www.jusletter.ch

Fabienne Graf / Liliane Obrecht / Soraya Weiner

Erste Erkenntnisse zu Transparenz, Diskriminierung und Manipulation

Rechtliche Rahmenbedingungen für Künstliche Intelligenz in der Schweiz

Algorithmische Systeme, häufig als «Künstliche Intelligenz» bezeichnet, nehmen immer grösseren Einfluss auf die Gesellschaft, den Alltag und individuelle Rechte. An international voranschreitenden Regulierungsbestrebungen ist die Schweiz im Rahmen des Rechtsetzungsvorhabens des Europarates beteiligt. Die Veranstaltung bot Einblicke in den «Zero Draft» der Europaratskonvention sowie in erste fundierte Erkenntnisse des laufenden Forschungsprojekts zur Transparenz, Diskriminierung und Manipulation. Den zweiten Schwerpunkt der Veranstaltung bildete eine Paneldiskussion mit verschiedenen Interessengruppen.

Beitragsart: Tagungsberichte
Rechtsgebiete: Informatik und Recht, Europäisches Privatrecht, Verwaltungsrecht, Europäisches Wirtschaftsrecht

Zitiervorschlag: Fabienne Graf / Liliane Obrecht / Soraya Weiner, Erste Erkenntnisse zu Transparenz, Diskriminierung und Manipulation, in: Jusletter 12. Dezember 2022

Inhaltsübersicht

1. Rahmen der Veranstaltung

[1] Am 8. November 2022 fand an der Universität Zürich die zweite öffentliche Veranstaltung des Projekts «Nachvollziehbare Algorithmen: ein Rechtsrahmen für den Einsatz von Künstlicher Intelligenz»[1] (nachfolgend: KI-Forschungsprojekt) statt. Dieses steht unter der Leitung von Prof. Dr. Nadja Braun Binder (Universität Basel) und Prof. Dr. Florent Thouvenin (Universität Zürich, Center for Information Technology, Society and Law, ITSL). Gefördert wird das KI-Forschungsprojekt von der Stiftung Mercator Schweiz,[2] die Projektlaufzeit beträgt drei Jahre (2021–2024). Den Auftakt bildete die Kick-Off-Veranstaltung im November 2021.[3] Die vorliegend besprochene zweite Veranstaltung widmete sich ersten Ergebnissen. Inhaltlich bildeten aus der Perspektive des Privatrechts Fragen zur Diskriminierung und Manipulation den Schwerpunkt, aus Sicht der öffentlichen Verwaltung die Schaffung breit angelegter, gesellschaftlicher Transparenz. Ergänzend kam eine internationale Perspektive hinzu, nachdem der Bundesrat kürzlich die Mitwirkung der Schweiz an den Verhandlungen des Europarats über einen verbindlichen Rechtsrahmen für die Anwendung von Künstlicher Intelligenz (KI) beschlossen hatte.[4]

2. Aus dem Europarat

[2] Botschafter Thomas Schneider (Leiter Dienst Internationales des BAKOM, Vorsitzender des Committee on Artificial Intelligence (CAI) des Europarates[5]) gab Einblick in die Verhandlungen des Europarats für eine Konvention zur Regulierung von KI. Das CAI habe den Auftrag, bis Ende 2023 einen über die EU bzw. Europa hinausgehenden geeigneten Rechtsrahmen zu schaffen.

1 Weitergehende Informationen unter https://ius.unibas.ch/de/e-piaf/nachvollziehbare-algorithmen/ und https://www.itsl.uzh.ch/de/Forschung-und-Beratung/Forschungsprojekte.html (alle Webseiten zuletzt besucht am 10. November 2022).

2 Weiterführende Informationen zur Stiftung Mercator Schweiz unter https://www.stiftung-mercator.ch/.

3 Siehe dazu Fabienne Graf/Liliane Obrecht, Rechtliche Rahmenbedingungen für Künstliche Intelligenz in der Schweiz: Tagungsbericht der Veranstaltung «Rechtliche Rahmenbedingungen für Künstliche Intelligenz in der Schweiz», in: Jusletter 29. November 2021.

