance: Regulatory Challenges, Zürich 2009 (WEBER, 2009).

- WEBER, ROLF H./HENSELER, SIMON, Regulierung von Algorithmen in der EU und in der Schweiz, EuZ 2020, 28–42 (WEBER/HENSELER, 2020) ([Text 7, S. 103 ff.](#))
- WEIDINGER, LAURA/UESATO, JONATHAN/RAUH, MARIBETH/GRIFFIN, CONOR/HUANG, PO-SEN/MELLOR, JOHN/GLAESE, AMELIA/CHENG, MYRA/BALLE, BORJA/KASIRZADEH, ATOOSA/BILES, COURTNEY/BROWN, SASHA/KENTON, ZAC/HAWKINS, WILL/STEPLETON, TOM/BIRHANE, ABEBA/HENDRICKS, LISA ANNE/RIMELL, LAURA/ISAAC, WILLIAM/HAAS, JULIA/LEGASSICK, SEAN/IRVING, GEOFFREY/GABRIEL, IASON, Taxonomy of Risks posed by Language Models, June 2022, <https://doi.org/10.1145/3531146.3533088> (accessed on 17 October 2023).
- ZHANG, YUE/LI, YAFU/CUI, LEYANG/CAI, DENG/LIU, LEMAO/FU, TINGCHEN/HUANG, XINTING/ZHAO, ENBO/ZHANG, YU/CHEN, YULONG/WANG, LONGYUE/LUU, ANH TUAN/BI, WEI/SHI, FRED/SHI, SHUMING, Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models, September 2023, <https://arxiv.org/pdf/2309.01219v1.pdf> (accessed on 17 October 2023).
- ZELLERS, ROWAN/HOLTZMAN, ARI/RASHKIN, HANNAH/BISK, YONATAN/FARHADI, ALI/ROESNER, FRANZISKA/CHOI, YEJIN, Defending against Neural Fake News, Advances in Neural Information Processing Systems, December 2020, <https://arxiv.org/abs/1905.12616> (accessed on 17 October 2023).

Human-Machine Cooperation through AI Agents?

Erstveröffentlichung in: Schweighofer et al. (Hrsg.), *Mensch-Maschine-Kooperation im Cyberspace* (IRIS 2026), Bern 2026, 19–27

AI agents play an increasingly important role in contractual relations. Such kind of digital assistants must be designed and programmed in a way that the traditional principal-agent problems can be avoided by optimal contracts' parameters. Recent international normative frameworks proposing rules for an appropriate AI agents' regime lay the foundation for a future human-machine cooperation, but a further refinement of the rules securing specific protection mechanisms is needed.

Table of Contents

1.	Notion of AI Agents	163
2.	Principal-Agent Situation and its Challenges	166
2.1.	Traditional Principal-Agent Problem	166
2.2.	Design of Optimal Contracts	167
3.	Legal Framework for Principal-Agent Relations	169
3.1.	ELI Model Law Principles	170
3.2.	Contents of the ELI Model Rules	171
3.3.	Comparison with the UNCITRAL Model Law	173
4.	Conclusion: Towards the Design of Optimal Contracts and of Adequate Agency Rules	174
5.	Bibliography	175

1. Notion of AI Agents

Artificial intelligence agents (AI agents), partly also called intelligent agents, are software or hardware components that are capable of autonomously performing tasks, processing data, making decisions, and interacting with other agents, humans or the environment. The envisaged designs intend to allow the creation and execution of plans that maximize the expected value

of the AI functions upon completion.¹ The AI agents are based on advanced algorithms and machine learning, making them superior to simple large language models such as ChatGPT. Functionally, AI agents can constantly learn and change conditions to improve their quality by tailoring the “results” to the needs of the concerned users.

The exact mode of operation of AI agents depends on the underlying algorithms and the specific application or their rationale, respectively. These capabilities are based on the possibility of using various tools such as databases, specific programs, etc. The AI agent’s function can theoretically be described along the applied incentive-intensity parameters influencing the behaviour of principals, such as equal treatment, risk aversion and effects’ observability factors.² Apart from legal rules, the empirical evidence plays an important role in assessing these factors.

In the meantime, academics developed notions of different classes of AI agents: (i) RUSSELL and NORVIG³ distinguish between (i) simple reflex agents, (ii) model-based reflex agents, (iii) goal-based agents, (iv) utility-based agents, and (v) learning agents. The last category, being most relevant for the legal considerations, not only encompasses learning elements, but also performance elements that gather feedback from “critical” reactions. (ii) WOOLRIDGE⁴ categorizes as follows: (i) Logic-based agents, (ii) reactive agents, (iii) belief-desire-intention agents, and (iv) layered architectures, allowing a decision-making across multiple software layers.

The quality of the achieved “results” depends on whether the AI agents are intelligent enough to learn how to apply various tools that can exercise different functions. Insofar, a key element concerns the availability of the relevant information or the absence of information asymmetry, being a crucial issue of contract theory.⁵ Only if sufficient information can be considered and assessed, an appropriate decision-making process is possible. Often, AI agents are also able to use neural networks to analyse images and recognize objects.

¹ BRINGSJORD/GOVINDARAJULU (2018), no. 2,

² To the theoretical framework see below Ch. [2.1](#).

³ RUSSELL/NORVIG (2003), pp. 46–52.

⁴ WOOLRIDGE (2013), Ch. 1.

⁵ See below Ch. [2.2](#).

AI agents do have technical or other relations to a variety of environments which provide input to the system. The respective relations can be designed as follows:⁶

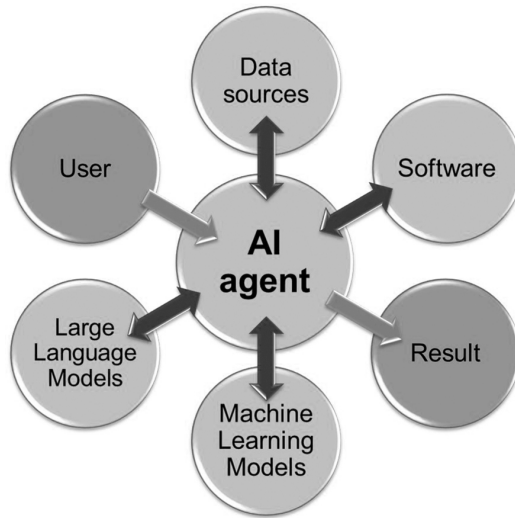


Fig. 1: How AI agents work

Training and quality of data have a crucial impact on the effectiveness of AI agents; as experience has shown, manipulation and data misuse must be avoided in this context. Therefore, the potential of transforming many aspects of daily life is big and the available technologies accelerate the respective ongoing efforts. Not surprisingly, however, the fast development of the AI agents' environment causes legal challenges having a completely new character or going back to already known tensions. As an example, the interrelations between a human and an AI agent could also become confronted with the “traditional” principal-agent challenges; the question to be discussed in more detail hereinafter concerns possible deviations from this crucial phenomenon.

⁶ Own illustration based on VAI LABS (October 2023), Navigating the AI Agent Landscape: Insights into Advancements and Opportunities, available at https://medium.com/@VAI_LABS/navigating-the-ai-agent-landscape-insights-into-advancements-and-opportunities-75c4e67ffc8e (accessed on 26 October 2025).

2. Principal-Agent Situation and its Challenges

2.1. Traditional Principal-Agent Problem

The principal-agent problem (partly also called the agency problem) relates to a potential conflict in interests and/or priorities that follow from the fact that a person or entity (the “agent”) is enabled to act on behalf of another person or entity (the “principal”). If the agent acts based on self-interest, the consequences caused to the principal consist in the so-called “agency costs”.

Almost 100 years ago the principal-agent problem already was discussed regarding the relations between the shareholders and the corporate management (BERLE/MEANS).⁷ More than fifty years ago, the so-called “institutional theory” analyzed the relations between the citizens and the elected government officers.⁸ The respective substantive analyses addressed aspects such as information asymmetry, uncertainty, and risks that influence the structure of corporations and the allocation of power. Two aspects merit particular attention:

(i) A discrepancy between the interests of the principal and of the agent can occur for various reasons. As academics have diagnosed over the last few decades, the principal-agent problem is mainly due to information asymmetries; since the agent very often has more or better information than the principal, decisions taken tend to turn out in favour of the agent. Other reasons for agency costs are given if moral hazard or adverse selection plays a role in the relation.

When a principal delegates tasks to an agent, normally the agent’s actions and, especially, the deviations to the principal’s best interest are not fully observable; therefore, the relationship is prone to agency problems; depending on the different goals of the two parties the agency costs will rise. With reference to contract theory, the principal-agent problem is exposed to the challenge that contracts without any restrictions often neglect the question how the agent’s reward is determined and cause the risk that the interests of the principal and the agent are misaligned.⁹ Furthermore, the design of the

⁷ BERLE/MEANS (1931–2009).

⁸ For an overview see MITNICK, BARRY M., *Origin of the Theory of Agency: An Account By One of the Theory’s Originators*, 10 October 2007, revised 2 June 2019, available at <https://dx.doi.org/10.2139/ssrn.1020378>.

⁹ The most-influential academic work discussing many aspects of the principal-agent issue stems from JENSEN/MECKLING (1976), pp. 305 et seq.

incentive structure can also impact the aspect of how detailed the assignment description was elaborated for the agent and how comprehensive the agent must report to the principal.

(ii) In economic theory, particularly in contract theory, two lines of modelling the principal-agent challenges exist, namely the adverse selection model and the moral hazard model. In the first case, the agent does have specific information, not available to the principal, prior to the conclusion of the contract; in the second case, the agent gets to know the preferential information after the conclusion of the contract.

To avoid moral hazard and adverse selection, contract theory focuses on designing contracts including incentive schemes that align the principal's and agent's interests to mitigate these problems.¹⁰ Contract theory takes the findings of the principal-agent theory into account and addresses the design of contracts; thereby, the focus lies on resolutions to answer the problem of asymmetric information. Thereby, attention must be paid to the sources and reasons causing information imbalances.¹¹ However, all methods coping with the occurring negative effects and the resulting agency costs in turn also have their "price". Therefore, an optimum balance between the costs of drafting the contract and the agency costs must be found. This calculation should consider not only the costs incurred during contract negotiations and drafting, but also potential follow-up costs from contractual disputes up to and including a dispute resolution.

In addition to the agency costs' issue that can have a substantial impact on the proper management of contractual relations, trust relationships equally do have a relevance in the principal-agent context.¹² Usual situations concern the relationship between the advice seeking person and the expert in the field, for example the attorney-at-law, the medical doctor, the IT specialist, etc. Trust relationships are of particular importance if the agent unfolds the activities based on automated algorithmic programs.

2.2. Design of Optimal Contracts

In the meantime, new technological developments are impacting the traditional principal-agent understanding even if technology as such does not solve

¹⁰ For more details see HART/HOLMSTRÖM (1987), pp. 71 et seq.

¹¹ To the details of possible contract designs from a principal-agent perspective see MILGROM/ROBERT (1992).

¹² To the trust relationship see BERG/DICKHAUT/MCCABE (1995), pp. 122 et seq.

some inherent challenges of contractual relations, for example the problem of information asymmetry and the trust issue. In 1963, ARROW¹³ introduced a contract theory being a useful instrument for modelling the behaviour of economic agents within a contractual relationship in case of asymmetric information due to “hidden information” or “hidden action”. Economists (apart from ARROW also AKERLOF¹⁴) have long argued that information asymmetry on the characteristics of the product (adverse selection) or on the behaviour of the provider (moral hazard) threatens the equilibrium of the contractual obligations.¹⁵ Only if such an equilibrium can be achieved, the respective outcomes also facilitate the overcoming of the principal-agent problem.¹⁶

The construction of consent between the parties is challenging if automatic signature procedures in the context of a weak technological understanding and information asymmetry prevail. The assumption of hypothetical opinions guiding the involved parties is mainly challenging in the case of incomplete smart contracts. If contractual relations are incomplete, the performance and enforceability entail substantial frictional costs. Therefore, a need for increased harmonization and predictability in this context is given.¹⁷

An optimal smart contract should incorporate product or service information and the utility functions of both, provider and user; an algorithm, based on the information produced, must then solve the generalized principal-agent model between the provider and the user. The development of such an algorithm is feasible, as recently published papers have shown;¹⁸ but in the design of the algorithm, it is important to realize an exchange of value that achieves a fair and balanced transfer in each direction since the appropriate value is the key to every human economic activity and collaboration.

Lately, in respect of information asymmetry, some economists expressed the opinion that contract theory provides tools that allow an appropriate “signalling and screening to limit adverse selection and design of optimal incentive mechanisms to mitigate moral hazard”.¹⁹ The theoretical approach is

¹³ ARROW (1963), pp. 941 et seq.

¹⁴ AKERLOF (1970), pp. 488 et seq.

¹⁵ WEBER (2026), Ch. 4.1.

¹⁶ OARDA (2024), p. 2.

¹⁷ WEBER (2026), Ch. 4.1.

¹⁸ OARDA (2024), pp. 1 et seq.

¹⁹ OARDA (2024), p. 2.

also contributing to the development of solutions by reducing the complexity of structure models and mitigating the principal-agent problem.²⁰

These (economically driven) contract theory tools are relevant in the digital environment of AI agents' use. The legal framework must incorporate those rules which have the potential to restrict the negative elements of the principal-agent relations (for example information asymmetry, moral hazard, adverse selection). Hereinafter, two already existing international instruments dealing with principal-agent relations are discussed in more detail.

3. Legal Framework for Principal-Agent Relations

Apart from crucial ethical concerns,²¹ the use of AI agent in the human-machine cooperation causes manifold legal challenges. As in all other AI contexts, data protection and privacy merit particular attention.²² Further legal issues are (i) AI safety, security, vulnerability, and risk of cybercrime,²³ (ii) weakened human oversight and algorithmic discrimination,²⁴ or (iii) explainability of IA agents' decisions and user manipulation.²⁵ These legal challenges will not be discussed hereinafter since the contribution is focused on the "cooperation" theme, i.e. the contractual relations between humans and AI agents (machines). As main example, the "Guiding Principles and Model Rules on Digital Assistants for Consumer Contracts" of the European Law Institute (ELI) of 2025, governing consumer-business contracts, will be discussed.²⁶ In addition, reference is also made to the UNCITRAL Model Law

²⁰ See also HOLMSTRÖM (1979), pp. 74 et seq.; GROSSMAN/HART (1983), pp. 7 et seq.

²¹ WEBER (2025a), pp. 27 et seq.

²² WEBER (2025b), pp. 749 et seq.

²³ To these topics see ZITTRAIN, JONATHAN L. (2024), We Need to Control AI Agents Now, Atlantic, July 2024, <https://www.theatlantic.com/technology/archive/2024/7/ai-agents-safety-risks/678864/> (accessed on October 26, 2025).

²⁴ To these topics see ZHUANG, MIKE/DESCRIJVER, ELIANE/RAMSEY, RICHARD/TUREL, OFIR (2025) Comparing discriminatory behavior against AI and humans, 15 Nature, Scientific Reports 2025, <https://doi.org/10.1038/s41598-025-94631-9> (accessed on October 26, 2025).

²⁵ To these topics see CRAWFORD, KATE (2024), AI Agents Will Be Manipulation Engines, <https://www.wired.com/story/ai-agents-personal-assistants-manipulation-engines/> (accessed on October 26, 2025).

²⁶ ELI Guiding Principles and Model Rules on Digital Assistants for Consumer Contracts (2025), <https://www.europeanlawinstitute.eu/projects-instruments/instruments/eli-guiding-principles-and-model-rules-on-digital-assistants-for-consumer-contracts/> (accessed on October 26, 2025).

on Automated Contracting of 2024, setting rules for business-business contracts.²⁷

3.1. ELI Model Law Principles

Most legal systems do know rules about principal-agent relations; relevant regulations can be found in the agency provisions, the mandate law and the shareholders-directors' rules. Agency provisions lead to a tripartite contractual "framework" in case of interventions by AI agents; the following figure as developed by the ELI in the "Model Rules" reflects the tripartite situation.²⁸

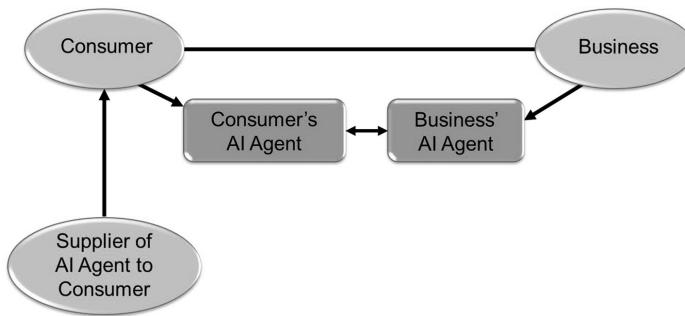


Fig. 2: ELI Model Rules

The ELI Model Rules are based on eight Guiding Principles outlining basic objectives for the design of contracts being concluded and executed by digital (AI) agents; these Principles must be observed and complied with.²⁹

²⁷ UNCITRAL Model Law on Automated Contracting (2024), <https://uncitral.un.org/en/mlac> (accessed on October 26, 2025).

²⁸ ELI Guiding Principles (supra note 26), Introduction, p. 3.

²⁹ ELI Guiding Principles (supra note 26), Introduction, p. 4.

Principle 1:	The actions of a digital assistant are attributed to the person using it.
Principle 2:	Consumer law extends to contracts concluded through a digital assistant.
Principle 3:	The obligation to provide pre-contractual information remains relevant.
Principle 4:	There should be no barriers to the use of digital assistants, and transactions arranged through a digital assistant should be legally effective.
Principle 5:	The use of a digital assistant must be disclosed by both parties.
Principle 6:	A digital assistant must be protected from manipulation.
Principle 7:	A consumer must remain in control of a digital assistant through the ability to set, review and amend all the parameters used by a digital assistant; through the ability to prevent the conclusion of a contract through the digital assistant; and by deactivating the digital assistant.
Principle 8:	Conflicts of interest arising from the use of a digital assistant must be disclosed.

These Guiding Principles are “converted” into 25 normative provisions constituting the Model Rules; the respective regulations will be analysed hereinafter. The ELI Model Rules use the term “digital assistant” defining it as an application linked to algorithmic systems; since algorithmic contracts are equally covered, the respective provisions can be similarly applied in case of AI agents.

3.2. Contents of the ELI Model Rules

The main principle of the legal “construction” of a principal-agent relation consists in the attribution of the AI agent to the person using it if certain conditions are met. Any kind of limitation to this attribution must be explicitly stated. If appropriately applied, this normative attribution leads to an adequate risk allocation reflecting the parties’ intentions.

A digital assistant can be defined as an application utilising algorithmic systems that, based on pre-set parameters, user inputs and data obtained from a range of sources, is designed to take actions to accomplish a set of predefined objectives for contractual relations.³⁰ The right to use a digital assistant for contractual relations is not limited as such; moreover, the business contract party is not entitled to impose a private rule on the consumer which directly or indirectly waives or restricts this right.³¹ Furthermore, the

³⁰ Art. 2 (2).

³¹ Art. 3 (1) and (2).

online interfaces must avoid that a design is chosen which excludes the users digital assistants.

A digital assistant intended to be used for contractual relations must allow the selection and modification, at any time, of a range of parameters and their relative importance by the digital assistant in performing its functions. As a minimum, such parameters should include a) price range, price criteria and other pricing elements, b) types and characteristics of products or services, c) identity or characteristics of the businesses entering the contractual relations, as well as d) for example duration, delivery, location and sustainability criteria.³² The ELI Model Rules also contain provisions related (i) to the functionality of preventing the conclusion of a contract (art. 6), (ii) to a deactivation of the digital assistant temporarily or permanently (art. 7) and (iii) to the disclosure requirements (art. 8). Finally, a digital assistant must be designed and operate in a way that does not manipulate nor otherwise materially distort or impair the possibility to make free and informed choices or decisions (art. 9).

The ELI Model Rules put much emphasis on transparency in respect of contracts for the supply of digital assistants: information, including contract terms, must be provided in a plain and intelligible manner, encompassing nine listed groups of different elements.³³ Main parameters are the characteristics of the digital assistant, its adaptive capability and any default settings. For obvious reasons, potential conflicts of interests must be disclosed before the contract is concluded and the information must be given separately from other information required to be submitted.³⁴ A digital assistant is to be designed in a way that it can comply with the operational requirements and functionalities as laid down in the Model Rules (art. 14). However, the business party does not have any obligation to monitor the transactions that the contracting parties are executing; only a duty to warn is foreseen if the business party has become aware of failures in the design of the digital assistant (art. 15).

A special chapter regulates the algorithmic contracts. The main principle concerns their legal recognition: An algorithmic contract shall not be denied validity or enforceability solely because a digital assistant was used (art. 17). Apart from “traditional” contractual disclosures (art. 20), the implementation

³² Art. 5 (1).

³³ Art. 11 (1) and (2).

³⁴ Art. 12 (1) and (2).

of a digital assistant for contractual tasks must be specifically made transparent to the other party in a clear and intelligible manner at the beginning of the contractual interaction (art. 19). Furthermore, compliance with obligations to provide information is equally applicable in case of algorithmic contracts (art. 21). A person who uses a digital assistant for contractual relations is bound by the actions taken by the digital assistant; specific attribution criteria try to delineate the respective responsibilities of the two contract parties and to protect the consumer as the weaker contract party.³⁵

3.3. Comparison with the UNCITRAL Model Law

The UNCITRAL Model Law governs the business-contract-relations, therefore, special rules protecting the weaker contract party (consumer) are not needed. Nevertheless, many provisions do have very similar contents. Art. 5 of the UNCITRAL Model Law states that a contract formed by use of an automated system is valid and enforceable even if no natural person reviewed or intervened in any action carried out in connection with the formation of the contract. The legal recognition of such contracts, in particular smart contracts, extends to the performance even if done automatically through the execution of a computer code.

Art. 6 of the UNCITRAL Model Law reiterates this validity principle in respect of further data exchanges: A contract is valid and enforceable even if its contents is contained in data messages in the form of computer code. In other words, computer coding that translates the terms into machine-readable instructions that can be executed by automated systems do not restrict the validity of the contract.³⁶ Furthermore, art. 6 includes dynamic information referring to information from a data source that changes periodically or continuously (such as information on data price or on the location of an object); the source may be external to a system (e.g. an oracle) or internal.

A key provision is art. 7 of the UNCITRAL Model Law: “as between the parties to a contract, an action carried out by an automated system is attributed in accordance with a procedure agreed to by the parties”. Consequently, the attribution can identify a (natural or legal) person bearing the responsibility and liability for the performance of the contract. The attribution does not cause a link to an automated system, but to a person.³⁷ Depending on the

³⁵ Art. 22 (1) and (3).

³⁶ UNCITRAL Model Law, Explanatory Report (supra note 28), p. 24.

³⁷ UNCITRAL Model Law, Explanatory Report (supra note 28), p. 27.

circumstances, a range of factors may be relevant in identifying the responsible person, for example (i) the person deploying the automated system, (ii) the control exercise over the operational parameters of the system and the action, (iii) the benefit or value derived from the action, (iv) the nature and purpose of the contract, and (v) trade usages and the practices established between the parties.³⁸

4. Conclusion: Towards the Design of Optimal Contracts and of Adequate Agency Rules

Academics have developed technological settings and theories that help to overcome the traditional principal-agent challenges and contribute to an appropriate design of optimal contracts. The description of the ELI Guiding Principles and the UNCITRAL Model Law has also shown that adequate legal rules for the use of AI agents can be introduced in the normative framework by stating the following theorems: (i) Contracts entered and performed by an AI are valid and enforceable, similarly to traditional contracts. (ii) The actions executed by an AI agent must be attributed to a natural or legal persons. The correctness of this attribution is crucial for the proper business conduct of an AI agent. In this respect, more academic research on an interdisciplinary level is needed.

Contract law is deeply intertwined not only with economic but also with societal aspects. For both reasons, the legal framework must contain guardrails for combatting “unexpected actions” of AI agents as well as technological errors that are not covered by the pre-agreed parameters. If the AI agent is able to significantly deviate from the consented contract program or exceed the parties’ expectations, normative barriers must come into place; the general objectives of the ELI Guiding Principles and the UNCITRAL Model Law do contain some “reservations”, however, further specifications appear to be desirable. For the time being, the existing legal instruments, particularly on the national level, leave certain gaps, and the preservation of the involved parties’ autonomy is not sufficiently secured; therefore, a call for action must be expressed in this context.

If the appropriate protection measures are not in place, the described principal-agent problem will accrue. To avoid the respective challenges, digital signalling and screening should limit negative outcomes due to the deception

³⁸ UNCITRAL Model Law, Explanatory Report (supra note 28), p. 29.

of trust in the AI agent. Consequently, legislators are called to implement safeguards protecting fair contractual relations, avoiding all kinds of manipulation and combatting future risks of innovative agentic solutions. Furthermore, the development of new product and service safety rules for AI agents is becoming a regulatory issue; apart from the general AI liability provisions, specific rules governing the contractual relations between individuals/businesses and AI agents will become imperative.

5. Bibliography

- AKERLOF, GEORGE A, The Market for “Lemons”: Quality Uncertainty and the Market Mechanism, 84 Quart. J. of Econ. 1970, pp. 488–500.
- ARROW, KENNETH J., Uncertainty and the Welfare Economics of Medical Care, 53 Am. Econ. Rev. 1963, pp. 941–973.
- BERG, JOYCE/DICKHAUT, JOHN/MCCABE, KEVIN, Trust, Reciprocity, and Social History, 10(1) Games and Economic Behavior 1995, pp. 122–142.
- BERLE, ADOLF A./MEANS, GARDINER C, The modern corporation and private property, New Brunswick/London, 1931, 10th print 2009.
- BRINGSJORD, SELMER/GOVINDARAJULU, NAVEEN SUNDAR, Artificial Intelligence, 2018, <https://plato.stanford.edu/entries/artificial-intelligence/> (accessed on October 26, 2025).
- GROSSMAN, SANFORD/HART, OLIVER D, An Analysis of the Principal-Agent Problem, 51(1) Econometrica, 1983, pp. 7–45.
- HART, OLIVER/HOLMSTRÖM, BENGT, The theory of contracts. In: Bewley (ed.), Advances in Economics and Econometrics, Cambridge Univ. Press, Cambridge 1987, pp. 71–155.
- HOLMSTRÖM, BENGT, Moral Hazard and Observability, 10(1) Bell J. of Econ., 1979, pp. 74–91.
- JENSEN WILLIAM H./MECKLING WILLIAM H., Theory of the firm: Managerial behavior, agency costs and ownership structure, 3 (4) J. of Fin. Econ. 1976, pp. 305–360.
- MILGROM, PAUL/ROBERT, JOHN, Economics, Organization & Management, Prentice Hall, New York 1992.
- OARDA, COSTIN, Optimal DAO: An Optimal Smart Contracts based Ecosystem Design for a Systematic Solution to the Principal-Agent Problem, 2024, https://optimalcontracts.ch/wp-content/uploads/2024/07/Optimal_DAO-DAWO24-20240611.pdf (accessed on October 26, 2025).
- RUSSELL, STUART J./NORVIG, PETER, Artificial Intelligence: A Modern Approach, 1st US ed. 2003, 4th US ed., Harlow/UK 2022.
- WEBER, ROLF H., Does Centralization Lead to a New Nature of Organizations. In: Mathis/Tor (Eds.), Law and Economics of Organizations, Springer, Cham 2016, *forthcoming* (cit. WEBER, 2026)
- WEBER, ROLF H., Rule of Law and Ethics-based Equality in the Age of Artificial Intelligence. In: Schweighofer et al. (Eds.), Human-Centred AI for Good – AI, Ethics & Law, Bern 2025, pp. 23–30. (cit. WEBER, 2025a).

WEBER, ROLF H., Datenschutz-Compliance beim Einsatz von KI-Systemen. In: Janisch et al. (Hrsg.), Am Glasfaserplus, Festschrift für Dietmar Jahnel, Salzburg 2025, pp. 749–762 (cit. WEBER, 2025b).

WOOLRIDGE, MICHAEL, Intelligent Agents, in: Weiss (ed.), Multiagent Systems, 2nd ed. Cambridge MA 2013, Ch. 1.

Herausforderungen beim Einsatz künstlicher Intelligenz in der Justiz

Erstveröffentlichung in: Fagnoli (Hrsg.), *Automatisierung von juristischen Entscheidungen*, Milano 2024, 43–62

Inhaltsübersicht

1.	Einleitung	177
2.	KI-Programme als Arbeitsinstrument	179
2.1.	Einsatzmöglichkeiten	179
2.2.	Herausforderungen	179
3.	Transparenz und Datensicherheit als wichtige Rahmenbedingungen	181
3.1.	Transparenz bzw. Nachvollziehbarkeit	181
3.2.	Datenschutz und Datensicherheit	182
3.3.	Koordination der Schutzziele	184
4.	Einsatz von KI-Systemen in der Justiz	186
4.1.	Digitalisierung und IT-basierte KI-Systeme in der Rechtsordnung	186
4.2.	Einsatzfelder in einzelnen Verfahrensstadien der Justiz	188
4.2.1.	Feststellung des faktischen Sachverhalts	188
4.2.2.	Auffinden von Rechtssätzen	189
4.2.3.	Anwendung von Rechtsnormen	190
4.3.	Vorteile der Automatisierung	191
4.4.	Kontrolle von Risiken bei der Automatisierung	193
4.4.1.	Maschinen statt Menschen?	193
4.4.2.	Einzelne Risiken	193
4.5.	Gegenseitige Abwägung von Vorteilen und Risiken	196
5.	Ausblick	197

1. Einleitung

Der weite Begriff «Künstliche Intelligenz» (KI), der z.B. Erscheinungen wie automatisierte Entscheidungen, algorithmische Programmierungen und auch juristische Sprachmodelle bzw. allgemein Large Language Models (LLM) umfasst, hat in den letzten Monaten erheblich an Bedeutung gewonnen. Algorithmen und LLM

sind in der Lage, die natürliche Sprache durch KI zu verarbeiten und gestützt darauf sachbezogene Texte zu entwickeln¹.

Die technischen Innovationen bzw. gar «Umwälzungen» führen aber auch zu neuen rechtlichen Herausforderungen, die rund um den Globus die Gesetzgeber haben aktiv werden lassen. Vorreiterin ist die Europäische Union mit dem Artificial Intelligence Act (AIA) gewesen²; der Europarat hat anfangs Mai 2024 eine Artificial Intelligence Framework Convention verabschiedet³; die Schweiz will bis Ende 2024 die Grundlagen für die normative «Bewältigung» der KI bereitstellen⁴. Wichtige Themen im KI-Bereich sind der vorliegend nicht zu vertiefende Urheberrechtsschutz⁵, die gegenüber den Betroffenen geschuldete Gewährleistung der Transparenz bzw. Nachvollziehbarkeit von KI-Vorgängen, die Vermeidung des Eintritts von Risikofaktoren und (angesichts der grossen Datensammlungen) die Aufrechterhaltung der Datensicherheit⁶.

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

¹ Vgl. R. WEBER, *Juristische Sprachmodelle zwischen Transparenz und Datensicherheit*, in *Sprachmodelle: Juristische Papageien oder mehr?* (Hrsg. E. SCHWEIGHOFER et al.), Bern, 2024, S. 19. ([Text 9, S. 149 ff.](#))

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

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

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

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

⁶ WEBER, zit. (Fn. 1), S. 21 ff.

⁷ Im Einzelnen dazu hinten [Ziff. 4.](#)

2. KI-Programme als Arbeitsinstrument

2.1. Einsatzmöglichkeiten

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

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

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

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

2.2. Herausforderungen

Der Einsatz von KI-Programmen in der Justiz hat jedoch unter Beachtung verschiedener Herausforderungen zu erfolgen, um den nicht zu unterschätzenden Risiken zu begegnen. Selbst wenn ein sprachlich fehlerfreier und oft auch inhaltlich überzeugender Text generiert wird, erweist sich nämlich ein unbegrenztes Vertrauen in die Aussagen von KI-Programmen als problematisch,

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

⁹ Vgl. hinten [Ziff. 4.3](#) und S. GLESS, *Künstliche Intelligenz in der Gerichtsbarkeit*, in ZSR 2023 I, S. 429, 445 f.

weil bei den entsprechenden automatisierten Abläufen nicht ein menschenähnliches Leistungsvermögen vorausgesetzt werden darf¹⁰. In Betracht fallen folgende Herausforderungen¹¹:

- *Fehlende Vertraulichkeit der eingegebenen Daten*: Bei der Nutzung von externen Schnittstellen fließen alle Eingaben zunächst an den Betreiber des Modells; sensible und vertrauliche Informationen, die – über das allgemeine Datenschutzrecht¹² hinaus – nicht zuletzt im Rahmen von Berufs- und Amtsgeheimnissen in den Verfahren der Justiz¹³ oft auftreten, bedürfen deshalb der Verknüpfung mit besonderen Schutzvorkehrungen.
- *Mangelnde Faktizität und Reproduzierbarkeit*: Die Generierung von Text beruht auf stochastischen Korrelationen, d.h. das «Wissen» wird aus (bereits gesehenen) Texten abgeleitet. Bezüge zur realen Welt existieren nicht, weshalb es bei für Menschen verständlichen Sachverhalten dennoch zu nicht korrekten Aussagen kommen kann; dieselbe Eingabe vermag auf Grund des wahrscheinlichkeitsbasierten Ansatzes zu unterschiedlichen Aussagen zu führen.
- *Ungenügende Aktualität*: Bietet das KI-Programm keinen Zugriff auf Live-Internetdaten (d.h. auf Informationen über aktuelle Ereignisse bzw. kürzlich ergangene Entscheide in der Justiz, national und grenzüberschreitend), werden ggf. aufgrund der in der Vergangenheit geprägten Datensammlungen nicht aktuelle Ergebnisse verarbeitet bzw. produziert.
- *Fehlerhafte Reaktion auf spezifische Eingaben*: Fehler vermögen aufzutreten, wenn Eingaben zu verarbeiten sind, die so stark von den Texten in den Trainingsdaten abweichen, dass sie sich nicht mehr korrekt in einem neuen Dokument verarbeiten lassen (im Schrifttum oft «Halluzination»¹⁴ als Ausdruck der Sinnestäuschung genannt); solche Abweichungen können zufällig auftreten oder auch absichtlich durch Interventionen Dritter als bewusster Täuschungsakt entstehen.

¹⁰ Vgl. auch R. WEBER, *Dürfen Menschen über Maschinen entscheiden?*, in *Schweizer Monat*, Februar 2019, S. 72, 73. ([Text 3, S. 37 ff.](#))

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

¹² Dazu hinten [Ziff. 3.2.](#)

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

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

- Anfälligkeit für «versteckte» Dateninfiltrationen mit manipulativer Absicht: Falls ungeprüfte Dokumente Dritter im KI-System vorhanden sind oder Situationen eintreten, in denen Angreifende für Nutzende unbemerkt gewisse Eingaben einzubringen in der Lage sind, leben Risiken auf (z.B. Angriff auf ein Chat-Tool).

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

3. Transparenz und Datensicherheit als wichtige Rahmenbedingungen

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

3.1. Transparenz bzw. Nachvollziehbarkeit

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

Selbst soweit es gesetzliche Transparenzanforderungen gibt, wie z.B. mit Blick auf die automatisierten Entscheidungen gemäss Art. 21 DSGVO bzw. Art. 22 DSGVO, ist deren Wirkung nicht zu überschätzen¹⁸. Empirische Untersuchungen zu den vorhandenen daten- und verbraucherrechtlichen Informationspflichten lassen nämlich daran zweifeln, ob die gesetzlich geforderte

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

¹⁶ Vgl. hinten [Ziff. 4.](#)

¹⁷ WEBER, zit. (Fn. 1), S. 21.

¹⁸ Vgl. dazu hinten [Ziff. 3.2.](#)

Transparenz stets einen Mehrwert für die betroffenen Personen generiert und ihren angedachten Zweck tatsächlich erfüllt¹⁹. Die zur Verfügung gestellten Informationen sind oft nicht ausreichend verständlich und überfordern angesichts der Informationsintensität (Länge, Detaillierungsgrad) die Empfänger/innen²⁰. Diese allgemeine Problematik wird dadurch verschärft, dass selbstlernende System- und Sprachmodelle oft einen hohen Komplexitätsgrad aufweisen; überdies ist eine sachdienliche Erklärung bei Machine Learning-Verfahren bisweilen technisch nur schwer zu bewerkstelligen²¹.

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

3.2. Datenschutz und Datensicherheit

KI-Programme, insbesondere in Kombination mit juristischen Sprachmodellen, nutzen regelmässig auch Trainingsdaten mit Personenbezügen, die nur unter Einhaltung der datenschutzrechtlichen Vorgaben bearbeitet werden dürfen²⁵. Bei der Ausgestaltung der Automatisierungsvorgänge ist deshalb si-

¹⁹ R. WEBER, S. HENSELER, *Regulierung von Algorithmen in der EU und in der Schweiz*, in EuZ, 2020, S. 28, 38. ([Text 7, S. 103 ff.](#))

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

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

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

²³ R. WEBER, *Künstliche Intelligenz: Regulatorische Überlegungen zum «Wie» und «Was»*, in EuZ, 2022, B1, B12. ([Text 2, S. 17 ff.](#))

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

²⁵ Grundlegend D. ROSENTHAL, *Datenschutz beim Einsatz generativer künstlicher Intelligenz*, in Weblaw Jusletter, 6. November 2023; vgl. auch R. WEBER, *Künstliche Intelligenz und Datenschutz*, in Weblaw IT-Jusletter, 4. Juli 2024 ([Text 8, S. 129 ff.](#)); F. THOUVENIN, ST. VOLZ, *White Paper: Datenschutz*, Juni 2024, <https://www.itsl.uzh.ch/de/Wissenstransfer/Publicationen.html#Positionspapiere>.

herzustellen, dass keine persönlichen Informationen über Datensubjekte, die in Gerichtsverfahren involviert sind, ggf. Drittpersonen zur Kenntnis gelangen. Im Rahmen von privatrechtlichen Verhältnissen lässt sich z.B. auf ein auf der OpenAI Website auffindbares Formular zurückgreifen, mit dem Widerspruch erhoben werden kann²⁶; materiell müssen also Auskunftsrechte, damit der Betroffene überhaupt Kenntnis von der Verwendung der Personendaten erhält, sowie Änderungs- und Löschungsrechte, welche den Betroffenen in die Lage versetzen, die Datenrichtigkeit herzustellen, vorhanden sein²⁷.

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

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

Ein wichtiger Aspekt des Datenschutzrechts ist die Datensicherheit (Art. 8 DSG bzw. Art. 32 DSGVO). Selbst wenn beim Stichwort «Datensicherheit» meist die Risiken im Vordergrund stehen, ist nicht zu übersehen, dass KI-Programme durchaus auch zur Verbesserung der IT-Sicherheit beitragen können; Beispiele sind (i) die Unterstützung bei der Detektion unerwünschter Inhalte, (ii) die Unterstützung bei der Berichterstellung zu Sicherheitsvorfällen, (iii) die Unterstützung bei der Erstellung und Analyse von Programmcodes (Untersuchung von Programmen auf Sicherheitslücken) sowie (iv) die Unterstützung bei der Analyse des Datenverkehrs³¹. Inwieweit die Verwendung von regelba-

²⁶ Vgl. https://share.hsforms.com/1UPy6xqxZSEqTrGDh4ywo_g4sk30.

²⁷ Vgl. auch CH. LUTZ, B. DOMENIG, A. FLÜCKIGER, *Datenschutzkonformer Einsatz von ChatGPT an Schulen*, in: *sic!*, 2024, S. 94, 96.

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

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

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

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

sierten Systemen (z.B. Experten-Entscheidungsbäume) zu einer Erhöhung der IT-Sicherheit beizutragen vermag, weil sie strukturierte Programmabläufe ermöglichen, erscheint derzeit noch als ungeklärt³².

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

Um eine ausreichende Datensicherheit zu gewährleisten, sind Massnahmen technischer und organisatorischer Art gegen mögliche Risiko- oder Missbrauchsfälle rechtzeitig einzurichten; Compliance-Vorkehren müssen die Authentizität von Textverfasser und Textinhalt sicherstellen bzw. den Nachweis bereit halten, welche Texte oder Informationen tatsächlich von einer bestimmten Person, Personengruppe oder Institution stammen³⁵.

3.3. Koordination der Schutzziele

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

- (i) Transparenz schafft grundsätzlich Klarheit und Nachvollziehbarkeit, die im Datenschutzbereich z.B. die Validierung von Personendaten (insbes. Datenrichtigkeit: Art. 6 Abs. 5 DSG bzw. Art. 5 Abs. 1 lit. d DSGVO) einschliesst und die auf die Rückverfolgung und Gewichtung der angewendeten Parameter ausreichenden Wert legt. Der Transparenz zu dienen vermögen auch neue Verfahren zur Authentisierung durch technische Prozesse³⁶. Gleichzeitig gilt der allgemeine (quantitativ zu verstehende)

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

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

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

³⁵ WEBER, zit. (Fn. 1), S. 23.

Grundsatz, dass je mehr Daten offengelegt werden, umso grösser auch das Risiko ist, dass Angriffe von Dritten erfolgreich sein können.

- (ii) Die Datensicherheit lässt sich z.B. erhöhen durch die Einschränkung der Nutzungsmöglichkeiten mit Bezug auf die zur Verfügung stehenden Datenbestände; dies liegt aber oft nicht im Interesse der auf KI-Systeme zurückgreifenden Betroffenen³⁷. Immerhin stehen heute automatisierte Detektionswerkzeuge mit Blick auf automatisch generierte Texte zur Verfügung; dadurch erhalten die Betroffenen, technische Kenntnisse vorausgesetzt, die Fähigkeit, einzelne Texte als maschinengeschrieben zu erkennen und damit ggf. ihre Authentizität und die Richtigkeit der enthaltenen Daten anzuzweifeln³⁸.

Mit Blick auf die Entwicklung transparenter und gleichzeitig sicherer KI-Programme sind verschiedene Risikovermeidungs- und Risikominderungsmaßnahmen in Betracht zu ziehen. Im Vordergrund steht die Sicherstellung der Datenqualität bei der Auswahl von Trainingsdaten. Die konkrete und transparente Bestimmung der Trainingsdaten ist wichtig, um ein hohes Qualitätsniveau mit Blick auf die algorithmischen Ergebnisse zu erreichen. Die verwendeten Trainingsdaten sind zudem Grundlage für die späteren Textgenerierungs-Modelle³⁹. Zur Datenqualität gehört weiter die Vermeidung von (versteckten) Diskriminierungen, die sich durch die Verwendung der Algorithmen perpetuieren könnten⁴⁰.

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

- *Privacy Attacks*: Möglichkeiten zur Verminderung der Anfälligkeit für solche Angriffe sind z.B. die manuelle Auswahl oder automatische Filterung bzw. Anonymisierung von Daten, die Entfernung von Dopplungen aus

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

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

³⁸ WEBER, zit. (Fn. 1), S. 24.

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

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

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

Trainingsdaten zwecks Herabsetzung der Wahrscheinlichkeit einer möglichen Rekonstruktion, die Anwendung von die Differential Privacy verbessernden Mechanismen, die Einschränkung der Zahl von Modellen bzw. des Zugriffs auf das Modell sowie die Vornahme von zusätzlichen Trainings bzw. eines besonderen Trainings für sensible Daten⁴².

- *Adversarial and Poisoning Attacks*: (i) Vorkehren gegen solche Angriffe zielen darauf ab, auch zwecks Erschwerung der Wahrnehmung einer Veränderung absichtlich nur leicht veränderte oder intransparent gestaltete Texte als Verfälschungen erkennbar zu machen⁴³. (ii) Möglichkeiten zur Herabsetzung der Anfälligkeit für solche Angriffe sind die Verwendung vertrauenswürdiger Quellen als Trainingsdaten, der Einsatz von geschultem und erfahrenem Personal, die intensivierte Analyse von Bewertungen und die Beschränkung der Auswirkungen des Einsatzes auf ein kontrollierbares Feld⁴⁴.

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

4. Einsatz von KI-Systemen in der Justiz

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

KI-Programme sind Teil der umfassenden Digitalisierung des Lebens und der beruflichen Tätigkeiten der letzten Jahre. Die Digitalisierung ist aber auch mit Bezug auf die Justizverfahren viel breiter. Das Projekt «Justitia 4.0» will den Aufbau und den Betrieb der Plattform «Justitia Swiss» realisieren⁴⁶. Rechtliche

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

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

⁴⁴ WEBER, zit. (Fn. 1), S. 25.

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

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

Grundlage soll das derzeit im Parlament zur Debatte stehende Bundesgesetz über die Plattformen für die elektronische Kommunikation in der Justiz (BEKJ) sein⁴⁷; der Gesetzesentwurf sieht ein Obligatorium für Justizbehörden und Anwaltschaft vor, diese Plattform in Justizverfahren zu benutzen. Die Grundversion der Plattform ist schon entwickelt worden; mit ihrem Einsatz ist aber frühestens 2025 zu rechnen⁴⁸. Diese Plattform stellt ein digitalisiertes Werkzeug zur Vereinfachung des Dokumentenaustausches dar, hat indessen wenig zu tun mit der Verwendung von KI-Programmen⁴⁹.

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

Bisher geht die überwiegende Lehre in der Schweiz und in Kontinentaleuropa indessen davon aus, dass IT-Systeme formal keine Rechtspersönlichkeit innehaben können⁵¹. Zwar wird erwogen, die Verleihung der Rechtspersönlichkeit an die von den Betroffenen zu übernehmende Verantwortung zu knüpfen, die von IT-Systemen generierten Rechtslösungen als Vergleichsvorschläge zu verstehen, welche von den Rechtsuchenden ggf. zu akzeptieren wären, was eine Zustimmung bedeuten würde⁵²; bisher hat dieser Gedankenansatz aber noch keine verbreitete Akzeptanz gefunden.

Die Rechtspersönlichkeit basiert zudem auch bei juristischen Personen letztlich auf personalen Entscheidungsmechanismen, d.h. natürliche Personen handeln für die «künstliche Einheit»; eine entsprechende Zuordnung müsste vergleichbar bei IT-Systemen bzw. Robotern möglich sein, um die Handlungs-

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

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

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

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

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

⁵² GLESS, zit. (Fn. 9), S. 461.

fähigkeit zu begründen⁵³. Überdies wäre ein Konzept zu entwickeln, das es ermöglichen würde, eine sachgerechte Haftung von IT-Systemen (mit einem entsprechenden Haftungssubstrat) vorzusehen⁵⁴.

4.2. Einsatzfelder in einzelnen Verfahrensstadien der Justiz

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

4.2.1. Feststellung des faktischen Sachverhalts

Der erste Schritt in einem Justizverfahren betrifft regelmässig die Feststellung des faktischen Sachverhaltes. Ein Gericht wird nur tätig, wenn ein entsprechender Antrag, der mit den relevanten Daten unterlegt ist, anhängig gemacht wird⁵⁷; dabei kann die Entscheidung über die Relevanz bzw. Irrelevanz der Daten, verbunden mit dem Problem der Gewährung des rechtlichen Gehörs⁵⁸, durchaus bereits eine anfängliche Herausforderung sein. Grundsätzlich ist aber davon auszugehen, dass KI-Programme bei der Feststellung des Sachverhalts zu helfen in der Lage sind⁵⁹, wie Beispiele aus der Praxis zeigen:

⁵³ WEBER, zit. (Fn. 10), S. 75.