4 Siehe https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-90367.html.

5 Dieses ist Nachfolgerin des Ad-hoc Committee CAHAI, das von 2019–2021 agiert. Weitere Informationen zum CAI unter https://www.coe.int/en/web/artificial-intelligence/cai.

Dazu entwickelte es im Juni 2022 einen «Zero Draft» für eine (Rahmen-)Konvention[6] zu KI. Es werde sich dabei um ein bindendes Instrument handeln, das Innovation fördern und gleichzeitig Menschenrechte, Demokratie und Rechtsstaatlichkeit hochhalten solle. Im Unterschied zum Verordnungsentwurf der EU-Kommission über Künstliche Intelligenz (EU-KI-VOE)[7] sei die Konvention kein Marktregulierungsinstrument, sondern eines, das eine Reihe von Grundprinzipien im Umgang mit KI festhalten solle.

[3] Der «Zero Draft» bestehe aus einer Präambel sowie fünf Kapiteln. In der Präambel würden Chancen und Risiken der KI genannt sowie auf relevante bestehende Rechtsinstrumente und Arbeiten anderer internationaler Organisationen im Bereich KI verwiesen. Das erste Kapitel enthalte allgemeine Bestimmungen wie Ziel und Zweck der Konvention, Definitionen, den Geltungsbereich, den Grundsatz der Nichtdiskriminierung sowie die Verpflichtung der Vertragsparteien, die Konvention auf die Gestaltung, Entwicklung und Anwendung von KI-Systemen während ihres gesamten Lebenszyklus anzuwenden. Über die breite Definition von KI werde sicherlich noch ausführlich diskutiert. Vom Anwendungsbereich ausgenommen sei die nationale Verteidigung, da der Europarat kein Mandat dafür habe. Die Konvention richte sich an Staaten, solle aber die Anwendung von KI durch private sowie öffentliche Akteurinnen und Akteure abdecken. Das zweite Kapitel stelle verschiedene Grundprinzipien wie die Vereinbarkeit mit Grundwerten demokratischer Gesellschaften, Verantwortung und Transparenz auf. Das dritte Kapitel führe eine Risiko- und Folgenabschätzung sowie die Möglichkeit von Moratorien oder Verboten eindeutig schädlicher Anwendungen ein. Die Umsetzungsmechanismen und die Anforderungen an Kooperationen zwischen Mitgliedstaaten seien in einem vierten Kapitel verankert. Das fünfte Kapitel enthalte Schlussbestimmungen.

[4] Mindestens gleich bedeutsam wie die (Rahmen-)Konvention sei gemäss Schneider die Methodologie ihrer nationalen Umsetzung. Sie müsse die Gesetzgebung der verschiedenen Staaten berücksichtigen und gleichzeitig Kohärenz schaffen. Dafür stelle das CAI auf die «HUDERIA»[8]-Methodologie ab. Es gehe um die Definition klarer, konkreter und objektiver Kriterien zur Ermittlung von Kontexten und Anwendungen, bei denen der Einsatz von KI-Systemen (oder Kombinationen solcher Systeme) wahrscheinlich erhebliche Risiken bergen würde. Ausserdem solle ein einheitliches Vorgehen bei der Ermittlung, Analyse und Bewertung dieser Risiken sowie der Bewertung der Auswirkungen solcher Systeme gewährleistet werden. So würden die nationalen Behörden bei der Festlegung der Verfahren und Mechanismen unterstützt. Sichergestellt würde die Einführung angemessener Protokolle für die Risikoanalyse, Folgenabschätzung, Folgenminderung, den Zugang zu Rechtsmitteln und eine Systemüberwachung. Zentral sei auch die Gewährleistung der Kompatibilität dieses Ansatzes mit den bestehenden Bewertungs-, Kontroll- und Einhaltungspraktiken der Industrie. Die Implementierung wäre praxisorientiert und dynamisch vorzunehmen. HUDERIA würde die Begleitung und iterative Überprüfung von Anwendungen während des gesamten Lebenszyklus sicherstellen. Das Ziel, vor Ende 2023 eine solche Konvention verhandelt zu haben, sei sehr ambitiös. Ob dies gelingen wird, hänge unter anderem

6 Noch unklar sei die alternative Bezeichnung als Konvention oder Rahmenkonvention.

7 Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz und zur Änderung bestimmter Rechtsakte der Union (SEC[2021] 167 final).