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

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

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

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

⁵⁸ Dazu hinten [Ziff. 4.4.2.](#)

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

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

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

4.2.2. Auffinden von Rechtssätzen

Das Auffinden und Auswerten möglicherweise anwendbarer Rechtssätze mit Blick auf die Entscheidung in einem bestimmten Rechtsfall ist ein Vorgang, der sich für Automatisierungen grundsätzlich eignet. Bereits heute gibt es das «Swiss Federal Supreme Court Dataset» (SCD)⁶³; «Webcrawler» bieten aufgrund publizierter Datensätze von Bundesgerichtsurteilen sachdienliche Auswertungsergebnisse an⁶⁴. Ohne menschliche Intervention geht es aber re-

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

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

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

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

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

gelmässig nicht, weil eine Feinjustierung im Verhältnis zwischen den anwendbaren Rechtsnormen und dem konkreten Sachverhalt meist notwendig ist⁶⁵.

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

4.2.3. Anwendung von Rechtsnormen

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

Gesamthaft betrachtet ist somit festzuhalten, dass sich in allen Ablaufschritten bzw. Verfahrensstadien (auch) KI-Programme einsetzen lassen, dass deren Nutzen jedoch von der konkreten Situation und vom gegebenen Komplexitätsgrad der in Frage stehenden Justizangelegenheit abhängig ist. Zudem muss jedenfalls sichergestellt sein, dass die Rahmenbedingungen der Rechtsordnung eingehalten werden, insbesondere die datenschutzrechtlichen Vorga-

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

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

⁶⁷ Vgl. Fn. 59.

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

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

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

ben⁷¹, aber auch die Bestimmungen des Lauterkeitsrechts zum Schutz vor der unmittelbaren Übernahme und Verwendung eines marktreifen Arbeitsergebnisses (Art. 5 lit. c UWG)⁷².

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

4.3. Vorteile der Automatisierung

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

- *Erwartung von kürzeren und kostengünstigeren Verfahren*: Die Effizienzgewinne durch Verfahrensstraffungen, sofern die verfahrensrechtlichen Grundprinzipien (z.B. rechtliches Gehör, Nichtdiskriminierung, usw.⁷⁷) eingehalten sind, dürften wohl ein unbestrittener Vorteil einer Automatisierung sein⁷⁸. Auch AILexA hat die Fähigkeit, in Sekundenbruchteilen alle Inhalte zu verarbeiten, indem unter allen vorhandenen Datensätzen jeweils die am besten geeignete Lösung gesucht wird. Der Effizienz-Aspekt ist übrigens kürzlich im Strafverfahren Vincenz und Konsorten sowohl von der Ankläger-Seite als auch von der Betroffenen-Seite angesprochen worden⁷⁹.

⁷¹ Vgl. vorne [Ziff. 3.2](#).

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

⁷³ Vgl. vorne [Ziff. 3.1](#).

⁷⁴ BERT steht für Bidirectional Encoder Representations from Transformers.

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

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

⁷⁷ Dazu hinten [Ziff. 4.4.2](#).

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

⁷⁹ Vgl. vorne Fn. 49.

- *Präzisere Rechtsanwendung:* Die automatisierte Subsumtion, basierend auf empirisch gesicherten Vorgaben sowie auf breiten, systematisch abgesuchten Informationssammlungen, ist ggf. der menschlichen Entscheidungsfähigkeit überlegen; eine nachträgliche Plausibilitätskontrolle erweist sich aber in der Regel dennoch als erforderlich⁸⁰. Immerhin lässt sich nicht übersehen, dass es bereits, insbesondere in den USA, recht präzise Vorhersagemodelle für mögliche Prozessergebnisse gibt⁸¹.
- *Grössere Einheitlichkeit der Rechtsanwendung:* Ja mehr Daten von früheren und parallelen Verfahren automatisch gesammelt und in KI-Programmen verarbeitet werden, umso einheitlicher dürfte die Rechtsanwendung durch die Gerichte sein⁸². Diese Einschätzung gilt schon für die nationale Rechtsprechung, noch viel mehr aber für die internationalen Gremien, die (verbindliche und unverbindliche) Entscheidungen fällen⁸³.
- *Ausschluss emotionaler menschlicher Elemente im Entscheidungsprozess:* Wenn Menschen arbeiten, ist es unumgänglich, dass auch personenbezogene Einflussfaktoren eine Rolle spielen. Beteiligte eines Gerichtsverfahrens können mehr oder weniger sympathisch sein, geäußerte Wertvorstellungen entsprechen mehr oder weniger den eigenen Ideen⁸⁴. Obwohl das Sachlichkeitsgebot von der Justiz fordern würde, dass solche Einschätzungen für die Entscheidungsfindung nicht relevant sein dürften, ist deren vollständige «Verbannung aus den Gerichtsräumen» eine Illusion⁸⁵. Automatisierungen und von KI-Programmen «produzierte» Ergebnisse können deshalb menschlichen Risikofaktoren durchaus entgegenwirken.

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

⁸⁰ Vgl. GLESS, zit. (Fn. 9), S. 446.

⁸¹ Vgl. vorne Fn. 69.

⁸² Vgl. GLESS, zit. (Fn. 9), S. 447.

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

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

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

4.4. Kontrolle von Risiken bei der Automatisierung

4.4.1. Maschinen statt Menschen?

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

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

4.4.2. Einzelne Risiken

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

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

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

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

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

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

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

⁹¹ Vgl. immerhin vorne [Ziff. 4.2.1](#) zu den Webcrawlern.

noch nicht ausreichend nachgewiesen. Immerhin lässt sich nicht übersehen, dass die Entscheidungsgründe menschlicher Richter oft auch nicht vollständig nachvollziehbar sind⁹².

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

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

⁹² Vgl. GLESS, zit. (Fn. 9), S. 458.

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

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

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

⁹⁶ Vgl. vorne [Ziff. 2.2.](#)

⁹⁷ Vgl. vorne [Ziff. 3.1.](#)

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

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

- (iii) Den automatisierten Entscheidungen in der Justiz stehen auch *verfassungsrechtliche Bedenken* entgegen, insbesondere die Argumente der Gewaltenteilung und der demokratischen Legitimation, z.B. mit Blick auf die Unabhängigkeit und die Unparteilichkeit der Richter/innen¹⁰⁰. Zweifellos würde der Einsatz umfassender KI-Programme in der Justiz aus verfassungsrechtlicher Sicht zu neuen Herausforderungen führen; dabei kann indessen nicht eine Anpassung der heutigen Rahmenbedingungen im Vordergrund stehen, sondern die zentralen Begrifflichkeiten sind der IT-Charakteristik entsprechend zu umschreiben.

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

- (iv) Als *weitere Risiken* werden in der Diskussion die «Versteinerung» der Rechtsentwicklung¹⁰⁴ und der Wegfall von menschlicher Kompetenz vorgebracht¹⁰⁵. Zwar dürften KI-Programme, selbst wenn die Daten regelmässig und sachgerecht aktualisiert werden¹⁰⁶, zum Teil weniger inno-

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

¹⁰¹ Zur Nachvollziehbarkeit vgl. vorne [Ziff. 3.1](#).

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

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

¹⁰⁴ Dazu GLESS, zit. (Fn. 9), S. 459.

¹⁰⁵ Dazu GLESS, zit. (Fn. 9), S. 460.

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

vativ sein als Menschen, doch ist die individuelle Bereitschaft, neue rechtliche Lösungen zu entwickeln, auch nicht zu überschätzen. Die menschliche Kompetenz ist selbstredend wichtig und muss auch beim Einsatz von KI-Programmen z.B. im Sinne einer Plausibilitätskontrolle ausgeübt werden; ihre teilweise Zurückdrängung hat aber auch den Vorteil, dass der erwähnte Ausschluss menschlicher Emotionen¹⁰⁷ eher möglich ist.

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

4.5. Gegenseitige Abwägung von Vorteilen und Risiken

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

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

¹⁰⁷ Vgl. vorne [Ziff. 4.3](#) mit Fn. 84/85.

¹⁰⁸ Vgl. vorne [Ziff. 4.1](#).

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

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

5. Ausblick

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

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

Wie bereits erläutert, ist die Nachvollziehbarkeit der IT-Prozesse und der KI-Systeme von grosser Bedeutung, d.h. eine «Explainable AI» muss transparent machen, wie ein bestimmter Rechtsanwendungsvorschlag zustande gekommen ist¹¹². Die semi-autonome, KI-unterstützte Anonymisierung von Personendaten in Gerichtsurteilen ist ein weiteres gutes Beispiel¹¹³.

¹¹⁰ Vgl. vorne Fn. 66.

¹¹¹ Vgl. vorne Fn. 74.

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

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

Der richtige Weg sollte also dahin gehen, im Einzelnen zu prüfen, in welchen Bereichen der Einsatz von KI-Systemen eine Entlastung für die Justiz im Sinne einer Ressourcenoptimierung bringen kann (z.B. Automatisierung schematisierter Rechtsbereiche) und wie das Design einer Automatisierung auszugestalten ist, um die notwendige Qualität der Rechtsprechung sicherzustellen¹¹⁴.

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

Ausblick

Digitalisierung und der Kampf ums Recht*

Erstveröffentlichung in: Dal Molin-Kränzlin/Schneuwly/Stojanovic (Hrsg.), *Digitalisierung – Gesellschaft – Recht*, Zürich 2019, 3–23

Inhaltsübersicht

I.	Einleitung: Wellen des technologischen Wandels	202
II.	Datafizierung als technologischer Trend	204
A.	Big Data	204
B.	RegTech	205
III.	Digitalisierung von rechtsrelevanten Tätigkeiten	206
A.	LegalTech als neue Herausforderung	206
B.	Anwendungsbeispiele	208
1.	Ausbildung und Forschung	208
2.	Anwaltsberuf	209
3.	Behördentätigkeiten	210
IV.	Disruption des Rechts durch Technologie?	210
A.	Code is Law	210
B.	Law is Code	211
V.	Weg der künstlichen Intelligenz	214
A.	Algorithmen statt Menschen als Entscheidungsträger?	214
B.	Maschinen als Rechtssubjekte?	217
VI.	Überlebensstrategien des Rechts	218
	Literaturverzeichnis	220
	Materialienverzeichnis	221

* Mit dem Titel «Digitalisierung – Gesellschaft – Recht» haben die Herausgeberinnen des Jubiläumsbandes zur 20. Herausgabe der Schriftenreihe APARIUZ (Analysen und Perspektiven von Assistierenden des Rechtswissenschaftlichen Instituts der Universität Zürich) eine sehr aktuelle und auch herausfordernde Rechtsthematik gewählt. Der einleitende Überblicksbeitrag kann von vornherein nicht alle Aspekte von «Digitalisierung – Gesellschaft – Recht» abdecken und schon gar nicht das bereits umfangreiche Schrifttum umfassend einfangen. Vielmehr geht es darum, aus der Perspektive eines interessierten Beobachters den Versuch zu wagen, die grossen Entwicklungslinien sowie die daraus sich ergebenden rechtlichen Herausforderungen mit beschränkten Fussnotennachweisen aufzuzeigen. Herrn Dr. ALFRED FRÜH, Geschäftsführer des Center for Information Technology, Society and Law (ITSL) an der Universität Zürich, danke ich für eine kritische Durchsicht des Textes.

I. Einleitung: Wellen des technologischen Wandels

Der Erfindungsgeist und die Innovationskraft der Menschheit führen ständig zu technologischen Entwicklungen, die aber durchaus unterschiedliche Auswirkungen auf die Gesellschaft haben können. Erst im Nachhinein, d.h. rückblickend, lässt sich jeweils beurteilen, wie stark sich eine neue Technologie auf die Gesellschaft ausgewirkt hat. Die Erfindung des Buchdrucks (durch JOHANNES GUTENBERG) oder der Dampfmaschine (vermeintlich durch JAMES WATT) sind zweifellos ganz grundlegende Technologiewandel-Erscheinungen gewesen. Ähnliches lässt sich wohl auch für die Entwicklung des Internets bzw. des *World Wide Web* (durch TIM BERNERS-LEE am CERN in Genf) sagen, das die Geschäftsmodelle seit anfangs der Neunzigerjahre des letzten Jahrhunderts stark verändert hat.¹

Immerhin bleibt nicht zu übersehen, dass neue Netzwerk-Infrastrukturen in der Regel vornehmlich einen zusätzlichen Kommunikationskanal anbieten (*Online-Kommunikation*), nicht aber einen eigentlichen Technologiewandel begründen. Auf der ersten Entwicklungsstufe ist der ursprüngliche Informationsaustausch, der in einer mündlichen Form (in Anwesenheit oder über Telefon) oder in einer schriftlich-physischen Form (als Dokumentenlieferung) erfolgte, durch eine unkörperliche Form (*Online-Übermittlung*) ergänzt worden; die Kommunikation bleibt aber «traditionell» verbal. Erst auf einer zweiten Entwicklungsstufe ist eine wesentliche technologische Veränderung durch die Digitalisierung (d.h. durch eine binäre Kodierung von Zeichen, die sich in Einsen und Nullen repräsentieren lässt) und die seit 2008 zur Verfügung stehende *Distributed Ledger Technology* (bzw. *Blockchain-Technologie*) mit der Folge des Vorhandenseins neuer Dateninfrastrukturen eingetreten.

Wenn neue Technologien wie Wellen über die Gesellschaft rollen, stellt sich die Frage, wie das Recht darauf zu reagieren gewillt ist und wie es dies auch zu tun vermag. Bei grundlegenden Veränderungen der Umwelt können sich nämlich disruptive Wirkungen einstellen. In welcher Art geht das Recht mit einem Technologiewandel um? Welche Rechtsbereiche sind für technologische Entwicklungen besonders anfällig?

Von der Theorie her betrachtet kommt dem Recht die Aufgabe zu, eine Steuerungsfunktion auszuüben.² Rechtliche Normen geben die Rahmenordnung vor,

¹ WEBER, *Information Law*, 7 ff. und 375 ff.

² Für einen Überblick am Beispiel des Finanzmarktrechts vgl. WEBER, *The Role of Law*, 253 f. m.w.H.

die von den Menschen bzw. der Gesellschaft in ihren Aktivitäten jedwelcher Art zu beachten ist. Das Recht ist also eine «Sollensordnung», die Verhaltensregeln festlegt. Kommt es zu starken technologischen Veränderungen, erscheint es als unumgänglich, dass sich das eher statische Recht entsprechend anpasst, um bessere Rahmenbedingungen setzen zu können.

Damit das Recht seine ihm zugedachte Steuerungsfunktion zu bewahren vermag, sind Anstrengungen notwendig, um die sachgerechten Anpassungen der normativen Ordnung durchzuführen. Im Jahr 1872 hat der damals sehr bekannte Professor RUDOLF VON JHERING in Wien einen Vortrag gehalten, den er «in wesentlich erweiterter und auf das grössere Lesepublikum zugeschnittener Gestalt unter dem Titel *«Der Kampf ums Recht»* der Öffentlichkeit übergab». Diese knapp 100-seitige Schrift hatte einen ungeheuren Publikumserfolg und wurde in 19 Jahren in 21 Sprachen übersetzt.³ Die Schrift beginnt mit folgender Aussage: «Das Ziel des Rechts ist der Friede, das Mittel dazu der Kampf.» In der Zwischenzeit ist der «Kampf» als Mittel des Rechts stark hinterfragt worden;⁴ dennoch soll nachfolgend der Versuch gewagt werden, die durch vielfältige technologische Veränderungen hervorgerufenen Schwierigkeiten der Rechtsgestaltung im Bild der sich nicht leicht in Einklang zu bringenden (i) technologischen Entwicklungen und (ii) eher starren Rechtsnormen zu reflektieren.

Diese herausfordernde Reise durch die mannigfaltigen potenziellen Disruptionen der Rechtsordnung, verursacht durch die Digitalisierung, wählt folgenden (wenn zwar durchaus auch anders wählbaren) Wanderweg in der theoretisch noch wenig untersuchten rechtlichen Landschaft: Vorerst wird die Datafizierung als technologischer Trend dargestellt, mit besonderem Augenmerk auf die neuen Erscheinungen von *Big Data* und *RegTech*. Hernach geht es um die Digitalisierung rechtsrelevanter Tätigkeiten, die schon eingetreten ist und weiter voranschreitet, auch ungeachtet der neuen Infrastruktur-Technologien (wie z.B. *Blockchain*). Einen eigentlichen (zum Teil auch disruptiven) Paradigmenwechsel erlebt das Recht angesichts des Einsatzes der modernen Infrastruktur-Technologien; seit einigen Monaten wird nicht nur von «*Code is Law*», sondern auch von «*Law is Code*» gesprochen. Eine weitere Stufe der Digitalisierung tritt mit der Verwendung von Algorithmen und dem Rückgriff auf künstliche Intelligenz ein, weil dadurch individuelle Entscheidungsmechanismen, die grundsätzlich den Menschen anvertraut sind, an Maschinen delegiert

³ VON JHERING, Vorwort zur 10. Auflage von 1891.

⁴ Vgl. im Einzelnen dazu BUCHER, FS Gauch, 53 ff.

werden. Abschliessend stellt sich dann die Frage, ob das heutige Recht überhaupt noch überlebensfähig ist.

II. Datafizierung als technologischer Trend

Den Begriff der «Datafizierung» (oder auch «Datifizierung», englisch «*datafication*») haben VIKTOR MAYER-SCHÖNBERGER und KENNETH CUKIER im Jahr 2013 mit Blick auf das neue Phänomen «*Big Data*» geprägt.⁵ Gemeint ist damit ein technologischer Trend, der viele Aspekte des modernen Lebens in computerisierte Daten zu verwandeln vermag.

A. Big Data

In der Informationsgesellschaft steigt nicht nur der Datenaustausch exponentiell, sondern die Bearbeitung der Daten erfolgt in immer effizienterer und ergebnisorientierterer Weise. Mit dem Schlagwort «*Big Data*» wird die Analyse einer regelmässig grossen Menge an aus vielfältigen Quellen erfassten Daten, die nach neuen Methoden und Technologien verarbeitet werden, beschrieben.⁶ *Big Data* zeichnet sich nach der üblichen Umschreibung durch vier Eigenschaften aus, nämlich die Menge der Daten, die Geschwindigkeit der Verarbeitung, die Vielfalt der Daten und die Richtigkeit der Daten. Das grosse Entwicklungspotenzial von *Big Data* ergibt sich auch aus der Möglichkeit, Daten immer wieder zu verwerten, was zu einer erheblichen Verlängerung des bisherigen durchschnittlichen Lebenszyklus der Daten führt. *Big Data*-Analysen erweisen sich aufgrund ihrer breiten Datenbasis meist als objektiv und ergebnisbasiert, doch liefern sie nicht rationale Entscheidungsgrundlagen, sondern stellen nur Wahrscheinlichkeitsrechnungen für eine optimalere Entscheidungsfindung zur Verfügung.⁷

Big Data-Analysen verursachen eine Reihe rechtlicher Herausforderungen. Erweist es sich deshalb als gerechtfertigt, den (fast) unbegrenzten technischen Möglichkeiten rechtliche Grenzen zu setzen? Drei Rechtsbereiche sind Gegenstand der Diskussionen:⁸

⁵ MAYER-SCHÖNBERGER/CUKIER, 50 ff.

⁶ WEBER, *Big Data*, 88 m.w.H.

⁷ Eine weitere Dimension erhalten *Big Data*-Analysen im Falle einer technologischen Verbindung mit Algorithmen bzw. künstlicher Intelligenz (dazu hinten Kap. V).

⁸ Für detailliertere Ausführungen zu diesen rechtlichen Herausforderungen vgl. WEBER, *Big Data*, 92 ff. m.w.H.

- *Big Data*-Analysen können in Konflikt mit verfassungsrechtlichen Grundnormen geraten: Im Vordergrund steht das Gleichbehandlungsgebot bzw. das Diskriminierungsverbot (Art. 8 BV und Art. 9 BV); Datenanalysen dürfen zum Beispiel nicht dazu führen, dass einzelne Personen benachteiligt werden. Ein weiterer rechtlicher Anknüpfungspunkt kann das Prinzip von Treu und Glauben sein (Art. 5 Abs. 3 BV, Art. 2 Abs. 1 ZGB).
- Der Hauptdiskussionspunkt aus rechtlicher Sicht hat bisher die Einhaltung der Datenschutzprinzipien im Falle von *Big Data*-Analysen betroffen: Dass *Big Data*-Analysen nur schwer in Einklang zu bringen sind mit einzelnen Datenschutzprinzipien (z.B. Datenminimierungs- und Verhältnismässigkeitsgrundsatz, Gebot der Zweckbindung), ist heute weitgehend anerkannt, auch wenn Anonymisierungen technisch möglich sind; aus diesem Grunde wird dafür plädiert, durch alternative Konzepte (wie z.B. die *Accountability*, ein *Sharing the Wealth*-Modell) oder durch verbesserte technische Schutzvorkehrungen dem Datenschutz zum Durchbruch zu verhelfen.
- Schliesslich sind verschiedene Spezialgesetze zu beachten, die Vorgaben für *Big Data*-Analysen formulieren: Im Vordergrund stehen insbesondere Gesetze im Versicherungs- und Gesundheitsbereich, weil *Big Data*-Analysen zwar viele positive Wirkungen herbeizuführen vermögen, aber auch erhebliche Herausforderungen, zum Beispiel mit Blick auf das Solidaritätsprinzip, zur Folge haben.

Big Data-Analysen bringen Chancen und Effizienzvorteile; um potenzielle Risiken zu vermeiden, braucht der Gesetzgeber zwar nicht unmittelbar etwas vorzukehren, aber er sollte die Entwicklungen sorgfältig verfolgen, um beim Auftreten von Missbräuchen eingreifen zu können.

B. RegTech

Der Begriff «RegTech», erstmals in England im März 2015 verwendet, lässt sich im engeren Sinne als Verfahren für die organisatorischen *Compliance*-Prozesse im Unternehmen auf Basis neuer digitaler Technologien umschreiben;⁹ im weiteren Sinne geht es bei *RegTech* um den digitalen Austausch von Dokumentationen zwischen den Behörden und den regulatorisch unterworfenen

⁹ Vgl. dazu weiter PETER HILTNER, *Regulatory Technology – Chancen, Risiken und die Frage nach der Aufsicht*, im eingangs erwähnten Sammelband, 305 ff.

Unternehmen (derzeit vor allem den Finanzintermediären).¹⁰ Der Einsatz von *RegTech* setzt das Vorhandensein kohärenter Technologien, vereinheitlichter Datenformate und kompatibler Programmschnittstellen voraus, weil maschinenlesbare Interaktionen ausgeführt werden müssen. *RegTech* hat einen breiteren Anwendungsbereich als *FinTech* und vermag in den verschiedensten Wirtschaftssektoren zu einem Paradigmenwechsel in der gesamten Regulierung zu führen.¹¹

Dass *RegTech* grosse Effizienzvorteile ausspielen kann, ist heute kaum mehr bestritten, doch fehlen weiterhin die ausreichenden internationalen Standardisierungen. Der Bund hat immerhin in einem 25-seitigen Bericht von Mitte 2018 zum Ausdruck gebracht, dass sich die Schweiz den neuen technologischen Entwicklungen nicht verschliessen dürfe.¹² Eine aktive Mitwirkung wäre insbesondere im Rahmen der International Standardization Organisation (ISO) denkbar.

RegTech bedeutet auch, dass der Nexus zwischen Daten und digitaler Identität an Bedeutung gewinnt. Relevant wird damit die Verknüpfung zwischen Personen und deren Daten bzw. deren Herkunft sowie die Verbindung zu anderen Daten. Die Schaffung digitaler Identitäten vermag den digitalen (persönlichen und beruflichen) Informationsaustausch zu erleichtern; die Beschreibung der jeweils handelnden «Personen» führt aber, etwa bei der Beurteilung eines Roboters als Rechtsobjekt, zu neuen rechtlichen Herausforderungen.¹³

III. Digitalisierung von rechtsrelevanten Tätigkeiten

A. LegalTech als neue Herausforderung

Die Entwicklungschancen dank der neuen Informationstechnologien und insbesondere die Einsatzmöglichkeiten neuer digitaler «Werkzeuge» haben schon und werden künftig vermehrt wichtige Veränderungen im Rahmen der rechtsrelevanten Tätigkeiten zur Folge haben. Verbreitet findet für diese The-

¹⁰ WEBER, *RegTech*, 10 f.; CONTRATTO, Rz. 7.

¹¹ WEBER, *RegTech*, 11.

¹² Bericht des Bundesrates, Einsatz innovativer Technologien im Bereich der Finanzmarktaufsicht und -regulierung (*RegTech*) vom 27. Juni 2018.

¹³ Vgl. auch WEBER, *RegTech*, 16 f. m.w.H. sowie hinten Kap. [V.A.](#)

matik in den letzten Jahren der Begriff «LegalTech» Anwendung. Als Elemente im Kontext von LegalTech lassen sich nennen:¹⁴

- Einsatz von algorithmischer Software bei der Rechtsausübung (künstliche Intelligenz, *Machine Learning*);
- Ausbau der Technologien, welche den Umgang mit oder den Zugang zu Recht automatisieren, erleichtern oder erheblich verändern;
- Verwendung von Datenverarbeitungsmethoden (inkl. *Software*- und *On-line*-Dienste), die das Rechtssystem, die Rechtsberatung, die Rechtsdienste und die juristischen Arbeitsprozesse unterstützen oder automatisch durchführen.

Ganz neu sind diese Entwicklungen nicht. Schon vor mehr als zehn Jahren wurde prognostiziert, dass gewisse rechtliche Dienstleistungen keiner individualisierten Erledigung mehr bedürfen, sondern als *Commodities* auch einem Angebot in Warenhäusern offenstehen. Zwischenzeitlich ist zudem noch die technologiebedingte Situation eingetreten, dass automatisierte Dienste die *Commodities* ersetzen. RICHARD und DANIEL SUSSKIND, Vorreiter dieser Entwicklung, haben schon vor Jahren ausgeführt: «*Nothing so special ... cannot be ... understandable on an online basis.*»¹⁵

Das neue technologische Umfeld trägt nicht nur zu einer effizienteren Gestaltung bisheriger Geschäftsmodelle bei, sondern vermag zu stark veränderten Rahmenbedingungen für die Erbringung von Dienstleistungen zu führen. Im Vordergrund steht dabei die Verwendung wertgenerierender Informationstechnologien, die sich in Aspekten wie *Know-how-Management/Sharing*, intelligente Dokumentenanalyse, automatisierte Daten-Analyse sowie Visibilität der angebotenen Produkte ausdrückt. *LegalTech* zeichnet sich weiter durch eine zunehmende Interdisziplinarität der Beteiligten aus; eine kontinuierliche Verbesserung der Ergebnisse ist zudem im Falle des Einsatzes von Algorithmen bzw. von künstlicher Intelligenz erreichbar.¹⁶

LegalTech ist in den verschiedensten Bereichen einsetzbar; nachfolgend kommen die Ausbildung, der Anwaltsberuf und die Behördentätigkeiten zur Sprache.

¹⁴ Die wohl aktivste Rolle beim Thema LegalTech spielt in der Schweiz die Swiss LegalTech Association (SLTA); für weitere Informationen vgl. <https://www.swisslegaltech.ch/>.

¹⁵ SUSSKIND/SUSSKIND, Ziff. 1.7.

¹⁶ Vgl. dazu hinten Kap. [V.A.](#)

B. Anwendungsbeispiele

1. Ausbildung und Forschung

Die Ausbildung auf allen Stufen kommt künftig um den Einsatz neuer Technologien nicht herum. Im Bereich der Rechtswissenschaften auf Universitätsstufe ist eine (teilweise) Abkehr von den traditionell eingesetzten Lehrmitteln und den üblicherweise vermittelten Lerninhalten anzustreben. Eine schon heute praktizierte Methode ist die Aufnahme einer Vorlesung auf *Podcast*, um die Studierenden in die Lage zu versetzen, deren Inhalt zeitverschoben und nach eigenen Präferenzen zur Kenntnis zu nehmen.

Ein *Podcast* kann aber lediglich ein erster Schritt sein. Wie das Beispiel der *Massive Open Online Courses* (MOOC) zeigt, stehen ganz neue Unterrichtsmethoden vor der Tür.¹⁷ Solche Angebote auf *Online*-Plattformen erlauben zudem in viel stärkerem Ausmass, Dozierende und Inhalte grenzüberschreitend kennenzulernen. Als Folge davon stellen sich grundlegende Fragen zur Kontrolle von Lerninhalten und zur Verwendung von Daten der Kursteilnehmer unter dem Gesichtspunkt von Datenschutz und Governance.

Als inhaltlich erforderlich erscheint zudem eine verstärkte Befassung mit den Schnittstellen zwischen dem Recht und den Technologien; ebenso ist die Integration digitaler Inhalte in die juristische Ausbildung erwünscht. Das Lernziel der Ausbildung kann nicht nur darin bestehen, das juristische Handwerkszeug zu beherrschen (d.h. die Leistung), sondern als notwendig erweist sich auch eine Ausrichtung auf die Leistungserbringung (Dienstangebot).¹⁸

In Ergänzung zur Ausbildung ist mit Blick auf die Forschung noch anzufügen: Das schon erwähnte Stichwort der Datafizierung als technologischer Trend ist ebenfalls im Rahmen von wissenschaftlichen Arbeiten verstärkt zu beachten; in interdisziplinärer Weise sollte statistisches Know-how dazu verwendet werden, Entstehungsprozesse von Rechtsetzung und Rechtsprechung genauer zu durchleuchten.¹⁹

¹⁷ Zu den *Massive Open Online Courses* und anderen neuen *Online*-Unterrichtsformen vgl. die Publikation von GLANZ.

¹⁸ HARTUNG, *Anwalts Revue* 2017, 257 ff.

¹⁹ Ein Beispiel dafür ist das vom Zürcher Professor TILMANN ALTWICKER durchgeführte SNF-Projekt mit dem Titel: «An Empirical Approach to Human Rights Interpretation – Hard Interpretative Choices in Judicial Decision Making Processes under the European Convention on Human Rights»; Gegenstand des Forschungsprojekts ist das verbesserte Verständnis der Entstehungsprozesse von Urteilen des Europäischen Menschenrechtsgerichtshofes.

2. Anwaltsberuf

Die Ausübung des Anwaltsberufs dürfte sich in den nächsten Jahren erheblich verändern. Traditionelle Geschäftsmodelle sind anzupassen: Schlagzeilen in den Medien lauten etwa: «Digitale Dienste ersetzen den Anwalt» oder «Dr. iur. Roboter». Anwaltskanzleien müssen parallel einerseits in preisgünstiger Weise die *Commodity*-Produkte und andererseits zu individuellen Tarifen die besonderen, auf die Klientschaft zugeschnittenen Dienstleistungen anbieten. Im Sinne einer Erweiterung des Geschäftsmodells ist auch an die Einrichtung von Plattformen zu denken; der Zugang kann offen (Information für jedermann, etwa über Rechtsentwicklungen im Ausland, z.B. zum Wettbewerbsrecht oder zu den börslichen Kotierungsvorschriften) oder spezifisch geschützt (z.B. für die Klientschaft) sein.²⁰

Die einzelnen Tätigkeitsschritte bei der anwaltlichen Dienstleistungserbringung bzw. Rechtsausübung lassen sich in einem Vergleich zwischen dem heutigen Zustand und den künftigen Anforderungen wie folgt charakterisieren.²¹

- *Information*: Von der Recherche und dem Aktenstudium zur Datenanalyse und zum Einsatz von Algorithmen;
- *Lösungssuche*: Von der Beratung, gegebenenfalls durch ein Spezialistenteam, zur Erarbeitung eines (zum Teil auch interdisziplinären) Lösungsdesigns;
- *Themenbearbeitung*: Von der rechtlichen Ausarbeitung und dem Projektmanagement zur automatisierten Rechtsanalyse (inkl. Compliance, Due Diligence) und zur zeitlich/örtlich flexiblen Prozessorientierung;
- *Präsentation der Lösung*: Von der Qualitätseinschätzung, beruhend auf Vertrauen und Erfahrung, zur Vorlage objektivierter Messbarkeitsfaktoren.

Anwaltstätigkeiten mutieren deshalb mittelfristig zum Angebot spezialisierter Rechtsausübungsdienste. Gewisse Eigenschaften, etwa das Vorhandensein von Vertrauen, bleiben aber auch in der neuen IT-Umgebung erhalten.

²⁰ [Tabelle des Originaltexts gelöscht.]

²¹ Statt vieler vgl. für *LegalTech* im Anwaltsberuf ALBERINI/BERNARD/BUGMANN, *Anwalts Revue* 2017, 253 ff.; QUADE, *Anwalts Revue* 2017, 281 ff.

3. Behördentätigkeiten

Sowohl in der Verwaltung als auch bei den Gerichten lassen sich Effizienzsteigerungen durch digitalisierte und in der Folge auch automatisierte Verfahren erreichen.²² Die Bemühungen erfolgen meist unter dem Stichwort «E-Government»²³ und decken eine breite Spanne an Aktivitäten ab. In seiner digitalen Strategie für die kommenden Jahre legt der Bundesrat ebenfalls viel Wert auf das E-Government.²⁴

Das politisch am meisten diskutierte Thema betrifft das E-Voting:²⁵ In einzelnen Kantonen ist es bereits heute möglich, elektronisch abzustimmen und zu wählen; der Bund unterstützt und fördert entsprechende Projekte. Gegenwind erhalten diese Bemühungen aber von Kritikern, die sich um die Datensicherheit und damit das Risiko der Verfälschung von Abstimmungen und Wahlen sorgen.

Die Digitalisierung geht noch einen Schritt weiter, wenn durch Algorithmen die Möglichkeit besteht, automatisierte (Einzel-)Entscheidungen zu fällen. Eine solche technologische Entwicklung lässt sich aber nur realisieren, wenn die rechtliche Rahmenordnung in verschiedenen Punkten eine Anpassung erfährt; im Vordergrund stehen insbesondere das Legalitätsprinzip und das Recht des Einzelnen auf Gehör, die verfassungsrechtlich vorgegeben sind.²⁶

IV. Disruption des Rechts durch Technologie?

A. Code is Law

Im Kontext der Entstehung des Internetrechts hat LAWRENCE LESSIG vor 20 Jahren in seinem Buch «*Code and other Laws of Cyberspace*» zu Beginn den folgenden kurzen Kapiteltitel gewählt: «Code is Law».²⁷ Mit dieser Begriffsverwen-

²² Vgl. GLASER, SJZ 2018, 181 ff.

²³ Vgl. dazu weiter LUKA MARKIĆ, Die elektronische Stimmabgabe im Lichte des Prinzips der Öffentlichkeit, im eingangs erwähnten Sammelband, 125 ff.

²⁴ Vgl. Bericht des Bundesrates Strategie «Digitale Schweiz» vom 5. September 2018, BBl. 2018, 5972 ff.

²⁵ Vgl. Bericht des Bundesrates zu Vote électronique vom 14. Juni 2013 und Medienmitteilung «Bundesrat beschliesst nächste Schritte zur Ausbreitung der elektronischen Stimmabgabe» vom 5. August 2017.

²⁶ Im Einzelnen dazu hinten Kap. [V.A.](#)

²⁷ LESSIG, Code, 3.

derung ist von LESSIG die Botschaft ausgegangen, dass mit dem Internetrecht ein grundsätzlich neues Rechtsgebiet entstehe, das nicht nur eine zusätzliche «Kategorie» zu den vielen bisherigen Rechtsgebieten darstelle. LESSIG wollte damit der Sache nach FRANK EASTERBROOK widersprechen, der gegen ein gesondertes Internetrecht mit der Begründung Stellung bezog, man hätte ja auch nicht ein «*Law of the Horse*» ins Leben gerufen.²⁸

LESSIG hat sich nicht im Einzelnen mit der Konkretisierung des Begriffs «*Code*» beschäftigt, wie dies einige Jahre zuvor insbesondere NIKLAS LUHMANN²⁹ getan hat, der im *Code* eine binäre Unterscheidung zwischen einem positiven und einem negativen Wert gesehen hat. Vielmehr ist LESSIG davon ausgegangen, das menschliche Verhalten würde reguliert durch komplexe Beziehungen zwischen vier Faktoren, nämlich Recht, Märkte, soziale Normen und Architektur. *Code*-Lösungen reflektieren, ähnlich wie rechtliche Regeln, die vorhandene Information und weisen Rechte daran den Einzelnen zu. Die Architektur als vierter Faktor kombiniert die Vorgaben von Physik, Natur und Technologie.³⁰

Dieses Konzept von LESSIG führt zur Erkenntnis, dass der *Code*, ähnlich wie das Recht, in der Lage ist, eine normative Rahmenordnung für das menschliche Verhalten vorzugeben. *Code* und Recht dürfen aber nicht gleichgestellt werden; vielmehr hat LESSIG in der zweiten, teilweise nicht nur ergänzten, sondern auch veränderten Auflage des ursprünglichen Grundlagenwerks im Jahr 2006 («*Code 2.0*») ausdrücklich wiederholt, dass der *Code* nicht Recht sei, sondern lediglich eine Wirkung auf die Umwelt zu entfalten vermöge und damit einen regulatorischen Einfluss ausüben könne.³¹ Weil der technische *Code* in der *Online*-Welt eine fast perfekte Kontrolle erlaubt, wenn alle technologischen Möglichkeiten ausgenutzt werden, bezeichnet LESSIG die Architektur als den mächtigsten Regulator. Aus diesem Grund erscheint es auch als zentral, festzulegen, wem die entsprechende Machtausübung zusteht.³²

B. Law is Code

Die neuesten Technologien, insbesondere die *Blockchain*- bzw. breiter formuliert die *Distributed Ledger Technologies* (DLT), haben in der Lehre zur Einschätzung geführt, aus der Formel von LESSIG «*Code is Law*» sei die neue For-

²⁸ Zur Kontroverse zwischen LESSIG und EASTERBOOK vgl. WEBER, *Regulatory Models*, 28 f.

²⁹ LUHMANN, 60–70.

³⁰ LESSIG, *Code*, 87 ff.

³¹ LESSIG, *Code 2.0*, 324.

³² LESSIG, *Code*, 130 und 206.

mel «*Law is Code*» abzuleiten.³³ Die Begründung liegt in der (technologischen) Annahme, dass mit dem Einsatz dieser 2008 von einer Person mit dem Pseudonym SATOSHI NAKAMOTO erfundenen neuen Infrastruktur der *Code* ein stabilerer Faktor geworden sei als das Recht. Die *Blockchain*-Technologie erlaubt den Beteiligten die Ausführung von irreversiblen Transaktionen auf einer dezentralen Basis; weil retroaktiv die abgewickelten, kryptografisch geschützten Blöcke nicht mehr rückgängig gemacht werden können, sind Interventionen durch – ihre Meinung ändernde – Transaktionsteilnehmer und insbesondere durch Dritte grundsätzlich nicht mehr möglich.³⁴

Den deutlichsten Ausdruck hat dieses neue Phänomen in den sog. *Smart Contracts* gefunden: Technisch handelt es sich dabei um computerisierte Transaktionsprotokolle, weshalb oft ironisch gesagt wird, ein *Smart Contract* sei weder «*smart*» noch «*contract*». Richtig ist an dieser Aussage, dass der *Smart Contract* sich in einem programmierten *Code* bzw. in einer Computerroutine ausdrückt; ein solcher Vertrag ist auch nicht verbal, denn die Programmiersprache (z.B. Java oder eine andere auf der Ethereum Virtual Machine abbildbare Programmiersprache) ist nur für die IT-Spezialisten verständlich. Das Recht steht somit vor der Herausforderung, die Zulässigkeit von *Smart Contracts* zu akzeptieren oder abzulehnen.³⁵

Die heutige klare Mehrheitsmeinung geht davon aus, dass die Voraussetzungen für einen Vertragsabschluss, insbesondere die übereinstimmenden gegenseitigen Willenserklärungen, auch im Falle der Einigung auf einen *Smart Contract* erfüllt sind, selbst wenn die Vertragsparteien die technische «*Niederschrift*» nicht verstehen. Argumentiert wird mit der Hypothese, wer den Willen äussere, einen *Smart Contract* einzugehen, sei auch damit einverstanden, die Programmiersprache nicht wie Prosa zu verstehen.³⁶ In dieser Perception sind somit *Code* und Sprache identisch; diese Gleichstellung würden Informatiker jederzeit bejahen, für Juristen ist eine solche Einschätzung aber doch eher ungewöhnlich.

³³ DE FILIPPI/HASSAN, *First Monday*, 2016:21:12, 6; WEBER, *Rose is a rose*, 703 f.; vgl. auch NATALIE POMPE, *The new regulatory power of algorithmic information selection*, im eingangs erwähnten Sammelband, 65 ff.

³⁴ Anstelle der schon sehr umfangreichen Literatur zur *Blockchain* sei auf ALEXANDRE LOMBARD, *Smart Property and the Blockchain*, im eingangs erwähnten Sammelband, 243 ff., verwiesen.

³⁵ Für einen allgemeinen Überblick vgl. WEBER, *Smart Contracts*, 291 ff. m.w.H.

³⁶ WEBER, *Smart Contracts*, 293 f.

Ohne auf die weiteren, nicht unerheblichen Herausforderungen für das Vertragsrecht im Kontext von *Smart Contracts* einzugehen, stellt sich aber noch die grundsätzliche rechtstheoretische Problematik, dass die Technologie letztlich über die Rechte und Pflichten der Vertragsparteien entscheidet (und damit auch eine gewisse Ähnlichkeit zur nachfolgend zu erläuternden automatisierten Fällung von Entscheidungen durch Algorithmen aufweist). Wenn das Recht durch die Technologie ausgedrückt wird, liegt die eingangs erwähnte Inversion zur Formel «*Law is Code*» vor bzw. zumindest ist das Recht *Code*-basiert.³⁷

Völlig unproblematisch ist diese Entwicklung indessen nicht; aus einer soziokulturellen Perspektive betrachtet, die nicht nur historisch begründet ist, sind rechtliche Regeln von Menschen für Menschen entwickelt worden.³⁸ Die menschliche Beurteilung der Angemessenheit erlassener Regeln hat also immer auch eine Rolle gespielt, was sich nicht zuletzt an den Auslegungsmethoden für Gesetze und Verträge zeigt, ist doch die historische Auslegung eine akzeptierte Interpretationsmethode. Dieses menschliche Element lässt sich nicht (mehr) realisieren, wenn ein technischer *Code* die Rechtsregeln festlegt, selbst wenn die Programmierer von künstlicher Intelligenz und Algorithmen versuchen könnten, entsprechende technologische «Befehle» aufzunehmen. Schutzvorkehrungen erweisen sich insbesondere als notwendig, um einen technologischen Missbrauch zu vermeiden. Zusammenfassend lässt sich die heutige Situation deshalb wie folgt umschreiben:³⁹

«Law must find ways to regulate the code in order to limit its disruptive potential in the hand of irresponsible code designers since technology tends to rely on code (rather than on law). For the time being, automated legal governance needs to be examined and applied with great caution in order to avoid that new prospects and novel scenarios lead to unforeseeable consequences. The delicate process of transposing legal rules into technical rules might have an important impact on the legal system and on the way we think about law.»

³⁷ WEBER, *Rose is a rose*, 705.

³⁸ DE FILIPPI/HASSAN, *First Monday*, 2016:21:12, 10 f.

³⁹ WEBER, *Rose is a rose*, 705 f.

V. Weg der künstlichen Intelligenz

A. Algorithmen statt Menschen als Entscheidungsträger?

Die neuen (zum Teil disruptiven) Technologien verändern nicht nur – wie eben angesprochen – die Infrastrukturen, sondern auch die Möglichkeiten der Abwicklung von digitalen Werteübertragungen und Dienstleistungen. Immer mehr verfügbar sind intelligente Maschinen und Systeme, die auf selbstlernender Software aufbauen. Die gebräuchlichen Stichworte lauten Algorithmen, künstliche Intelligenz (*Artificial Intelligence*) und *Machine Learning*.⁴⁰

Die rechtliche Herausforderung, die sich hinter diesen neuen Technologien verbirgt, betrifft die Tatsache, dass an sich von Menschen zu fällende Entscheidungen vermehrt an Maschinen bzw. automatisierte Datenauswertungsanlagen übertragen werden (können).⁴¹ Lernende Systeme oder künstliche neuronale Netze ermöglichen algorithmische (Einzel-)Entscheidungen, die sich nicht ohne Probleme in die geltende Rechtsordnung einzuflügen vermögen, weil der Grundsatz, dass das Recht die Leitplanken für das menschliche Verhalten und die gesellschaftliche Organisation festlegt, ins Wanken gerät.⁴²

Das Verhältnis zwischen den Menschen und den Maschinen lässt sich rechtlich in verschiedener Weise regeln.⁴³ Eine sehr klare (wenn zwar auch einseitige) Regelung enthält die nach dem Zweiten Weltkrieg unter dem Eindruck der Industrialisierung geschriebene Verfassung der Freien Hansestadt Bremen; deren Abs. 12 Abs. 1 statuiert das folgende Prinzip: «Der Mensch steht höher als Technik und Maschine [...]». Soweit ersichtlich ist diese Vorschrift in der Rechtslandschaft einzigartig; Rechtsprechung zum Vorrang des Menschen scheint bisher nicht vorzuliegen; würde diese Anordnung wörtlich genommen, wäre die Übertragung von Entscheidungen auf Maschinen ausgeschlossen.

Andere Rechtsnormen sind hingegen wesentlich offener: (i) So verpflichtet die Behindertenrechts-Konvention der Vereinten Nationen (UNO) die Staaten dazu, die Forschung und Entwicklung neuer Technologien zu fördern, um den behinderten Menschen eine bessere Hilfe anbieten zu können (Art. 4 Abs. 1 lit. g); sofern diese Voraussetzung zum Beispiel für automatisierte, mit künst-

⁴⁰ HÖFER, Rz. 2.

⁴¹ Vgl. dazu weiter HANNO MENGES, *Moralische Entscheidungen autonomer Technologien: eine Aufgabe für den Gesetzgeber?*, im eingangs erwähnten Sammelband, 45 ff.

⁴² Grundüberlegungen zu diesem Kapitel sind enthalten in WEBER, *Maschinen*, 72 ff.; dieser Beitrag wird hernach nicht vertieft zitiert. ([Text 3, S. 37 ff.](#))

⁴³ Vgl. DJEFFAL, 17, 18.

licher Intelligenz betriebene Geräte zutrifft, wäre eine staatliche Unterstützung angebracht. (ii) Art. 41 der Europäischen Grundrechtecharta legt fest, dass die Bürgerinnen und Bürger ein Recht auf eine gute Verwaltung haben; sofern in den konkreten Umständen davon auszugehen ist, dass die Algorithmen die Verwaltungsaufgaben besser, schneller und/oder für die Betroffenen einfacher erledigen können, wäre ihr Einsatz in der Verwaltung geboten.