8 HUDERIA steht für «human rights, democracy and rule of law impact assessments».

davon ab, wie schnell sich die 27 EU-Mitgliedstaaten und die EU-Kommission darauf einigen, wie und mit welchen Positionen sie in den Verhandlungen zur Konvention teilnehmen sollen.

3. Erste Erkenntnisse aus dem Forschungsprojekt

[5] Zu Beginn des KI-Forschungsprojekts wurden fünf Bereiche identifiziert, die als Herausforderungen beim Einsatz von algorithmischen Systemen rechtlich erfasst werden sollten: Transparenz (i.S.v. Erkennbarkeit und Nachvollziehbarkeit), Diskriminierung, Manipulation, Schutz der Privatsphäre bzw. Datenschutz (inklusive Datensicherheit) sowie Haftung.[9] Im Verlauf des KI-Forschungsprojekts werden Erkenntnisse in White Papers und detailliert in wissenschaftlichen Aufsätzen publiziert. Gegenstand der Veranstaltung waren Teilbereiche der drei erstgenannten Herausforderungen.

3.1. Transparenz durch ADM-Register?

[6] Prof. Dr. Nadja Braun Binder betonte, dass die Digitalisierung und insbesondere der Einsatz von Algorithmen dem Staat und seiner Gesellschaft viel Mehrwert böten. Von verschiedenen Seiten, namentlich der Politik und zivilgesellschaftlicher Organisationen, würden derzeit die Forderungen nach Transparenz über den KI-Einsatz im Staat immer deutlicher. Der Staat hingegen warte mit einer Regulierung auf Bundesebene bis anhin zu.

[7] Aus rechtswissenschaftlicher Perspektive stelle sich als erstes die Frage, wie die rechtliche Herstellung von Transparenz zu begründen sei. Die besondere Vertrauensbeziehung, die zwischen Staat und Bevölkerung bestehe, müsse auch im Kontext der Digitalisierung erhalten bleiben. Bei der öffentlichen Verwaltung gehe es dabei insbesondere um einen rechtsstaatskonformen und diskriminierungsfreien Einsatz von Algorithmen. Dabei könne die Tatsache, dass nicht bekannt sei, wo und wie KI von Behörden eingesetzt werde, zu grundsätzlichem Misstrauen der Bevölkerung gegenüber staatlichen Handlungen führen. Deshalb bedürfe es Mechanismen, um staatliches Handeln kontrollierbar zu machen. Transparenz könne eine individuelle Komponente enthalten, indem sie über eine individuelle Entscheidung hergestellt werde. Allerdings sei in einem gesamtgesellschaftlichen Kontext Transparenz darüber, welche algorithmischen Systeme der Staat überhaupt einsetze, ebenso wichtig. Ein öffentliches Register über den Einsatz algorithmischer Systeme könne dies verwirklichen. Es sollten aber nicht alle in der öffentlichen Verwaltung eingesetzten algorithmischen Systeme erfasst werden, sondern nur jene, die tatsächlich Auswirkungen auf die betroffene Person haben, indem sie ihre Rechte und Pflichten berühren. Gesprochen werde von algorithmenbasierten Entscheidungssystemen («automated decision making», ADM-Systeme). Ebenso sollten Systeme erfasst werden, deren Entscheidungsstruktur schwer bzw. nicht nachvollziehbar sei. Dazu zählten insbesondere Systeme, die gänzlich ohne Mitwirkung einer natürlichen Person funktionieren und solche, bei denen unklar sei, wie ein bestimmtes Ergebnis zustande kommt. Verschiedene Erlasse auf Bundesebene sähen bereits Informations- und Kommunikationspflichten der Regierungen und Verwaltungen vor (z.B. das Regierungs- und Verwaltungsorganisationsgesetz, RVOG). Diese liessen es bereits zu, ein solches Register (freiwillig) zu

[9] Florent Thouvenin et al., Ein Rechtsrahmen für Künstliche Intelligenz, Positionspapier November 2021, abrufbar unter https://www.dsi.uzh.ch/de/research/projects/strategy-lab/strategy-lab-21.html.

führen. Um Vollständigkeit und Vereinheitlichung zu schaffen, sei die verpflichtende Einführung von Registern jedoch zwingend. Das Kompetenznetzwerk des Bundes für KI (CNAI)[10] sehe bereits ein Register vor, es lägen jedoch nur vage Informationen darüber vor und für die Öffentlichkeit sei nicht voll ersichtlich, wo und welche KI-Systeme eingesetzt werden. Schliesslich böte ein öffentliches Register eine allgemeine, gesellschaftliche Kontrollmöglichkeit, rationales Staatshandeln sowie behördliche Selbstkontrolle zu fördern. Dabei müsse adressatengerechte Transparenz geschaffen werden, um den Zugang der breiten Öffentlichkeit zu ermöglichen.

3.2. Verhinderung von Diskriminierung durch KI

[8] Dr. Stephanie Volz zeigte auf, dass nicht nur Menschen, sondern auch algorithmische Systeme diskriminieren können. Dabei bestünden besondere Herausforderungen. Ihr Einsatz führe zu einer Skalierung, da sie darauf ausgerichtet seien, massenhaft Entscheidungen zu treffen und sich so ein einzelnes System auf eine Vielzahl von Personen auswirken könne. Dazu komme, dass bestehende Diskriminierungen beim Einsatz von algorithmischen (insbesondere selbstlernenden) Systemen aufgrund von Rückkoppelungseffekten («feedback loops») verstärkt werden können, wenn die Systeme mit Daten trainiert werden, die bereits Diskriminierungen enthielten und die Ergebnisse als Trainingsdaten in neue Systeme einflössen. Erschwerend wirke sich aus, dass sich Diskriminierung durch algorithmische Systeme oft nur schwer erkennen und nachweisen lasse. Algorithmische Systeme seien jedoch auch eine Chance zur Verhinderung von Diskriminierungen. Anders als bei menschlichen Entscheidungen könnten bestehende Diskriminierungen nicht nur erkannt, sondern durch eine Anpassung der algorithmischen Systeme auch behoben werden. Dies möge technisch zwar anspruchsvoll sein und nicht immer gelingen. Im Vergleich zum gesellschaftlichen Prozess, der erforderlich sei, um Diskriminierungen durch Menschen zu verhindern, erweise sich jedoch die Anpassung von Systemen ungleich einfacher, schneller und kostengünstiger.

[9] Im Vordergrund stehe bei algorithmischen Systemen die Proxy-Diskriminierung (Stellvertreterdiskriminierung). Dabei hänge eine Entscheidung von Daten ab, die neutral erschienen, faktisch aber in einer Diskriminierung aufgrund sensibler Merkmale resultierten. So könne bspw. die Zugehörigkeit zu einer Gruppe in Social Media als Proxy für das Geschlecht einer Person genutzt werden. Ursachen für Diskriminierungen fänden sich bei der Entwicklung dieser Systeme, in den Daten und der Anwendung.

[10] Die Schweizerische Bundesverfassung verbiete zwar Diskriminierung durch den Staat (Art. 8 Abs. 2 BV). Diskriminierung durch Private sei in der Schweiz bisher rechtlich nur fragmentarisch erfasst. Das breite Anwendungsfeld algorithmischer Systeme lasse dies zunehmend problematisch und eine rechtliche Erfassung notwendig erscheinen. Eine mögliche Lösung bilde die Einführung eines allgemeinen Gleichbehandlungsgesetzes, das bspw. auf dem bestehenden Behindertengleichstellungsgesetz aufbauen und sachlich auf Arbeits- oder Konsumverhältnisse beschränkt werden könnte. Weitere Optionen stellten die Statuierung eines neuen Grundsatzes der Datenbearbeitung im Datenschutzgesetz und die Erfassung der Diskriminierung als Persönlichkeitsverletzung über das allgemeine Persönlichkeitsrecht (Art. 28 ZGB) dar.