Die Verwendung von Algorithmen hat aber in Einklang mit der geltenden Rechtsordnung zu erfolgen. Im Vordergrund steht dabei die Verfassung. Nach Art. 7 BV ist der Grundsatz der Menschenwürde unantastbar und schützt vor Eingriffen in die physische oder psychische Integrität des Einzelnen. An sich gilt nach der Rechtsprechung und Lehre die sog. «Objektformel»,⁴⁴ d.h. die Gefährdung der Menschenwürde durch algorithmische Entscheidungen darf nicht davon abhängen, ob in einer abstrakten Beurteilung die automatisierten Entscheidungen besser oder weniger gut sind als die menschlichen Entscheidungen. Vielmehr kommt es «nur», aber immerhin, auf die potenzielle Beeinträchtigung der Menschenwürde an.

An Bedeutung gewinnen dürfte im Kontext algorithmischer Entscheidungen das Gleichbehandlungsgebot bzw. Diskriminierungsverbot (Art. 8 BV).⁴⁵ Das Risiko, dass es beim Einsatz von Algorithmen zu Fehlentwicklungen zulasten von potenziell benachteiligten Personen kommt, folgt aus der Tatsache, dass die von intelligenten Systemen verwendeten Daten nicht repräsentativ sind oder keinen relevanten Bezug zur konkreten Datenauswertung aufweisen. Beispielsweise sei an Anstellungsgespräche gedacht, die auf Arbeitgeberseite von einem Roboter geführt werden.⁴⁶

Die Bundesverfassung enthält weiter für staatliches Handeln eine Reihe von Prinzipien, die auch im Falle algorithmischer Entscheidungen zu beachten sind. Im Vordergrund steht das Legalitätsprinzip, d.h. automatisierte Einzelentscheidungen sind nur zulässig, wenn eine gesetzliche Grundlage vorliegt, im Falle eines schwerwiegenden Grundrechtseingriffs sogar in der Gestalt eines formellen Gesetzes.⁴⁷ Diese Voraussetzung ist heute nur in beschränktem Masse erfüllt. Der Entwurf zum neuen Datenschutzgesetz (E-DSG) will nun aber einzelne Lücken des Legalitätsprinzips stopfen: Im Anhang zum E-DSG ist vorgesehen, dass die Festsetzung der Schwerverkehrsabgabe, des Steuerbetrags nach Tabaksteuergesetz und Mineralsteuergesetz sowie die

⁴⁴ MAHLMANN, AJP 2013, 1311.

⁴⁵ Zur Diskriminierungsproblematik vgl. schon vorne Kap. [III.A](#).

⁴⁶ Vgl. dazu WILDHABER, AJP 2017, 213 ff.

⁴⁷ RECHSTEINER, Rz. 14 ff.

Veranlagung nach Zollgesetz und nach Biersteuergesetz automatisiert bzw. durch einen Algorithmus erfolgen darf. Weiter sollen die mit der Durchführung der Unfall- und der Militärversicherung betrauten Organe pauschal die Ermächtigung erhalten, automatisierte Verfügungen zu erlassen.⁴⁸

Weiter zu beachten sind für staatliches Handeln die allgemeinen Verfahrensgarantien, insbesondere der Anspruch auf rechtliches Gehör (Art. 29 Abs. 2 BV).⁴⁹ Beim automatisierten Erlass einer Verfügung fehlt in vielen Fällen die Möglichkeit, dem Betroffenen vorgängig ein Recht auf Äusserung einzuräumen; sofern eine Verfügung zu seinem Nachteil ergehen soll, ist das Recht auf Gehör aber zu gewähren. Eine weitere Anforderung von Art. 29 Abs. 2 BV betrifft das Recht auf Begründung staatlichen Handelns; ob eine algorithmische Entscheidung, gestützt auf künstliche Intelligenz, eine rechtsgenügende Begründung zu liefern vermag, ist zumindest ungewiss, weil Algorithmen regelmässig nicht alle Kausalitäten zu erkennen vermögen. Aus diesen Gründen dürften automatisierte Entscheidungen insbesondere in stark standardisierten Verfahren zur Anwendung kommen.

Die grösste Aufmerksamkeit im Zusammenhang mit automatisierten Entscheidungen hat bisher das Grundrecht auf informationelle Selbstbestimmung in seiner Konkretisierung im Datenschutzrecht erfahren.⁵⁰ Dieses Grundrecht auf Gewährleistung der Privatsphäre schützt jeden Umgang mit persönlichen Daten, d.h. mit solchen Daten, die einen gewissen Bezug zu einer bestimmten oder zu einer bestimmbar Person haben. Die Aufrechterhaltung des Schutzes der Personendaten muss auch gelten, wenn Maschinen die Entscheidungen automatisiert fällen.

Schon das französische Datenschutzgesetz von 1978 enthielt ein grundsätzliches Verbot automatisierter Einzelentscheidungen. Die Datenschutz-Grundverordnung der Europäischen Union (DSGVO) geht in Art. 22 ebenfalls von deren Unzulässigkeit aus, enthält aber gewisse Ausweichklauseln: Ausnahmen sind etwa vorgesehen mit Blick auf den Abschluss oder die Erfüllung eines Vertrages und im Falle einer ausdrücklichen Einwilligung der betroffenen Person, sofern sie auf einer ausreichenden Information beruht. Obwohl in der Schweiz der E-DSG grundsätzlich dem Konzept der DSGVO folgt, wird in Art. 19 nicht ein Verbot von automatisierten Einzelentscheidungen vorgesehen, sondern

⁴⁸ Botschaft zum Bundesgesetz über die Totalrevision des Bundesgesetzes über den Datenschutz und die Änderung weiterer Erlasse zum Datenschutz vom 15. September 2017, BBl. 2017, 6941, 7138 ff.

⁴⁹ Eingehender dazu RECHSTEINER, Rz. 19 ff.

⁵⁰ Vgl. statt vieler THOUVENIN/FRÜH/GEORGE, Rz. 19 ff.; DAEDELLOW, Rz. 13 ff.

lediglich eine (relativ detaillierte) Informationspflicht zugunsten des Betroffenen und ein Recht auf sog. menschliches Gehör, d.h. dem Betroffenen ist die Möglichkeit einzuräumen, sich zum automatisierten Entscheidungsvorgang zu äussern und sich allenfalls dagegen auszusprechen. Ähnlich wie in der EU sind zudem Ausnahmen vorgesehen.⁵¹ Wie sich das Parlament im Laufe des Jahres 2019 zu diesem Vorschlag des Bundesrates stellt, ist schwer prognostizierbar.

B. Maschinen als Rechtssubjekte?

Einer natürlichen Person (und bei Einhaltung der gesetzlichen Vorgaben auch einer juristischen Person) verleiht die Rechtsordnung eine jeweils gesetzlich umschriebene Rechtspersönlichkeit, d.h. die Person ist ein Rechtssubjekt. Jedes Rechtssubjekt hat nach seiner Entstehung grundsätzlich verschiedene (gesetzlich vorgesehene oder im Falle einer Nichtregelung privat festlegbare) Rechte, aber auch korrespondierende Pflichten. Ob automatisierten Datenauswertungsanlagen und selbstlernenden Systemen, weil sie in intelligenter und oft auch effizienter Weise selbstständig Entscheidungen zu treffen in der Lage sind, ebenfalls die Rechtspersönlichkeit zuerkannt werden soll, ist eine erst in den Anfängen erläuterte Rechtsfrage.⁵²

Eine (frühe) bemerkenswerte Rolle hat die Diskussion zur Rechtspersönlichkeit von «Maschinen» im Steuerrecht gespielt; staatliche Behörden (Steuerämter) sind interessiert daran, eine Anknüpfung der Steuererhebung bei den Maschinen ins Auge zu fassen, weil immer mehr Arbeiten nicht mehr von den Menschen erledigt werden (sog. Roboter-Steuer).⁵³ Ein anderer Rechtsbereich, der in der Lehre ebenfalls schon angesprochen wurde, ist die Haftungsordnung: Die Frage der Rechtssubjekteigenschaft spielt eine Rolle, wenn es um die Beurteilung des Entstehenmüssens für einen Schaden geht, den eine automatisiert handelnde Maschine verursacht hat.⁵⁴

Bisher ist in der juristischen Lehre überwiegend, aber nicht einstimmig, die Auffassung vertreten worden, dass den Maschinen keine Rechtspersönlichkeit zukomme. Diese Einschätzung mag auch mit der Tatsache zusammenhängen, dass es sich bei der künstlichen Intelligenz, die algorithmische Entscheidungen ermöglicht, um eine von der Technikgestaltung abhängige Querschnitts-

⁵¹ THOUVENIN/FRÜH/GEORGE, Rz. 23 ff.

⁵² BECK, AJP 2017, 186 ff.; vgl. auch TEUBNER, ACP 2018, 160 ff. und HILGENDORF, Schweizer Monat, Oktober 2017, 68–71.

⁵³ OBERSON, AJP 2017, 236 ff.

⁵⁴ HÄNSENBERGER, Rz. 9 ff.; BLESKIE, Rz. 10 ff.; DENG, Computer & Recht 2018, 70 ff.

materie handelt. Die Quantität der an Maschinen delegierten Entscheidungen darf für die Beurteilung der Rechtssubjekteigenschaft aber ebenso wenig ausschlaggebend sein wie die Qualität der automatisierten Entscheidungen, die im Lichte immer ausgeklügelterer Technologien mittelfristig besser werden sollte. Vielmehr geht es um eine rechtliche Grundsatzfrage, die in ihrer ganzen Tragweite zu beurteilen ist.⁵⁵

Angesichts der Bedeutung dieser Rechtsfrage hat die Europäische Kommission umfangreiche Arbeiten, unterstützt durch Experten verschiedener Disziplinen, aufgenommen, um die Eckpunkte einer normativen Rahmenordnung für die künstliche Intelligenz auszuloten. Im Vordergrund stehen ethische Richtlinien bei der Verwendung von Algorithmen. Zur Vertiefung der Diskussionen hat die Kommission ein «Algorithmic Awareness Building Project» ins Leben gerufen.⁵⁶

Im Sinne eines interdisziplinären Blicks auf die Ethik fällt auf, dass in ähnlicher Weise wie im Recht derzeit auch über eine «Ethik der Algorithmen»⁵⁷ bzw. über eine «Roboter-Ethik» diskutiert wird. Inhaltlich geht es darum, ob die autonomen Maschinen selber über irgendeine Form von ethikrelevantem Bewusstsein oder Moral verfügen.⁵⁸ Eine solche Annahme würde voraussetzen, dass Maschinen auch eine moralische Verantwortung für ihr Tun (bzw. ein Bewusstsein in dieser Hinsicht) zu übernehmen in der Lage wären. Die Mehrheitsmeinung in der Ethik scheint deshalb heute von der Annahme auszugehen, dass anstelle der Schaffung einer Roboter-Ethik vielmehr eine Erweiterung der auf den Menschen bezogenen Ethik sachgerecht wäre.⁵⁹

VI. Überlebensstrategien des Rechts

RUDOLF VON JHERING hat die Frage nach dem «Kampf ums Recht» angesprochen und in seinem historischen Kontext konkretisiert. Waffen sind heute sicherlich nicht mehr notwendig und der «Kampfgedanke» erscheint als überholt. Viel-

⁵⁵ Vgl. auch WEBER, Maschinen, 73.

⁵⁶ European Commission, Medienmitteilung vom 25. April 2018.

⁵⁷ DIEFFAL, 13.

⁵⁸ IEEE, Ethically Aligned Design 20 ff.; HELBING, Rz. 1 ff.

⁵⁹ Vgl. auch die Kontroverse zwischen BESCHORNER, NZZ vom 1. November 2018, 10, BRENNER, NZZ vom 30. November 2018, 10, und SEELE, NZZ vom 13. Dezember 2018, 10; mit einer ähnlichen Fragestellung im eingangs erwähnten Sammelband beschäftigt sich HANNO MENGES, Moralische Entscheidungen autonomer Technologien: eine Aufgabe für den Gesetzgeber?, 45 ff.

mehr geht es darum, das Recht so zu positionieren, dass die normative Ordnung nicht von den Technologien überrollt wird. Oder um etwas näher an VON JHERING zu bleiben: Der Vorwand, das Recht könne mit den sich schnell ändernden Technologien nicht mehr Schritt halten und damit sei der Verzicht auf die Durchsetzung von Recht und Ordnung entschuldigt, vermag nicht zu überzeugen.⁶⁰

Das Recht ist weiter aufgerufen, die Leitplanken für das menschliche Verhalten und die Organisation der Zivilgesellschaft festzulegen. Auch wenn sich der technologische Fortschritt nicht aufhalten lässt, bedarf es einer normativen Grundordnung, die Vorgaben zum Zusammenleben in der Gemeinschaft macht. Diese Einschätzung bedeutet nicht, dass als Instrument des Rechts allgemeine Verbote in den Vordergrund gerückt werden sollten: Die Geschichte lehrt, dass obstruktive Grenzen gegen neue Technologien keine lange Lebensdauer aufweisen: So hat etwa das Verbot des Autofahrens im Kanton Graubünden (1900–1925) nach nicht allzu langer Zeit dem Druck der Gesellschaft weichen müssen, die neue Technologie auf den Strassen zuzulassen.

Die Veränderungen in der Umwelt können indessen nicht spurlos am Recht vorübergehen. Schon in der Ausbildung, aber auch während der gesamten Berufstätigkeit, muss die interdisziplinäre Betrachtung von Einzelfragen an Bedeutung gewinnen. Die in den letzten Jahrzehnten eingetretene Spezialisierung, auch im Recht, bedarf der «Durchbrechung» insoweit, als andere Fachrichtungen, insbesondere die Informatik, Eingang in die tägliche Arbeit zu finden haben. Wer die Technologie nicht zumindest in den Grundsätzen versteht, wird bei der Beurteilung konkreter Rechtsfragen und bei der Ausübung rechtsrelevanter Tätigkeiten künftig vermehrt in Schwierigkeiten geraten.

Das interdisziplinäre Verständnis ist insbesondere von Bedeutung, wenn es um *Distributed Ledger Technologies* (DLT) oder um den Einsatz von Algorithmen geht. Die Rahmenbedingungen des Rechts müssen insoweit als Grundpfeiler die Menschenwürde, das Diskriminierungsverbot und den Schutz der Privatsphäre (informationelle Selbstbestimmung) auch im Kontext neuester Technologien gewährleisten. Diese verfassungsrechtlichen Prinzipien bedürfen der Konkretisierung in einzelnen Bereichen, insbesondere wenn es um die Frage geht, wann und unter welchen Umständen die Maschinen über Menschen entscheiden dürfen. Solchen Herausforderungen ist nicht auszuweichen, sondern es lohnt sich, die spannenden Aufgaben, die sich in der Rechtswissenschaft stellen, an die Hand zu nehmen.

⁶⁰ BUCHER, FS Gauch, 60.

Literaturverzeichnis

- ALBERINI ADRIEN/BERNARD FRÉDÉRIC/BUGMANN LIONEL, Legal Marketplaces: Opportunité et/ou menace pour les avocats et les ordres d'avocats, *Anwalts Revue* 2017, 253 ff.
- BECK SUSANNE, Der rechtliche Status autonomer Maschinen, *AJP* 2017, 183 ff.
- BESCHORNER THOMAS, Roboterethik – eine Schlüsselfrage des 21. Jahrhunderts, *NZZ* vom 1. November 2018, 10
- BLESKIE NICOLAI, Künstliche Intelligenz und haftungsrechtliche Fragen, *Jusletter IT* vom 24. Mai 2018
- BRENNER ANDREAS, «Roboterethik» – Verantwortung setzt Bewusstsein voraus, *NZZ* vom 30. November 2018, 10
- BUCHER EUGEN, Gegen Jherings «Kampf um's Recht», in: Pierre Tercier/Marc Amstutz/Alfred Koller/Jörg Schmid/Hubert Stöckli (Hrsg.), *Gauchs Welt: FS Peter Gauch*, Zürich 2014, 45 ff.
- CONTRATTO FRANCA, «RegTech»: Digitale Wende für Aufsicht und Compliance, *Jusletter* vom 15. August 2016
- DAEDELLOW ROMY, Wenn Algorithmen (unfair) über Menschen entscheiden ..., *Jusletter* vom 26. November 2018
- DE FILIPPI PRIMAVERA/HASSAN SAMER, Blockchain Technology as a Regulatory Technology: From Code is Law to Law is Code, *First Monday*, 2016:21:12, <<http://firstmonday.org/ojs/index.php/fm/article/view/7113/5657>> (19. Dezember 2018)
- DENGA MICHAEL, Deliktische Haftung für künstliche Intelligenz, *Computer & Recht* 2018, 69 ff.
- DJEFFAL CHRISTIAN, Künstliche Intelligenz in der öffentlichen Verwaltung, *Berichte des NEGZ*, Nr. 3, Berlin 2018
- GLANZ UDO, Digitale Diskurskultur in der Bildung. Kommunikationstheoretische Grundlagen, Analyse von Online-Foren und Folgerungen hinsichtlich der Konzeption eines Bildungsportals, Freiburg i.B. 2016
- GLASER ANDREAS, Einflüsse der Digitalisierung auf das schweizerische Verwaltungsrecht, *SJZ* 2018, 181 ff.
- HÄNSENBERGER SILVIO, Die Haftung für Produkte mit lernfähigen Algorithmen, *Jusletter* vom 26. November 2018
- HARTUNG DIRK, Juristische Ausbildung in Zeiten der Digitalisierung, *Anwalts Revue* 2017, 257 ff.
- HELBING DIRK, Societal, Economic, Ethical and Legal Challenges of the Digital Revolution, *Jusletter IT* vom 21. Mai 2015
- HILGENDORF ERIC, Strafrecht für Roboter, *Schweizer Monat*, Oktober 2017, 68 ff.
- HÖFER SEBASTIAN, Algorithmen, maschinelles Lernen und die Grenzen der KI, *Jusletter* vom 26. November 2018
- LESSIG LAWRENCE, *Code and other Laws of Cyberspace*, New York 1999 (zit. LESSIG, Code)
- LESSIG LAWRENCE, *Code 2.0*, New York 2006 (zit. LESSIG, Code 2.0)
- LUHMANN NIKLAS, *Das Recht der Gesellschaft*, Frankfurt a.M. 1993
- MAHLMANN MATTHIAS, Die Garantie der Menschenwürde in der Schweizerischen Bundesverfassung, *AJP* 2013, 1307 ff.

- MAYER-SCHÖNBERGER VIKTOR/CUKIER KENNETH, *Big Data. A Revolution That Will Transform How We Live, Work and Think*, London 2013
- OBERSON XAVIER, *Taxer les robots?*, AJP 2017, 232 ff.
- QUADE BENNO, *Digitale Rechtsabteilung: Und der Anwalt?*, Anwalts Revue 2017, 281 ff.
- RECHSTEINER DAVID, *Der Algorithmus verfügt. Verfassungs- und verwaltungsrechtliche Aspekte automatisierter Einzelentscheidungen*, Jusletter vom 26. November 2018
- SEELE PETER, *Die Theorie der moralischen Stämme*, NZZ vom 13. Dezember 2018, 10
- SUSSKIND RICHARD/SUSSKIND DAVID, *The Future of Professions: How Technology will Transform the Work of Human Experts*, Oxford 2015
- TEUBNER GUNTHER, *Digitale Rechtssubjekte?*, AcP 2018, 155 ff.
- THOUVENIN FLORENT/FRÜH ALFRED/GEORGE DAMIAN, *Datenschutz und automatisierte Entscheidungen*, Jusletter vom 26. November 2018
- VON JHERING RUDOLF, *Der Kampf ums Recht*, 10. Aufl. 1891
- WEBER ROLF H., *Regulatory Models for the Online World*, Zürich 2002 (zit. WEBER, *Regulatory Models*)
- WEBER ROLF H., *Information Law in Swiss Legal Culture*, Bern 2017 (WEBER, *Information Law*)
- WEBER ROLF H., *The Role of Law in Constituting Financial Markets*, 32 *Journal of International Banking Law and Regulation* 2017, 253 ff. (zit. WEBER, *The Role of Law*)
- WEBER ROLF H., *RegTech as a new legal challenge*, CAPCO *Journal of Financial Transformation* 2017, 10 ff. (zit. WEBER, *RegTech*)
- WEBER ROLF H., *Big Data – Rechtliche Grenzen von unbegrenzten Möglichkeiten*, in: Stephan Fuhrer (Hrsg.), *HAVE, Jahrbuch SGHVR* 2018, Zürich 2018, 87 ff. (zit. WEBER, *Big Data*)
- WEBER ROLF H., *Smart Contracts, Vertrags- und verfügungsrechtlicher Regelungsbedarf?, sic!* 2018, 291 ff. (zit. WEBER, *Smart Contracts*)
- WEBER ROLF H., *«Rose is a rose is a rose is a rose» – what about code and law?*, *Computer Law & Security Review* 2018 (34), 701 ff. (zit. WEBER, *Rose is a rose*)
- WEBER ROLF H., *Dürfen Maschinen über Menschen entscheiden?*, *Schweizer Monat*, Februar 2018, 72 ff. (zit. WEBER, *Maschinen*) ([Text 3, S. 37 ff.](#))
- WILDHABER ISABELLE, *Robotik am Arbeitsplatz: Robo-Kollegen und Robo-Bosse?*, AJP 2017, 213 ff.

Materialienverzeichnis

- IEEE (INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS), *Ethically Aligned Design, a Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems*, Version 2, July 2018, <<https://ethicsinaction.ieee.org>> (19. Dezember 2018) (zit. IEEE, *Ethically Aligned Design*)

Synthetische Persönlichkeit und der Schutz des Selbst

Wie künstliche Intelligenz die Grenzen des Menschseins verwischt

Sarah Montani* / Rolf H. Weber**

Erstveröffentlichung in: Weblaw, 1. Dezember 2025

Die Kamera war das Werkzeug des 20. Jahrhunderts, der Prompt ist das Werkzeug des 21. Jahrhunderts. Bislang war Kreativität an Fähigkeiten gebunden: Wer erschaffen wollte, musste wissen, wie; ein Bild zu malen, verlangte Technik. Ein Kunstwerk war bis jetzt untrennbar mit der Leistung des Menschen verbunden, seiner Zeit, seiner Disziplin, seiner Intention. Mit KI-Generatoren verschiebt sich diese Achse radikal: Nicht mehr die Hand, der Pinsel oder die Kamera erschaffen, sondern die Sprache, der Befehl als Prompt. KI-Videogeneratoren verlagern den Akt des Erschaffens vom Handwerk zur Instruktion.

Inhaltsübersicht

1.	Einführung: Was ist SORA und wie funktioniert es?	224
2.	Generierung eines SORA-Videos und rechtlicher Überblick	226
	2.1. Beschreibung des Videos	226
	2.2. Vom Video betroffene Rechtsgebiete	227
3.	Analyse der wichtigsten Rechtsgebiete am SORA-Beispiel	228
	3.1. Persönlichkeitsschutz (ZGB)	228
	3.2. Datenschutz (DSG)	229
	3.3. Urheberrecht/Markenrecht	230
	3.4. Wettbewerbsrecht (UWG)	231
	3.5. Vertrags- und Haftungsrecht	231
	3.6. Strafrecht	232

* SARAH MONTANI, lic.iur., Künstlerin.

** Wir danken Frau Dr. med. Léa Kiener-Schmidlin, Psychiaterin und Chefärztin am CHUV, für die sorgfältige Überprüfung der psychologischen Einschätzungen. Alle Internetseiten wurden zuletzt am 11. November 2025 geprüft.

4.	Weitere Rechtsgebiete und Risiken	232
4.1.	Verfassungsrecht	232
4.2.	Biometrie-, Gentechnik- und Gesundheitsrecht	233
4.3.	Geheimnisschutzrecht	233
4.4.	Kinderrechte und Jugendschutz	234
4.5.	Telekommunikations- und Plattformrecht	234
4.6.	Bank- und Finanzmarktrecht	235
4.7.	Strafrechtliche Spezialdelikte	235
4.8.	Rechtsverletzungen durch KI-basierte Rekonstruktionen	235
4.9.	Haftung und Systemverantwortung	236
5.	Ausserrechtliche Faktoren mit Beeinflussungspotenzial auf das Recht	237
5.1.	Juristisches Multiversum und die «Will-Smith-Benchmark»	237
5.2.	Psychologische Implikationen: Von der Fiktion zur Erlebniswelt	238
5.3.	Synthetische Personen	239
5.4.	«Synthetisches Selbst» und seine Verwundbarkeit	241
5.5.	Neurales Fundament digitaler Verführung	242
6.	Bevorstehende Herausforderungen für die Rechtsordnung	243
6.1.	Umbruch des Schöpfer-Begriffs	243
6.2.	Mensch als Datenschleife?	244
6.3.	Würde im Datenrauschen	245

1. Einführung: Was ist SORA und wie funktioniert es?

[1] SORA ist ein von OpenAI entwickeltes generatives Text-zu-Video-Modell, das als multimodale KI in der Lage ist, auf Basis von Texteingaben («Prompts») oder vorhandenen Bildern/Videos neue Videosequenzen zu erzeugen. Mit SORA kann man sich in nahezu beliebige Szenarien versetzen, allein oder gemeinsam mit anderen. Das Modell erzeugt Videos aus Text oder Bildvorgaben und kann so Personen, Orte und Handlungen frei kombinieren. SORA ermöglicht es, sich selbst in einer natürlichen Umgebung zu zeigen (etwa beim Spaziergang am Meer, auf einem Berggipfel oder bei Sonnenuntergang in der Wüste), in fiktiven Szenen zu inszenieren (als Astronautin, Künstlerin, Redner oder Forscherin) oder in symbolischen oder emotionalen Momenten darzustellen (z.B. bei einer Preisverleihung, in einem Streitgespräch oder beim Wiedersehen mit jemandem).

[2] SORA kann mehrere Personen darstellen, real existierende oder erfundene. So lassen sich Szenen erzeugen wie ein gemeinsamer Tanz auf dem Bundesplatz, eine Diskussionsrunde mit historischen Figuren, ein Familientreffen, das nie stattgefunden hat, oder eine Kooperation mit Prominenten (z.B. Sam Alt-

man oder Musikerinnen). In all diesen Fällen bleibt offen, was «real» und was «inszeniert» ist; dadurch wird SORA zum kreativen, aber auch rechtlich und psychologisch sensiblen Werkzeug. Die Grenze zwischen Erinnerung, Wunsch und Wirklichkeit verschmilzt zu einer einzigen, synthetischen Szene.

[3] SORA kann auf Basis eines Rauschzustandes schrittweise durch einen Diffusionsprozess («Denoising») ein kohärentes Video generieren; dabei nutzt das Modell eine Kombination aus Diffusionsmechanismen und Transformer-Architekturen, um räumlich-zeitliche Konsistenz über mehrere Frames hinweg zu gewährleisten.¹ SORA verwendet eine Technik namens «Recaptioning», bei der visuelle Trainingsdaten mit beschreibenden Captions versehen werden, damit das Modell Textanweisungen besser folgen kann.² Als wesentliche technische Besonderheiten und Metaphern sind zu nennen:

(i) *Start mit zufälligem Rauschen, sukzessive Umwandlung in ein visuelles Signal (Frame), wie beim Entwickeln eines Fotos in der Dunkelkammer:* Zuerst ist alles nur graues Rauschen, dann treten langsam Konturen, Formen und Farben hervor, bis das Bild scharf wird. SORA «entwickelt» sozusagen aus reinem Zufall ein Video, indem es schrittweise das Rauschen entfernt und Strukturen sichtbar macht.

(ii) *Analyse mehrerer Einzelbilder (Frames) gleichzeitig, damit Objekte, die sich kurzzeitig aus dem Bild bewegen, bei der Bewegung realistisch wieder auftauchen und nicht «verschwinden», d.h. SORA denkt nicht Bild für Bild (Frame für Frame), sondern agiert wie ein Regisseur, der beim Filmen schon weiss, wohin sich die Schauspieler gleich bewegen werden.* Anstatt nur das aktuelle Bild zu «sehen», schaut das Modell ein Stück in die Zukunft und Vergangenheit des Videos, um Bewegung und Kontinuität zu wahren. Dadurch «vergisst» es nicht, dass eine Person, die kurz hinter einer Säule verschwindet, gleich wieder auftauchen wird.

(iii) *Möglichkeit, nicht nur Videos direkt aus Text zu generieren, sondern auch bestehende Bilder zu animieren, Videos zu verlängern oder Zwischenbilder zu interpolieren:* SORA kann ein statisches Bild «in Bewegung versetzen», Lücken zwischen Szenen füllen oder eine Geschichte weiterführen, die ursprünglich abrupt endete, wie ein Filmrestaurator, der alte Filmrollen ergänzt oder ein Programm, das alte Fotografien plötzlich zum Leben erweckt.

¹ SORA steht (oder stand) zunächst bestimmten Abonnent*innen von ChatGPT Plus/Pro zur Verfügung. Die Zugänglichkeit zum Zeitpunkt der Publikation ist regional limitiert (z.B. teils noch nicht in der Schweiz).

² Weitere Anbieter sind u.a. Hunyuan Video 混元AIGC视频生成 von Tecent oder Kling, Mini-Max als Teil von Hailuo AI, MOCHI/HAIPER und andere Tools.

(iv) Vorhandensein eingebauter Sicherheitsmechanismen wie sichtbare Wasserzeichen und C2PA-Metadaten zur Herkunftskennzeichnung: Unsichtbaren Metadaten und Wasserzeichen dienen als Echtheits- und Herkunftsnachweis, damit erkennbar bleibt, dass ein Inhalt KI-generiert wurde, ähnlich wie ein Signaturstempel unter einem Dokument oder einem digitalen Fingerabdruck.

2. Generierung eines SORA-Videos und rechtlicher Überblick

2.1. Beschreibung des Videos

[4] In einem Experiment haben wir mit SORA ein Video generiert, in dem das Gesicht der Autorin und das Gesicht von ChatGPT-Gründer Sam Altman verwendet wird. Um dies zu ermöglichen, hat sich die Autorin in der Applikation «gecameout».³ Ein Cameo in diesem Kontext ist ein digitaler Gesicht- oder Körper-Scan, der es ermöglicht, dass eine Person in KI-generierten Inhalten wiedererkennbar dargestellt wird, also ohne real vor der Kamera zu stehen. Für die Nutzung von SORA wird aktuell ein Gesichtsscan vorgenommen, in dem drei zweistelligen Nummern vorgelesen werden und man das Gesicht von links nach rechts und umgekehrt bewegt. Technisch gesehen wird dabei ein biometrisches Profil erstellt, das mindestens aus folgenden Daten besteht: Gesichtsgometrie, Mimik und Gestik, Stimme oder Sprachmuster sowie Körperhaltung, Haarfarbe, Hauttextur usw. Dieses Profil dient als visuelle Identitätsschablone, anhand derer ein KI-Modell die Person in beliebigen Szenen darstellen kann, tanzend, sprechend, diskutierend, ohne dass sie tatsächlich dort ist. Rechtlich ist der Cameo besonders sensibel, weil er biometrische Personendaten enthält.⁴ Wer ein Cameo-Video hochlädt, gibt damit also die Kon-

³ Der Begriff «Cameo» stammt ursprünglich aus der Filmwelt und bezeichnet einen kurzen Auftritt einer bekannten Person in einem Werk. Das englische Substantiv «Cameo» wird im Deutschen inzwischen oft unverändert benutzt, ähnlich wie gepostet, geliket, gedownloaded.

⁴ Technisch dazu was folgt: KI-Modelle müssen lernen, wie unser Mund, unsere Lippen und unsere Stimme zusammenarbeiten, wenn wir sprechen. Das Vorlesen von Zahlen oder zufälligen Worten sorgt dafür, dass unser Gesicht sich in verschiedenen Positionen und Lautformen bewegt (z.B. beim «f», «o», «m» oder «a»). Dadurch kann das System u.a. unsere Gesichtsbewegung mit der Tonspur synchronisieren, authentische Mundbewegungen für jedes Wort generieren und uns später realistisch sprechen oder singen lassen. Zahlen sind ideal, weil sie sprachlich neutral, leicht variierbar und universell verständlich sind. Das Vorlesen einer zufälligen Zahlenfolge stellt zudem sicher, dass die Person selbst vor der Ka-

trolle über ein digitales Abbild und seiner Stimme preis und vertraut darauf, dass die Plattform es weder weiterverwendet noch trainiert.

[5] Das Experiment hat sich wie folgt abgespielt: Zusammen mit anderen Menschen, die T-Shirts mit der Aufschrift «WEBLAW» trugen, tanzten die synthetischen Persönlichkeiten der Autorin und Sam Altmann bis zum Abendrot auf dem Bundesplatz. Das Video wurde anschliessend auf einem sozialen Netzwerk mit einem Disclaimer versehen und gepostet.⁵ Wir analysieren hernach dieses Szenario nach Schweizer Recht.

2.2. Vom Video betroffene Rechtsgebiete

[6] Beim Einsatz des beschriebenen KI-Videogenerators ergeben sich zahlreiche rechtliche Fragestellungen, die sich quer durch verschiedene Rechtsgebiete ziehen:

(i) *Datenschutz und Persönlichkeitsrechte*: Wenn das Modell mit Bildmaterial trainiert oder betrieben wird, das Gesichter enthält, stehen biometrische Merkmale im Raum, die besonders geschützt sind. Die Verwendung von Personenabbildern erfordert grundsätzlich eine wirksame Einwilligung ([Art. 28 ZGB](#), [Art. 6 DSGVO](#)). Weil viele Rechenzentren im Ausland betrieben werden, liegt ggf. eine Datenübermittlung ins Ausland vor ([Art. 16/17 DSGVO](#)). Betroffene müssen informiert werden, wie ihre Daten verwendet werden (Transparenzpflichten, [Art. 19 DSGVO](#)).

(ii) *Urheber- und Immaterialgüterrecht*: Der KI-Trainingsprozess gilt als Vervielfältigung ([Art. 10 URG](#)) und setzt mangels anwendbarer Schrankenregulierung eine Rechtfertigung oder eine Lizenzerteilung voraus. Zudem sind KI-generierte Werke nur dann geschützt, wenn eine menschliche schöpferische Leistung erkennbar ist ([Art. 2 URG](#)). Zwar ist der Stil an sich nicht geschützt, doch kann eine übermässige Nähe (Motivnachahmung) zu einem geschützten Werk eine Verletzung begründen.

mera sitzt und kein Foto oder eine andere KI als Deepfake und die Aufnahme echt ist, man spricht hier von einer «Liveness Verification», also einem Test, der prüft, ob ein echter Mensch interagiert.

⁵ LinkedIn Post Sarah Montani 13. Oktober 2025: Das Video ist eine künstlerische, experimentelle Darstellung zu Forschungszwecken. Es besteht keine geschäftliche Beziehung zwischen SARAH MONTANI, Weblaw, SAM ALTMAN oder OpenAI. Die Darstellung des Gesichts von SAM ALTMAN erfolgt mit ausdrücklicher Einwilligung für die Nutzung in SORA, https://www.linkedin.com/posts/sarah-montani-10146713_sora-legaltech-kiundrecht-activity.

(iii) *Vertrags- und Haftungsrecht*: Anbieter wie OpenAI schliessen üblicherweise vertraglich die Haftung aus oder begrenzen sie. Wenn durch einen erzeugten Inhalt individuelle Rechte Dritter verletzt werden (z.B. Persönlichkeitsrechte, Urheberrechte), kann eine deliktische Haftung ([Art. 41 OR](#)) aufleben. Falls eine KI-Anwendung ein «Produkt» darstellt, sind bei Fehlern oder Schäden die Produkt- oder Softwarehaftungsregeln einschlägig.

(iv) *Wettbewerbs- und Markenrecht*: Wenn ein KI-Video den Eindruck erweckt, dass eine geschäftliche Verbindung oder eine Sponsoringbeziehung besteht (Täuschung, Rufausbeutung), sind die Regeln zum unlauteren Wettbewerb anwendbar. Die Nutzung einer geschützten Marke erfordert eine Erlaubnis, insbesondere wenn sie werblich oder im Wettbewerb eingesetzt wird.

(v) *Medien- und Rundfunkrecht*: Bei einer Verbreitung über öffentliche Kanäle (Streaming, Social Media, TV) könnten medienrechtliche Vorschriften greifen, z.B. Impressumspflicht, Jugendschutz.

(vi) *Strafrechtliche Aspekte*: Deepfakes berühren persönliche Rechte und begründen ggf. Delikte wie Beleidigung, üble Nachrede ([Art. 173 StGB](#)) oder Identitätsverletzung. Illegale Inhalte sind strafbar; Täuschung oder Urkundenfälschung kommen zum Zuge, wenn Inhalte als echt dargestellt werden, obwohl sie fiktiv sind.

3. Analyse der wichtigsten Rechtsgebiete am SORA-Beispiel

[7] Wie erwähnt berührt die Generierung eines SORA-Videos viele Rechtsgebiete. Weil das technische «Produkt» mit den anwendbaren gesetzlichen Bestimmungen kompatibel sein muss, analysieren wir nachfolgend detailliert die einzelnen relevanten Rechtsgebiete.⁶

3.1. Persönlichkeitsschutz (ZGB)

[8] Der Einsatz von Gesichtern (auch digital) ist ein Eingriff in die Persönlichkeit ([Art. 28 ZGB](#)). Eine Einwilligung dazu ist notwendig, gebunden an Zweck und Umfang. Die Autorin hat sich selbst dargestellt (Selbsteinwilligung). Sam

⁶ Der Beitrag präsentiert sehr aktuelle technologische Entwicklungen und analysiert die davon betroffenen Rechtsgebiete; materiell geht es dabei um einen ersten allgemeinen Überblick, ohne dass jede Rechtsfrage mit einem grossen wissenschaftlichen Apparat vertieft wird.

Altman hat sein Gesicht ausdrücklich zur Verfügung gestellt, d.h. er hat aktiv eingewilligt, dass es in SORA-Videos verwendet werden darf. Aktuell ist das Netzwerk SORA deshalb mit Videos von Sam Altman überschwemmt. Wenn Sam Altman seine Zustimmung für «SORA-Trainings- und Demo-Zwecke» gegeben hat, bedeutet dies aber nicht automatisch, dass jede Drittnutzung (z.B. mit einer Marke oder Firma wie Weblaw) erlaubt ist. Wird das Gesicht in einem Kontext gezeigt, der über den Zweck hinausgeht oder einen kommerziellen Bezug herstellt, dürfte die Verwendung durch die Einwilligung nicht mehr gedeckt sein.⁷

[9] Vermutlich sind es genau diese Probleme, die wegen der vielen Nutzenden in Zukunft zu Nachbesserungen führen. Der Prompt «Die tanzenden Menschen tragen ein T-Shirt der Firma Weblaw» hatte nicht funktioniert, die Anweisung «Die tanzenden Menschen tragen ein T-Shirt mit dem Schriftzug Weblaw» brachte dann das gewünschte Resultat. Das Video wurde anschliessend gepostet, d.h. die Autorin hat als «Verantwortliche» gemäss [Art. 5 lit. j DSG](#) die Pflicht, transparent zu informieren (wer, zu welchem Zweck, wo veröffentlicht).

[10] Wenn das Video die Person in einem unerwünschten Licht darstellt, kann eine Ehrverletzung vorliegen ([Art. 28 ZGB](#)/[Art. 173 StGB](#)). Solange das Video nicht diffamierend, herabwürdigend oder verfälschend ist, besteht hingegen kein Ehrenschutzproblem. Tanzen auf dem Bundesplatz ist neutral oder sogar sympathisch. Das Video kann aber suggerieren, dass Sam Altman in einer kommerziellen Kampagne für «Weblaw» mitwirkt (Kontextverfälschung); dies wäre dann überschüssend und zwecküberschreitend ([Art. 28 Abs. 2 ZGB](#)/[Art. 6 Abs. 3 DSG](#)).

3.2. Datenschutz (DSG)

[11] [Art. 5 DSG](#) (Grundsätze der Datenbearbeitung) verpflichtet zur *Rechtmässigkeit, Transparenz und Zweckbindung der Datenbearbeitung*. Beim Erstellen und Veröffentlichen eines SORA-Videos werden verschiedene *Personendaten* bearbeitet, etwa das Gesicht, die Stimme oder der Name von Sam Altman sowie Metadaten der Autorin.

⁷ Zur angemessenen Information, die für eine rechtmässige Einwilligung notwendig ist, vgl. eingehend OFK DSG-FANGER, in: Bieri/Powell (Hrsg.), Kommentar zum DSG, Zürich 2023, Art. 6 Rz. 18 ff.

[12] Die Autorin handelt als *Verantwortliche* im Sinne von [Art. 5 lit. j DSGVO](#) und ist daher verpflichtet, die betroffenen Personen klar zu informieren, wozu die Daten verwendet werden ([Art. 19 DSGVO](#)).

[13] Die Datenbearbeitung muss zudem *zweckgebunden* sein ([Art. 6 Abs. 3 DSGVO](#)): Wenn Sam Altman seine Einwilligung «für SORA-Demozwecke» gegeben hat, umfasst sie nicht automatisch die Nutzung für eine Darstellung mit Firmenbezug («Weblaw»).

[14] SORA wird ausserhalb der Schweiz betrieben; damit liegt eine *Datenübermittlung ins Ausland* im Sinne von [Art. 16/17 DSGVO](#) vor. Die Übermittlung ist nur zulässig, wenn im Empfängerland ein angemessenes Datenschutzniveau besteht oder geeignete Garantien (z.B. Standardvertragsklauseln) vorliegen. Die Autorin muss sich daher vergewissern, dass OpenAI als Anbieterin die Anforderungen des Schweizer Datenschutzrechts erfüllt.

3.3. Urheberrecht/Markenrecht

[15] Das Schweizer *Urheberrechtsgesetz* (URG) schützt schöpferische Werke der Literatur und Kunst ([Art. 2 URG](#)), wozu auch audiovisuelle Werke zählen. Ein von SORA erzeugtes Video kann urheberrechtlich relevant sein, wenn es eine *individuelle geistige Schöpfung* darstellt, z.B. durch originelle Auswahl, Choreografie oder Szenengestaltung. In der Praxis ist der Schutz von KI-generierten Werken aber umstritten, weil das Gesetz von einer *menschlichen Urheberschaft* ausgeht.⁸

[16] [Art. 10 Abs. 2 URG](#) verbietet die unbefugte *Vervielfältigung geschützter Werke*. Falls SORA für das Training urheberrechtlich geschützte Videodaten verwendet wird, liegt eine rechtlich problematische Vervielfältigung vor.⁹ Solange nicht transparent ist, aus welchen Quellen SORA trainiert wurde, besteht das Risiko einer URG-Verletzung.

[17] Das *Markenschutzgesetz* (MSchG), insbesondere [Art. 13 MSchG](#), schützt eingetragene Marken gegen unbefugte Nutzung. Die Marke «Weblaw» ist markenrechtlich geschützt. Wird sie im Video sichtbar verwendet, ohne dass die Markeninhaberin zugestimmt hat, kann dies eine *Markenrechtsverletzung* darstellen, bei kommerziellem Kontext (z.B. Branding, Werbung) ggf. parallel zu einem Verstoß gegen das UWG.

⁸ Vgl. dazu ROLF H. WEBER, *Kreatives Schaffen im Crypto Art Kosmos*, in: Thouvenin et al. (Hrsg.), *Festschrift für Reto Hilty*, Heidelberg 2024, 165, 166 f.

⁹ Vgl. FLORENT THOUVENIN, *White Paper, Urheberrecht, ITSL*, Juni 2025, 3 f.

3.4. Wettbewerbsrecht (UWG)

[18] Das Gesetz gegen den unlauteren Wettbewerb (UWG) schützt vor Täuschung und Rufausbeutung im geschäftlichen Verkehr. Nach [Art. 3 Abs. 1 lit. b UWG](#) handelt unlauter, wer durch irreführende Angaben den Eindruck erweckt, es bestehe eine geschäftliche Verbindung, obwohl dies nicht zutrifft. Wenn das SORA-Video den Anschein erweckt, Sam Altman tanze im Auftrag oder im Rahmen einer Kooperation mit «Weblaw», liegt eine *Irreführung über geschäftliche Verhältnisse* vor.¹⁰

[19] Zudem besteht das Risiko, dass gemäss [Art. 3 Abs. 1 lit. e UWG](#) eine Rufausbeutung angenommen wird: Das Video könnte vom positiven Image Altmans profitieren, um Aufmerksamkeit oder Glaubwürdigkeit zu gewinnen. Selbst wenn kein wirtschaftlicher Gewinn erzielt wird, vermag bereits die Anlehnung an seinen Ruf unlauter zu sein.¹¹

[20] Schliesslich verlangt [Art. 2 UWG](#) die Beachtung von Treu und Glauben im Wettbewerb. Auch ausserhalb klassischer Konkurrenzverhältnisse darf eine Darstellung keine Täuschung oder Imageübernahme fördern. Wird durch ein KI-Video der Eindruck einer geschäftlichen Kooperation erweckt, ohne dass eine solche besteht, liegt unter Umständen *Unlauterkeit* im Sinne des UWG vor, selbst bei künstlerischer Motivation.

3.5. Vertrags- und Haftungsrecht

[21] Wird das Video an Dritte weitergegeben oder öffentlich verbreitet (z.B. auf LinkedIn), entstehen *vertragliche und deliktische Verantwortlichkeiten*. Die Autorin gilt *als Verbreiterin* im Sinne von [Art. 41 OR](#) (unerlaubte Handlung).¹² Falls durch das Video ein Schaden entsteht, etwa wegen Persönlichkeitsverletzung, Markenmissbrauch oder falscher Darstellung, kann sie *schadenersatzpflichtig* werden. Bei kommerzieller Verwertung (z.B. Nutzung auf einer Firmenwebsite) müssten zudem *vertragliche Zusicherungen* über Rechte und Haftung geklärt sein.

[22] Weil das Werk mit einer externen KI erstellt wurde, bestehen *unklare Eigentumsverhältnisse* und potenzielle *Haftungslücken*. Ein allgemeiner Haf-

¹⁰ Vgl. DIKE UWG-BLATTMANN, in: Heizmann/Loacker (Hrsg.), UWG-Kommentar, 2. Aufl. Zürich 2025, Art. 3 Abs. 1 lit. b, Rz. 43 ff.

¹¹ Vgl. DIKE UWG-STAUER/ISKIC, in: Heizmann/Loacker (Fn. 10), Art. 3 Abs. 1 lit. e, Rz. 9 ff.

¹² Nicht nur die Verursachung eines Delikts, sondern auch die Mitwirkung daran, kann widerrechtlich sein.

tungsausschluss bietet keinen sicheren Schutz, wenn Rechte Dritter tangiert sind. Wer ein SORA-Video veröffentlicht, trägt die rechtliche Verantwortung für Inhalt, Kontext und allfällige Folgen, selbst wenn die KI den Inhalt generiert hat.

3.6. Strafrecht

[23] *Ehrverletzungsdelikte* ([Art. 173 ff. StGB](#)): Wird durch das Video eine Person in ihrer Ehre verletzt, indem falsche Tatsachen behauptet oder entwürdigende Szenen dargestellt werden, liegt eine *strafrechtlich relevante üble Nachrede oder Verleumdung* vor, selbst wenn das Video offensichtlich fiktional ist, sofern die Darstellung für Dritte glaubwürdig wirkt.

[24] *Unzulässige Bildaufnahme* ([Art. 179^{quater} StGB](#)): Diese Bestimmung erfasst Bildaufnahmen, die den höchstpersönlichen Lebensbereich verletzen. Bei KI-generierten Videos ohne reale Aufnahme wäre sie nur *analog anwendbar*, könnte aber greifen, wenn der Anschein einer realen, intimen Aufnahme entsteht (z.B. Deepfake mit erotischem Inhalt).

[25] *Pornografische Darstellungen Minderjähriger* ([Art. 197 StGB](#)): Bei generativen Modellen gilt eine *strikte Nulltoleranz*. Selbst die Simulation solcher Szenen wäre strafbar, wenn ein realistischer Eindruck entsteht, dass Minderjährige betroffen sind, was aber bei einer Tanzszene nicht zutreffen dürfte.

4. Weitere Rechtsgebiete und Risiken

[26] Je stärker die KI-Technologie in der Lage ist, visuelle Realität zu erzeugen oder zu imitieren, desto weiter verschiebt sich der juristische Horizont. Heute noch ist SORA ein künstlerisches Werkzeug – morgen könnte es eine rechtliche Fallstudienmaschine werden. Denn mit jedem Fortschritt entstehen neue Grenzüberschreitungen zwischen Technik, Gesellschaft und Recht. Im Folgenden sprechen wir weitere Rechtsgebiete an, die künftig (auch) betroffen sein könnten, wenn KI-Video-Technologien wie SORA eskalieren.

4.1. Verfassungsrecht

[27] Die *Bundesverfassung* (BV) garantiert in Art. 16 und Art. 21 die *Meinungs- und Kunstfreiheit*. Gleichzeitig schützt [Art. 13 BV](#) die *Privatsphäre*, und [Art. 10 BV](#) die *persönliche Freiheit*. Wenn ein KI-Video veröffentlicht wird, das einer realen Person Eigenschaften zuschreibt oder sie satirisch darstellt, muss eine *Güterabwägung* zwischen der Kunstfreiheit und dem Schutz der Persönlich-

keit stattfinden. Das Bundesgericht¹³ verlangt eine *Verhältnismässigkeitsprüfung*. Bei SORA-Videos ist diese Abwägung besonders komplex, weil sie zwischen *Kunst, Technik und Identität* verläuft; der digitale Kunstschaffende greift zum Algorithmus und das Recht sucht nach der Grenze zwischen Ironie und Eingriff.

4.2. Biometrie-, Gentechnik- und Gesundheitsrecht

[28] Damit SORA genutzt werden kann, muss ein Cameo des eigenen Gesichts vorgenommen werden, welches das System auf Gesichtserkennung usw. trainiert. Mit zunehmender Präzision kann KI somit nicht nur Gesichter darstellen und erkennen, sondern auch biometrische oder genetische Merkmale simulieren. Damit geraten wir in die Nähe des *Gentechnikrechts* (GTG) und der geplanten Regelungen zu einer biometrischen Identitätskarte. Wenn etwa SORA hypothetisch ein Video generiert, das Menschen mit genetischen Merkmalen verändert (z.B. Hautfarbe, Augenform oder Körperbau), verursacht die Kunst schnell ein *Diskriminierungs- oder Manipulationspotenzial*, das ethische und medizinrechtliche Fragen, insbesondere bei der *Verwendung von Gesichtern Minderjähriger* oder der Simulation körperlicher Beeinträchtigungen, auslösen könnte. KI, die Menschen verändert, ohne dass sie es wissen, berührt nicht nur das Persönlichkeitsrecht, sie tastet die *biologische Identität* an.

4.3. Geheimnisschutzrecht

[29] Art. 321a OR und Art. 162 StGB schützen vertrauliche Informationen vor unbefugter Offenlegung. Wenn durch SORA KI-Szenen entstehen, die vertraulichen Produktionsprozessen, Prototypen oder Abläufen ähneln, dürfte bereits der *Eindruck einer realen Preisgabe* schaden. Ebenso könnte ein Prompt wie «Zeige das geheime Labor von Firma X» eine faktische Rufschädigung oder eine Geschäftsgeheimnisverletzung darstellen, selbst wenn das Video fiktiv ist. In der Wahrnehmung zählt nicht die Wahrheit, sondern die Glaubwürdigkeit der Fiktion.

¹³ Gemäss BGE 129 III 715, E. 4.1, greift jede erkennbare Abbildung einer Person in die geschützte Persönlichkeitssphäre nach Art. 28 ZGB ein. Eine Veröffentlichung ist nur zulässig mit Einwilligung oder bei überwiegendem öffentlichem Interesse. Der Fall betraf die Veröffentlichung von Fotos eines Prominenten ohne Zustimmung in einem Magazin. Das *Recht am eigenen Bild* schützt nicht nur vor diffamierenden, sondern auch vor unerwünschten, neutralen Darstellungen, wenn sie ohne Zustimmung erfolgen.

4.4. Kinderrechte und Jugendschutz

[30] Die Darstellung von Kindern und Minderjährigen in synthetischen Videos ist besonders heikel. Ausgangspunkt ist [Art. 11 BV](#), der Kindern und Jugendlichen einen besonderen Schutz ihrer Unversehrtheit und eine Förderung ihrer Entwicklung zusichert. Seit 1. Januar 2025 gilt zudem das Bundesgesetz über den Jugendschutz in den Bereichen Film und Videospiele (JSFVG)¹⁴ samt Verordnung (JSFVV)¹⁵. Diese Regelung adressiert u. a. Alterskennzeichnungen, Alterskontrollen und Bereitstellungs-/Zugänglichkeitsvorgaben für Filme und Videospiele.

[31] Unabhängig vom Jugendschutzrecht sind strafrechtliche Grenzen zu beachten: Pornografische KI-Darstellungen, die sexuelle Handlungen mit Minderjährigen zum Inhalt haben, auch virtuelle/nicht-tatsächliche Darstellungen, sind nach [Art. 197 Abs. 4–5 StGB](#) strafbar (Herstellung, Verbreitung, Besitz, Konsum); Rechtsprechung und Lehre bestätigen die Einbeziehung virtueller Kinderpornografie.

[32] Nicht-pornografische KI-Darstellungen können zivilrechtlich unzulässig sein, wenn sie die Persönlichkeitsrechte des Kindes verletzen (z.B. entwürdigende/peinliche Szenen, Deepfakes, Blossstellungen). Zur Anwendung kommen [Art. 28 ff. ZGB](#) (Unterlassung, Beseitigung, Widerruf, Schadenersatz/Genugtuung); ergänzend ist das Datenschutzrecht einschlägig, weil Bild/Video Personendaten sind.

[33] Schliesslich ist die von der Schweiz ratifizierte UNO-Kinderrechtskonvention (KRK) als Auslegungsrahmen zu berücksichtigen (u.a. Schutz vor Ausbeutung, Achtung der Privatsphäre und Würde); sie verstärkt die Pflicht zu einem kindeswohlorientierten Umgang mit KI-Inhalten.

4.5. Telekommunikations- und Plattformrecht

[34] Sobald SORA als soziales Netzwerk agiert, könnte das *Fernmeldegesetz* (FMG) oder eine *NetzDG-ähnliche Regulierung* greifen. Plattformbetreiber wären danach verpflichtet, KI-generierte Inhalte klar zu kennzeichnen, auf Meldung hin zu löschen und Wiederholungsveröffentlichungen zu verhindern. Die

¹⁴ <https://www.fedlex.admin.ch/eli/cc/2024/331/>.

¹⁵ <https://www.fedlex.admin.ch/eli/cc/2024/332/>.

EU verfügt über entsprechende Regeln im *Digital Services Act (DSA)*,¹⁶ der auch für Schweizer Anbieter gilt, sobald sie EU-Nutzer bedienen.

[35] Die Schweiz ist erst daran, die Rechte von Nutzern im digitalen Raum zu stärken. Der Bundesrat hat am 29. Oktober 2025 den Entwurf für ein neues Gesetz über Kommunikationsplattformen und Suchmaschinen in die Vernehmlassung gegeben.¹⁷ Ziel des Gesetzes ist vornehmlich die Bekämpfung illegaler Inhalte auf sozialen Netzwerken; hingegen werden KI-Bots (wie z.B. ChatGPT) nicht geregelt.

4.6. Bank- und Finanzmarktrecht

[36] KI-Videos könnten in Zukunft zur *Marktmanipulation* oder Täuschung von Investoren eingesetzt werden, etwa durch gefälschte CEO-Ansprachen, Finanzberichte oder Produktvorstellungen. Das würde sowohl das *Finanzmarktinfrastrukturgesetz* ([Art. 143 FinfraG](#)) als auch das *Strafrecht* ([Art. 146 StGB](#), *Betrug*) tangieren. Ein überzeugendes Video, das «den CEO der UBS» über eine angebliche Fusion sprechen lässt, könnte Milliarden bewegen.

4.7. Strafrechtliche Spezialdelikte

[37] Neben den klassischen Ehrverletzungs- und Pornografie-Bestimmungen könnten künftig auch Spezialtatbestände relevant werden: [Art. 251 StGB](#) (*Urkundenfälschung*), wenn ein KI-Video als «Beweis» eingesetzt wird, [Art. 271 StGB](#) (*Verbotene Handlungen für einen fremden Staat*) bei propagandistischen Deepfakes und [Art. 258 StGB](#) (*Schreckung der Bevölkerung*) bei simulierten Katastrophen oder Angriffen.

4.8. Rechtsverletzungen durch KI-basierte Rekonstruktionen

[38] Welche Rechtsnormen sind tangiert, wenn SORA reale Bauten oder Schauplätze digital rekonstruiert, etwa das Bundeshaus oder das Matterhorn? Beim Matterhorn handelt es sich um ein national bedeutsames Naturdenkmal/Landschaftsobjekt des BLN.¹⁸ Virtuelle Darstellungen lösen in diesen Regimen keine unmittelbaren Tatbestände aus. Historische Stätten wie das Bundeshaus sind in der Schweiz rechtlich geschützt, der Schutz greift aber

¹⁶ Verordnung 2022/2065 über einen Binnenmarkt für digitale Dienste vom 19. Oktober 2022, ABl 2022 L 277/1 vom 27. Oktober 2022.

¹⁷ Vgl. <https://www.news.admin.ch/de/newnsb/6TmEAd4htulaWG9CWYtK>.

¹⁸ Bundesinventar der Landschaften und Naturdenkmäler von nationaler Bedeutung.

vornehmlich bei räumlichen/baulichen Eingriffen und Planung (Substanz, Umgebung, Nutzung), nicht bei rein virtuellen Darstellungen. KI-Videos, die reale Bauwerke verändern, verfremden oder in fiktiven Zerstörungsszenen zeigen (z.B. das Bundeshaus in Flammen), lösen für sich keine Tatbestände des Kulturgüter- oder Heimatschutzrechts aus. Urheberrechtlich kann eine entstellende Bearbeitung problematisch sein, sofern das betroffene Werk (z.B. ein zeitgenössisches Gebäude oder eine Skulptur) noch geschützt ist (Recht auf Werkintegrität, [Art. 11 URG](#)). Beim Bundeshaus ist der urheberrechtliche Schutz hingegen abgelaufen,¹⁹ sodass die Werkintegrität heute nicht mehr greift. Eine andere Beurteilung kann greifen, wenn ein solcher Clip als real ausgegeben wird und geeignet ist, Angst in der Bevölkerung zu erzeugen (z.B. *Bundeshaus brennt jetzt!*), dann ist ggf. [Art. 258 StGB](#) (Schreckung der Bevölkerung) einschlägig.

4.9. Haftung und Systemverantwortung

[39] Wenn KI-Systeme zunehmend autonom handeln, stellt sich die Frage nach einer *Produkthaftung* für KI. Das heutige Produkthaftungspflichtgesetz ([PrHG](#)) setzt ein fehlerhaftes Produkt voraus, das einen Schaden verursacht. Doch was geschieht, wenn der Schaden von einem Algorithmus ausgeht, der auf Basis von Wahrscheinlichkeiten entscheidet? Die EU hat im Jahre 2024 die (alte) Produkthaftungs-Richtlinie erheblich erweitert, aber den Entwurf für eine spezifische KI-Haftungs-Richtlinie mit *Beweislastumkehr* und *verschärfte Sorgfaltspflichten* wegen des politischen Widerstands im Jahre 2025 wieder zurückgezogen.²⁰ Eine ähnliche Diskussion wird in der Schweiz unvermeidlich sein, sobald generative Systeme wie SORA massenhaft zum Einsatz kommen.²¹

¹⁹ Architekt Hans Wilhelm Auer, † 1906; Schutzdauer 70 Jahre nach dem Tod.

²⁰ Vgl. dazu ISABELLE WILDHABER/FÉDÉRIC BARTH, Regulierung der künstlichen Intelligenz in der Schweiz, AJP 2025, 598, 601 f.

²¹ Der Bundesrat hat sich im Zusammenhang mit der sog. «Auslegeordnung» (Bericht vom 12. Februar 2025) zur KI-Regulation in der Schweiz nicht detailliert zur KI-Haftung geäußert, kurz darauf aber eine Teilrevision des Produktsicherheitsgesetzes angekündigt (vgl. WILDHABER/BARTH [Fn. 20], 604, 606).

5. Ausserrechtliche Faktoren mit Beeinflussungspotenzial auf das Recht

5.1. Juristisches Multiversum und die «Will-Smith-Benchmark»

[40] In der Geschichte der KI-Videoentwicklung hat sich eine banale Szene zum inoffiziellen Massstab für Fortschritt entwickelt: «Will Smith eating spaghetti». Das erste entsprechende Video aus dem Jahr 2023, erzeugt mit einem frühen Text-zu-Video-Modell, zeigte eine grotesk verformte Figur, die klebrige Spaghetti in eine formlose Masse verwandelte, eine Szene zwischen Komik und Unbehagen.²² Sie wurde zum Symbol für die damaligen Grenzen künstlicher Bildkonsistenz: Gesichter zerflossen, Objekte verformten sich, Bewegungen wirkten wie aus einem Albtraum. Mit fortschrittlichen Modellen wie SORA, Hunyuan Video oder Kling ist die Darstellung heute weitgehend stabil und realistisch geworden. In der KI-Community gilt: «Wenn die KI Will Smith endlich korrekt Spaghetti essen lassen kann, hat sie die Welt verstanden.»²³

[41] Während KI-Videos heute schon bestimmte Szenarien (z.B. Sonnenuntergänge, ruhige Portraits, bekannte Persönlichkeiten in eher statischen Szenen) so gut erzeugen können, dass sie auch von Menschen mit hoher Medienkompetenz kaum als künstlich erkannt werden, bleiben komplexere Szenarien mit Bewegung, Interaktion und semantischem Tiefgang herausfordernd. Doch genau diese Herausforderungen werden durch die breite Nutzung von SORA beschleunigt, sodass die Grenzen zwischen «echt» und «künstlich» sehr schnell verschwinden und damit auch die juristischen und ethischen Fragestellungen exponentiell steigen.

[42] Wo früher Kamera und Schauspieler standen, stehen heute Daten bzw. Codes, und das Recht tastet sich durch den digitalen Schatten auf der Suche nach dem neuen Licht. Je weiter die KI-Generalisierung fortschreitet, desto stärker verwischen sich die Grenzen zwischen Fiktion, Realität und Verantwortung. SORA ist nicht nur ein Werkzeug zur Videoproduktion, sie ist ein juristischer Katalysator, der alte Normen auf den Prüfstand stellt.

²² DANIEL DI PLACIDO, 2025. Google's AI Passed the «Will Smith Eating Spaghetti» Test. Forbes, forbes.com/sites/danidiplacido/2025/05/22/google-passed-the-will-smith-eating-spaghetti-test.

²³ ANIL INAMDAR, 2025. Will Smith Eating Spaghetti: Why We Keep Underestimating AI. Medium, <https://medium.com/@anilinandar/will-smith-eating-spaghetti-why-we-keep-underestimating-ai-beb65092ff85>.

5.2. Psychologische Implikationen: Von der Fiktion zur Erlebniswelt

[43] Wenn eine Person sich selbst in einer fiktiven Umgebung sieht, etwa applaudiert für Handlungen, die sie nie begangen hat,²⁴ in einem Kriegsgebiet unter Dauerbeschuss flüchtet oder auf hoher See ertrinkt, können psychologische Effekte entstehen; diese reichen von Identitätsveränderungen über kognitive Verzerrungen bis hin zur emotionalen Belastung.

[44] Die Grenze zwischen dem «realen Selbst» und dem «synthetischen Selbst» kann verschwimmen. Eine aktuelle Panel-Studie zeigt, dass die Identifikation mit einem virtuellen Avatar im Metaverse sowohl das Selbstwertgefühl zu steigern vermag als auch bei Inkohärenz zwischen Offline- und Online-Selbst Identitätsdisruption und verminderte Lebenszufriedenheit nach sich zieht.²⁵ Deshalb kann ein KI-Video, in welchem man sich imaginär sieht, die eigene Identität zumindest kurzfristig verändern. Wenn die Szenen sehr realistisch sind, behandelt das Gehirn sie als verinnerlichte «Erinnerung» oder integriert sie assoziativ in die persönliche Biografie. So zeigte eine Studie,²⁶ dass Teilnehmende nach VR-Erlebnissen virtuelle Ereignisse fälschlich als echte Erinnerungen bezeichneten. Ein hochwertiges KI-Video kann somit potenziell eine falsche Erinnerung erzeugen, mit Auswirkungen auf die Entscheidungsfindung, das Selbstbild oder das Vertrauen in die eigene Wahrnehmung.

[45] Applaus oder Reaktionen in virtuellen Umgebungen erzeugen ein Gefühl von Anerkennung, das in der realen Welt nicht existiert. Ein solcher Effekt würde ein Selbstbild aufbauen, das reale Rückkopplungen jedoch nicht bestätigt. In dieser Situation greift der sogenannte «Proteus-Effekt»: Die Gestaltung eines Avatars beeinflusst das tatsächliche Verhalten der Person.²⁷ Übertragen auf ein KI-Video heisst das: Wenn ich mich selbst als gefeierte Figur

²⁴ Die Autorin zeigt sich in einem Video, in welchem ihr ein fiktiver Nobelpreis für Literatur verliehen wird, LinkedIn Post vom 11. Oktober 2025, https://www.linkedin.com/posts/sarah-montani-10146713_buchtaufe-soranobelpreisimherzen-activity.

²⁵ YANG SOEUN/KIM HAESOO/SONG MINWOO/LEE SEUNGHYUN/JANG JEONG-WOO, 2024. The Double-Edged Influence of Self-Expansion in the Metaverse: A Two-Wave Panel Assessment of Identity Perception, Self-Esteem and Life Satisfaction, *Journal of Virtual Environments*, 8(2), 112–128.

²⁶ MILA BUJIĆ/MIKKO SALMINEN/JUHO HAMARI, 2023. Effects of Immersive Media on Emotion and Memory: An Experiment Comparing Article, 360-Video, and Virtual Reality, *International Journal of Human-Computer Studies*, 103118, 179.

²⁷ NICK YEE/JEREMY N. BAIENSON, 2007. The Proteus effect: The effect of transformed self-representation on behavior, *Human Communication Research* 33(3), 271–290.

sehe, könnte ich beginnen, mich selbst so zu erleben, obwohl meine reale Situation davon abweicht.

[46] Wenn man in virtuellen Szenarien gewisse Handlungen nachspielt, die man im wahren Leben abgelehnt hätte, können Schuldgefühle oder Projektionen entstehen, etwa weil man sich selbst im Film sieht, wie man eine gewalttätige Handlung begeht, oder einer fremden Person die Hand hält und flirtet. Solche Erlebnisse sind oft belastend, weil sie das ethische Selbstbild infrage stellen.

[47] Die Faszination, die eigene Fantasie visuell realisiert zu sehen, kann zur Flucht in simulierte Erlebnisse führen. Der Rückzug aus der realen Welt in eine konstruktive, aber nicht reale Umgebung wird durch immersive Medien belegt.²⁸ Wer später realisiert, dass Vieles virtuell war, kann Frustration, Enttäuschung oder existenzielle Zweifel erleben («Was war wirklich?»). Eine Studie zu Virtual Reality (VR)²⁹ zeigte, dass sich die *wahrgenommene* Selbst-Position («self-location») manipulieren lässt, was Auswirkungen auf das Körper- und Identitätsgefühl hat. Eine starke immersive oder KI-gestützte Darstellung kann das Gefühl von «Ich bin hier gewesen» oder «Ich war das» erzeugen, eine «synthetische Erinnerung».

[48] Diese Effekte sind nicht nur theoretisch: Untersuchungen zu immersiven Medien zeigen, dass intensive virtuelle Erlebnisse das Realitätsgefühl beeinflussen. In Kombination mit derart realitätsgetreuen KI-Videos, wie sie jetzt von jeder Person ohne grosse Kenntnisse generiert werden können, dürfte sich dieses Phänomen verstärken, und zwar insbesondere deshalb, weil man die Illusion hat, sich selbst im Video zu sehen. Wenn Fiktion zur Erlebniswelt wird, steht nicht nur das Bild infrage, sondern die eigene Identität.

5.3. Synthetische Personen

[49] Die Menschen, die in einem SORA-Video vorkommen (Gesichter, Körper, Bewegungen), existieren in der Regel nicht als reale Individuen. Sie sind das Ergebnis von statistischer Synthese, also einer Musterneukombination aus

²⁸ JAMES J. CUMMINGS/JEREMY N. BAILENSON, 2016. How immersive is enough? A meta-analysis of the effect of immersive technology on user outcomes, *Media Psychology* 19(2), 272–309, <https://vhil.stanford.edu/sites/g/files/sbiybj29011/files/media/file/cummings-mp-how-immersive.pdf>.

²⁹ ESTELLE NAKUL/NICOLAS ORLANDO-DESSAINTS/BIGNA LENGGENHAGER/CHRISTOPHE LOPEZ, 2020. Measuring perceived self-location in virtual reality, *Scientific Reports* 10, 19879. <https://doi.org/10.1038/s41598-020-63643-y>.

Millionen visueller Trainingsdaten, die das Modell gelernt hat. Wie eine Malerin, die tausende Porträts studiert hat und nun ein neues Gesicht malt, das so aussehen könnte wie ein Mensch, aber keine Kopie von jemand Bestimmtem ist.

[50] Wenn die Personen nicht real sind, sind, kommt kein klassischer Persönlichkeitsschutz ([Art. 28 ZGB](#)) zum Tragen, weil keine identifizierbare, natürliche Person betroffen ist. Falls ein Gesicht aber versehentlich einer echten Person ähnelt, liegt ggf. eine Persönlichkeitsverletzung vor (z.B. bei starker Ähnlichkeit). OpenAI selbst schreibt in der SORA System Card, dass: «The model does not generate real people unless explicitly instructed and permissioned; generated humans are synthetic and not based on specific individuals.»³⁰ Standardmässig werden synthetische Personen erzeugt, reale Personen (z.B. Sam Altman) dürfen nur erscheinen, wenn eine ausdrückliche Freigabe vorliegt. OpenAI verwendet hierfür sogenannte Identity Filters, die bekannte Gesichter oder Prominente blockieren, es sei denn, sie seien im Modell autorisiert. Der Effekt für Betrachter ist dennoch täuschend echt: Menschen reagieren auf diese Figuren emotional wie auf reale Menschen; Mimik, Blick, Körpersprache wirken zwar manchmal noch ein wenig hölzern, aber dennoch authentisch. Das Gehirn unterscheidet nicht zwischen «statistisch erfundenem» und «wirklich existierendem» Gesicht, solange die Darstellung konsistent ist.

[51] Jeder Mensch trägt ein einzigartiges genetisches Muster (ausgenommen eineiige Zwillinge). Aber: Das Gesicht entsteht aus einer Kombination von mehreren hundert Genen und die möglichen Kombinationen sind endlich. Forschende haben festgestellt, dass das Phänomen von «Menschen, die sich sehr ähnlich sehen, aber nicht verwandt sind» (so-genannte «Look-Alikes» oder «Doppelgänger»), nicht nur eine Kuriosität ist, sondern auch genetische Grundlagen haben kann.³¹ Mit einer globalen Bevölkerung von rund 8 Milliar-

³⁰ Vgl. openai.com/index/sora-system-card.

³¹ Beispielsweise fanden RICKY S. JOSHI/MARIA RIGAU/CARLOS A. GARCÍA-PRIETO/CASTRO DE MANUEL MOURA/DAVID PIÑERO/SEBASTIAN MORAN/VERONICA DAVALOS/PABLO CARRIÓN/MANUEL FERRANDO-BERNAL/IÑIGO OLADE/CARLES LALUEZA-FOX/ARCADI NAVARRO/CARLES FERNÁNDEZ-TENA/DECKY ASPANDI/FEDERICO M. SUKNO/XAVIER BINEFA/ALFONSO VALENCIA/MANEL ESTELLER, 2022, im Rahmen einer Studie mit 32 Paaren von nicht-verwandten Personen, die von Gesichtserkennungs-Algorithmen als sehr ähnlich klassifiziert wurden, heraus, dass neun dieser Paare mehr als 19'000 gemeinsame Single-Nukleotid-Polymorphismen (SNPs) in rund 3'730 Genen aufwiesen, [https://www.cell.com/cell-reports/fulltext/S2211-1247\(22\)01075-0](https://www.cell.com/cell-reports/fulltext/S2211-1247(22)01075-0).

den Menschen sind zufällige Übereinstimmungen im Aussehen³² wahrscheinlicher, als man intuitiv glaubt.

[52] Im Rahmen eines KI-Video-Modells wie SORA entsteht eine zusätzliche Dimension: Einerseits kann eine reale Person ihr eigenes Bild oder Gesicht zur Verfügung stellen und wird somit Teil einer «synthetischen Person». Andererseits lassen sich multiple fiktive oder synthetische Personen erzeugen, etwa Publikum, Politiker, Passanten, die nie real existiert haben. In dieser Welt der generativen KI verschwimmt die Linie zwischen echten «Look-Alikes» und «synthetischen Doppelgängern». Wenn eine KI-Person dargestellt wird, die einer real existierenden Person verblüffend ähnlich ist, aber nicht diese Person ist, entsteht eine synthetische Doppelgänger-Situation.

[53] Wenn ein KI-Video eine Person zeigt, die wirklich existiert oder dieser zum Verwechseln ähnlich sieht, stellt sich wiederum die bereits angesprochene Frage nach Persönlichkeitsrechten, Einwilligung und Missbrauch (Rz. 8). Für den dargestellten Menschen kann es belastend sein, eine «alternative Version» ihrer selbst zu sehen, und sich zu fragen: «Habe ich irgendwo einen echten Doppelgänger?» oder «Bin ich hier wirklich?». SORA ist kein rechtloser Raum, diese KI ist der neue Prüfstein dafür, ob die Rechtsordnung die Würde des Menschen auch im digitalen Spiegel erkennt.

5.4. «Synthetisches Selbst» und seine Verwundbarkeit

[54] Wenn wir unser Gesicht und unsere Stimme digital zur Verfügung stellen, geben wir nicht nur biometrische Daten preis, sondern auch einen Teil unserer inneren Sehnsüchte und damit einen Zugang zu unserer emotionalen Architektur. In der digitalen Welt sind nicht mehr nur die Daten über uns wertvoll, sondern das, was wir wünschen, wenn niemand hinschaut.

[55] Wenn wir uns selbst in einer KI-generierten Szene sehen, reagiert unser Gehirn nicht neutral. Visuelle und emotionale Zentren (insbesondere Amygdala und präfrontaler Cortex) verarbeiten das Bild, als wäre es zumindest teilweise autobiografisch real. Dies führt zu sog. impliziter autobiografischer Integration: Unser Gehirn schreibt das virtuelle Bild in die «innere Biografie» ein.³³

³² ALLEGRA ROSENBERG, 22. November 2024, Do you have a celebrity twin? The surprising science behind doppelgängers, National Geographic, <https://www.nationalgeographic.com/science/article/look-alikes-doppelgangertwin-stranger>.

³³ VILAYANUR S. RAMACHANDRAN, 2012, The Tell-Tale Brain: Unlocking the Mystery of Human Nature, New York: Vintage Books, <https://www.penguinrandomhouse.com/books/307739/the-tell-tale-brain-by-v-s-ramachandran/>.

Diese kognitive Dissonanz, zwischen «Ich weiss, das ist nicht echt» und «Mein Gehirn fühlt, dass es echt ist», kann sowohl Euphorie als auch Stress erzeugen.

[56] Solche Reaktionen verändern die Selbstwahrnehmung. Man erlebt sich als handelnde Figur, obwohl man in Wahrheit Zuschauerin ist. Wenn wir uns selbst als Star auf der Bühne, als gefeierte Forscherin oder als geliebte Heldin darstellen, zeigen wir nicht nur Kreativität, wir offenbaren tiefe persönliche Bedürfnisse: nach Anerkennung, Einfluss, Zugehörigkeit, Schönheit oder Liebe. Die digitalen Selbstinszenierungen werden zu einem Projektionsraum des Unbewussten. Das Gesicht, die Stimme, die Pose, all das wird zu einem emotionalen Datenprofil, das KI-Systeme analysieren und potenziell zur Beeinflussung nutzen können. Wer seine innersten Wünsche offenlegt, wird psychologisch «durchsichtig», ein idealer Ansatzpunkt für algorithmische Beeinflussung, Werbung oder politische Meinungsmache. Die Psychologin Shoshana Zuboff beschreibt das als «Verhaltensüberschuss»: Wir liefern Daten nicht nur über das, was wir tun, sondern über das, was wir begehren, und genau darin liegt die neue Macht der Überwachungskapitalismen.³⁴

[57] Diese Situation bildet die psychologische Basis von *affektiver Profilierung*: Emotionen werden zum Rohstoff des Marketings. Die EU-KI-Verordnung (Art. 5 Abs. 1 lit. a) verbietet KI-Systeme, die Menschen «substanziell manipulieren». Das Schweizer Recht kennt bisher lediglich indirekte Schranken über den Persönlichkeitsschutz ([Art. 28 ZGB](#)) und das Datenschutzgesetz ([Art. 6, Art. 31 DSG](#)). Hingegen schützt das Datenschutzgesetz bislang nur «Personendaten», nicht jedoch ausdrücklich emotionale oder psychologische Profile. Doch sobald sich solche Profile auf konkrete Personen zurückführen lassen, sind sie personenbezogen und ihre unautorisierte Nutzung wäre eine Verletzung des Persönlichkeitsrechts ([Art. 28 ZGB](#) i.V.m. [Art. 6 DSG](#)).

5.5. Neutrales Fundament digitaler Verführung

[58] Selbstwahrnehmung ist kein neutraler Akt, sondern ein emotionaler Reflex. Die KI hat nun gelernt, ihn auszulösen. Wenn wir unser eigenes Gesicht sehen, reagiert unser Gehirn mit einem «Erkennungssignal»: Das limbische System, insbesondere der präfrontale Kortex und das ventrale Striatum, regis-

³⁴ SHOSHANA ZUBOFF, 2019, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York: Public Affairs, 455 ff. <https://www.publicaffairs-books.com/titles/shoshana-zuboff/the-age-of-surveillance-capitalism/9781610395694/>.

trieren: «Das bin ich – vertraut, sicher, vorhersehbar».³⁵ Diese Wiedererkennung löst Dopamin aus, also denselben Neurotransmitter, der auch bei Belohnung oder Erfolg aktiviert wird.³⁶ Unser Gehirn behandelt das eigene Gesicht wie ein sicheres Zuhause. Man fühlt sich «bei sich selbst». Das erklärt, warum man dem eigenen Bild nicht neutral begegnen kann, es ist neurologisch «markiert» als selbstrelevant und belohnend. Schon Kleinkinder erkennen sich im Spiegel³⁷ und reagieren auf ihr Bild nicht wie auf eine fremde Person, sondern wie auf ein Teil ihres Selbst. Das Selbstbild ist also nicht nur visuell, sondern emotional aufgeladen.

[59] Genau hier liegt auch die Brisanz für KI-Videos wie SORA: Wir können ein KI-Video, in dem unser Gesicht vorkommt, kaum *nicht* mögen, weil das «Selbst» tief im Gehirn mit positiven Emotionen verknüpft ist. Diese neuronale Bindung ist so stark, dass Technologien wie SORA an der emotionalen Architektur unserer Identität arbeiten. Das Selbst zu sehen bedeutet, sich lebendig zu fühlen und genau das macht das eigene Bild zur stärksten aller digitalen Drogen. Das schafft Manipulationspotenzial, ähnlich wie bei personalisierter Werbung, nur viel intensiver. Im Sinne des Datenschutzrechts ([Art. 5 lit. c DSGVO](#)) handelt es sich hier nicht nur um biometrische, sondern auch um affektiv verwertbare Daten: das, was wir an uns selbst lieben. Eine solche emotionale Rückkopplung könnte als Verhaltensbeeinflussung gelten und unter Umständen gegen das Prinzip der Verhältnismässigkeit und Zweckbindung ([Art. 6 Abs. 3 DSGVO](#)) verstossen.

6. Bevorstehende Herausforderungen für die Rechtsordnung

6.1. Umbruch des Schöpfer-Begriffs

[60] Menschliche Kreativität wird immer mehr zu einem dialogischen Akt mit der Maschine. Wir entwerfen nicht mehr das Werk, sondern die Anweisung

³⁵ MOTOAKI SUGIURA/YUKO SASSA/HYEONJEONG JEONG/KEISUKE WAKUSAWA/KAORU HORIE/SHIGERU SATO/RYUTA KAWASHIMA, 2008, Self-face recognition in social context in Human Brain Mapping 33(6), 1364–1374, 2012, <https://doi.org/10.1002/hbm.21290>.

³⁶ LUCINA Q. UDDIN/JONAS T. KAPLAN/ISTVAN MOLNAR-SZAKACS/ERAN ZAIDEL/MARCO IACOBONI, 2005. Self-face recognition activates a frontoparietal «mirror» network in the right hemisphere, NeuroImage 25(3), 926–935, www.sciencedirect.com/science/article/abs/pii/S1053811904007669?via=ihub.

³⁷ Das berühmte Rouge-Test-Experiment.

zum Werk. Statt innerer Auseinandersetzung genügt künftig Prompt-Kompetenz: Wer weiss, wie man die Maschine «reizt», gilt als kreativ. Die Wertschätzung verschiebt sich von der Fertigkeit zur Idee. Das Handwerkliche verliert an Bedeutung, während das Konzeptuelle gewinnt. Nach [Art. 2 URG](#) ist schutzwürdig, was eine geistige Schöpfung mit individuellem Charakter darstellt. Aber was soll gelten, wenn die schöpferische Leistung nicht mehr in der Ausführung liegt, sondern in der Steuerung einer KI?

[61] Juristisch deutet sich eine Entwertung des individuellen Werkbegriffs an.³⁸ Kunst wird zunehmend prozedural, nicht mehr personal: Der Akt des Erschaffens ist nicht einmalig, sondern reproduzierbar. Kreativität entsteht traditionell durch Begrenzung und Mühe: aus der Spannung zwischen Idee und Umsetzung. Wenn diese Reibung verschwindet, geht auch der emotionale Reifungsprozess des Schaffens verloren. Es gibt keinen Kampf mehr mit dem Material, keine gedankliche Auseinandersetzung, kein Scheitern. Die KI befreit uns vom Handwerk und raubt uns den Widerstand, an dem Kunst wächst. Der kreative Prozess wird leer: Man generiert Inhalte, aber nicht mehr Erkenntnis.

[62] Gleichzeitig entsteht eine neue Kunstform: die Meta-Kreativität. Sie liegt nicht mehr im Malen, Schreiben oder Filmen, sondern im Kombinieren, Reflektieren und Kuratieren der generierten Welten. Die Kunst verschiebt sich von der Leinwand ins Interface, vom Atelier in die Sprachzeile. Damit eröffnet sich auch juristisch eine neue Kategorie: «Promptkunst», ein schöpferischer Akt, der nicht im Werk selbst liegt, sondern in der Sprache, mit der das Werk entsteht. Mit der Generator-Ära beginnt ein neues kulturelles Zeitalter: Wir schaffen nicht mehr mit der Hand, sondern mit der Wahrscheinlichkeit. Unsere Werkzeuge denken mit, und das Denken selbst wird zum Werkzeug.

6.2. Mensch als Datenschleife?

[63] In der Frühphase des Internets war Teilen das Ziel. Mit YouTube, Facebook und Instagram wurde Erschaffen das Ideal. Menschen produzierten Inhalte als Ausdruck von Können, Meinung oder Zugehörigkeit. Influencer, Blogger und Videokünstler wurden zu neuen Berufsbildern: *Der Mensch als Medium*. Heute aber verschiebt sich der Schwerpunkt: Wir produzieren nicht mehr selbst, sondern lassen mit der KI produzieren. In der Welt von TikTok, Instagram oder X (Twitter) galt Authentizität lange als Währung. Man zeigte sein Leben, seine Stimme, sein Gesicht, seine Identität und Handschrift.

³⁸ Vgl. auch WEBER (Fn. 8), 174 f.

[64] Mit der Einführung von KI-Generatoren wie SORA (Video), ChatGPT (Text) oder Suno/Udio (Musik) wird diese Handschrift algorithmisch reproduzierbar. Jede Person kann in Sekunden ein Video drehen, ohne Kamera und ohne jegliche Kenntnisse. Soziale Medien verwandeln sich in synthetische Bühnen; nicht mehr das Individuum steht im Mittelpunkt, sondern das Ergebnis des Generators.

[65] Der Wandel vom Content-Erzeuger zum Content-Generator markiert einen epochalen Umbruch: Kreativität wird massenhaft, Identität wird reproduzierbar und Wissen wird zugleich grenzenlos – und fragil. Wikipedia war der erste Beweis, dass kollektives Wissen mehr schaffen kann als Einzelne. SORA ist der Beweis, dass kollektive Daten Welten schaffen können, die so echt wirken, dass sie das Reale überblenden. Wir leben nicht mehr in einer Welt des «Content», sondern in einer Welt der synthetischen Wirklichkeiten.

[66] Wenn SORA (oder ein zukünftiges OpenAI-Produkt) tatsächlich ein eigenes soziales Netzwerk wird, entsteht eine fundamentale Veränderung der Dynamik zwischen KI-Modell, Daten und Nutzern. Ein soziales Netzwerk liefert kontinuierliche, authentische, emotional markierte Daten, also genau das, was heute allen KI-Unternehmen fehlt. Wenn ChatGPT (bzw. OpenAI) über SORA analysiert, was Menschen teilen, sagen, fühlen, liken und gleichzeitig das Videomaterial (Gesichter, Stimmen, Emotionen) auswertet, dann entsteht der vollständige Feedback-Kreislauf: Prompt → KI-Output → menschliche Reaktion → neues Training. Das ist für KI-Unternehmen strategisch zentral, denn es schliesst die Lücke zwischen Datenquelle und Produktoptimierung. Soziale Netzwerke zeigen nicht nur, *was* Nutzer produzieren, sondern *wie sie reagieren*. Diese Daten sind wertvoller als klassische Trainingsdaten: Daraus lässt sich lernen, was menschliche Resonanz erzeugt. KI-Videos werden dadurch *emotionaler und manipulativer*. Wenn SORA als soziales Netzwerk funktioniert, hat OpenAI direkten Zugriff auf eine kritische Masse echter Nutzerinhalte und ist nicht mehr auf fremde Plattformen (TikTok, YouTube, Instagram) angewiesen.

6.3. Würde im Datenrauschen

[67] «Handle so, dass du die Menschheit sowohl in deiner Person, als in der Person eines jeden anderen jederzeit zugleich als Zweck, niemals bloss als Mittel brauchst.» schreibt Immanuel Kant 1785 in *Grundlegung zur Metaphysik der*

Sitten.³⁹ Mit SORA und vergleichbaren Systemen tritt die Menschheit in eine neue epistemische Epoche ein: Wir sind nicht länger nur Urheber oder Betrachtende von Bildern, sondern selbst Teil der Daten, aus denen sich neue Welten synthetisieren. Der Mensch wird im wörtlichen Sinne zum Trainingsmaterial seiner eigenen Simulation. Wo früher das Subjekt den Blick in die Welt richtete, richtet die Welt der Daten nun den Blick auf das Subjekt zurück – und macht es zum Objekt.

[68] Der Mensch steht nicht mehr handelnd in der Welt, sondern erzeugt sie mit Prompts und Worten. Seine Welt wird simuliert, sein Tun abstrahiert, seine Spur ist ein Datensatz. Diese Entkörperlichung führt zu einem paradoxen Zustand: Wir erscheinen überall und sind nirgends wirklich. Der digitale Doppelgänger ist präsent, das Subjekt wird abwesend. Der französische Philosoph Jean Baudrillard sah in den Simulationen des 20. Jahrhunderts bereits die «Vernichtung des Realen durch seine perfekte Reproduktion».⁴⁰ SORA bringt diese Diagnose ins audiovisuelle Zeitalter: Was wir sehen, ersetzt, was wir glauben.

[69] Die menschliche Würde, in der Schweiz durch [Art. 7 BV](#) verfassungsrechtlich garantiert, beruht auf der Idee, dass jeder Mensch ein unverfügbarer Träger von Sinn und Eigenwert ist. In der digitalen Ökonomie der Gefühle aber verwandelt sich dieser Eigenwert in Datenwert. Was wir fühlen, begehren und darstellen, wird zur Ressource im System.

[70] Mit jedem Klick, jedem Video, jedem synthetischen Selbstporträt trainieren wir nicht nur ein Modell, sondern geben auch einen Teil unserer affektiven Struktur preis: was wir lieben, was wir fürchten, was wir als «wir» empfinden. Damit wird die Persönlichkeit, einst der unantastbare Kern der Würde, ökonomisch reproduzierbar. Der Mensch entwickelt sich zur Schleife im Lernprozess der Maschine, zur Rückkopplungseinheit einer algorithmischen Zivilisation.

[71] Die Ambivalenz der KI liegt in ihrer doppelten Natur: Sie ist Werkzeug der Befreiung und der Unterwerfung zugleich. Sie befreit von Mühe, schafft Zugang zu Kreativität, aber sie nimmt uns auch die Erfahrung des eigenen Schaffens. Der Mensch als *Promptgeber* bleibt scheinbar souverän, doch sein Befehl nährt ein System, das ihn zugleich studiert, spiegelt und optimiert.

³⁹ IMMANUEL KANT, 1785. Grundlegung zur Metaphysik der Sitten (Akademie-Ausgabe IV) S. 429. Riga: Johann Friedrich Hartknoch, <https://korpora.org/kant/aa04/429.html>.

⁴⁰ Vgl. JEAN BAUDRILLARD, *Simulacres et simulation*. Éditions Galilée, Paris 1981.

[72] Die Rechtsordnung steht damit vor einer ethischen Aufgabe; sie muss die digitale Person als Trägerin von Rechten schützen, auch wenn sie nur als Datenkonstellation existiert. [Art. 28 ZGB](#) schützt die Persönlichkeit vor Verletzungen durch Dritte. Falls die Verletzung aber nicht durch einen Menschen, sondern durch ein Modell erfolgt, das uns spiegelt und mit uns in Resonanz tritt, fehlt ein Schutz. Überdies ist die klassische Vorstellung von Schuld und Handlung verwischt, Verantwortung wird verteilt, verdünnt, algorithmisiert.

[73] Die Zukunft der Würde des Menschen liegt somit nicht im Algorithmus, sondern in der Fähigkeit, ihn zu befragen und zu hinterfragen, ebenso wie in der bewussten Entscheidung, nicht alles zu generieren, was generierbar ist.

Veröffentlichungen des Center für Information Technology, Society and Law (ITS�) der Universität Zürich

erschienen bei Schulthess Juristische Medien AG, Zürich

- Vol. 1: Werbung – Online**
FLORENT THOUVENIN/ROLF H. WEBER (Hrsg.) – Zürich 2017
- Vol. 2: Transatlantic Data Protection in Practice**
ROLF H. WEBER/DOMINIC N. STAIGER – Zürich 2017
- Vol. 3: Endorsements and Behavioral Advertising in Social Media under EU, Swiss, and US Law**
MANE SARGSYAN – Zürich 2017
- Vol. 4: Nutzung von Cloud-Diensten durch Anwältinnen und Anwälte/Utilisation des services de cloud par les avocates et avocats**
CHRISTIAN SCHWARZENEGGER/FLORENT THOUVENIN/BURKHARD STILLER – Zürich 2019
- Vol. 5: Die Informationspflichten des Unternehmers im E-Commerce**
DOMINIC OERTLY – Zürich 2019
- Vol. 6: Der IT-Outsourcingvertrag im schweizerischen Recht**
ALEXANDER SCHMID – Zürich 2019
- Vol. 7: Elemente einer Datenpolitik**
FLORENT THOUVENIN/ROLF H. WEBER/ALFRED FRÜH – Zürich 2019
- Vol. 8: Prinzipien und Rechtmässigkeitsbedingungen im privaten Datenschutzrecht**
DAMIAN GEORGE – Zürich 2021

erschienen bei EIZ Publishing, Zürich

- Vol. 9: Cybersicherheit und Cyber-Resilienz in den Finanzmärkten**
ROLF H. WEBER/OKAN YILDIZ – Zürich 2022
- Vol. 10: DLT-Handelsplattformen**
ROLF H. WEBER – Zürich 2022
- Vol. 11: Automatisierte Einzelentscheidungen im (Schweizer) Datenschutzrecht – Phänomen, Entstehungsgeschichte, Regelungsziele und Begriff**
FABIENNE SUTER – Zürich 2024
- Vol. 12: Ein Rechtsrahmen für KI in der Schweiz: Perspektive öffentliches Recht**
NADJA BRAUN BINDER, FLORENT THOUVENIN (Hrsg.) – Zürich 2025
- Vol. 13: Datenschutz beim Kredit scoring von Auskunftseien – Eine Untersuchung zum Profiling (mit hohem Risiko) und zur automatisierten Entscheidung**
SIMON HENSELER – Zürich 2025

Die rasanten technologischen Entwicklungen im Bereich der künstlichen Intelligenz (KI) treiben die digitale Transformation voran und stellen das Recht vor grundlegende neue Herausforderungen. Der vorliegende Band vereint neu verfasste sowie in den vergangenen Jahren entstandene Beiträge, die sich mit der rechtlichen Rahmenordnung von KI-Systemen und KI-gestützten Prozessen auseinandersetzen. Neben vertieften Analysen ausgewählter sektorieller Rechtsgebiete rücken insbesondere übergreifende Fragen der Verankerung von KI-Regulierungen in einer humanistischen und menschenzentrierten Rechtsordnung in den Fokus.