10 Weiterführend https://cnai.swiss/.

3.3. Verhinderung von Manipulation durch KI

[11] Prof. Dr. Florent Thouvenin stellte fest, dass überall dort, wo Menschen mit algorithmischen Systemen interagieren, die Befürchtung einer Beeinflussung des menschlichen Denkens und Handelns durch diese Systeme bestehe. Allerdings hätten sich noch keine konkreten Vorstellungen über das «Wie» des Einwirkens und dessen rechtlicher Regelung verfestigt. Hier bestehe noch grosser Bedarf nach einem öffentlichen Diskurs und rechtswissenschaftlicher Forschung. Drei Praxisbeispiele veranschaulichten, wie algorithmische Systeme in verschiedenster Weise auf Konsumentinnen und Konsumenten einwirken können: die Anreise mittels App-basierter Routenplanung; der Konsum personalisierter Inhalte auf Videoplattformen; das Nicht-Wählen eines Opt-outs von verrechneten Kosten im Interface Design eines Onlineshops. Die Beispiele liessen drei Grundfragen erkennen: Zum einen sei zu klären, welche Formen und welche Intensität des Einwirkens auf menschliches Denken und Handeln problematisch seien. Zweitens sei fraglich, ob es sich bei Manipulation um ein spezifisches Problem algorithmischer Systeme handle oder um allgemeine Phänomene der Beeinflussung bei der Interaktion mit Menschen oder Maschinen, bspw. durch die Ausgestaltung von User Interfaces. Drittens sei fraglich, ob nur dann eine rechtlich relevante Beeinflussung vorliege, wenn diese zum Nachteil der betroffenen Personen geschehe.

[12] Für die rechtliche Einordnung der Fragen dränge sich die Unterscheidung zwischen Willensbildung und Willensumsetzung auf. Die Willensbildung beruhe auf dem Verarbeiten von Information durch Menschen. Dieser Prozess könne namentlich durch das Verbreiten von falschen oder irreführenden Informationen beeinflusst werden. Die Verbreitung von Information sei schon im geltenden Recht dicht geregelt, bspw. durch das Irreführungsverbot des UWG (Art. 3 lit. b UWG) und verschiedene Bestimmungen des Straf- und Zivilgesetzbuchs. Zudem sei die Meinungs- und Informationsfreiheit grundrechtlich geschützt. Weitere Eingriffe auf Ebene der Informationsverbreitung seien deshalb nicht angezeigt. Anders sähe es bei der Willensumsetzung, als Proprium der Manipulation, aus: deren Beeinflussung sei in der Rechtsordnung noch kaum erfasst. Eine mögliche Umsetzung könne dabei nicht auf einen Vor- oder Nachteil abstellen. Vielmehr bilde die Autonomie der Menschen den Schutzgegenstand und rechtlichen Anknüpfungspunkt. Gemeint sei dabei nicht eine uneingeschränkte Autonomie. Das Beispiel eines kurzzeitig versperrten Fussgängerübergangs zeige die Schwelle zur minimal-invasiven Einwirkung. Im Ergebnis beinhalte die rechtliche Erfassung drei Elemente: (1.) ein Einwirken auf die Willensumsetzung; (2.) ein Auswirken auf das Handeln, wobei die Geeignetheit genügt; (3.) eine gewisse Intensität der Einwirkung. Zwei bestehende Generalklauseln bildeten die Werkzeuge, mit denen die Gerichte die vielfältigen Konstellationen bereits heute als Manipulation erfassen könnten. Für Art. 28 ZGB sei die Autonomie naheliegendes Persönlichkeitsrechtsgut. Ergänzend sichere Art. 2 UWG die Autonomie von Konsumentinnen und Konsumenten und das Funktionieren des Wettbewerbs.

4. Panel- und Plenumsdiskussion

[13] Im Panel waren verschiedene Interessengruppen vertreten: Dr. Lisa Bechtold (Zurich Versicherungen), Dr. Sophia Ding (AWK Group), Ivette Djonova (SWICO), Stephanie Gygax (Stiftung Mercator Schweiz), Matthias Mazenauer (Statistisches Amt Kanton Zürich), Botschafter

Thomas Schneider (BAKOM), Daniel Schönberger (Web 3 Foundation) und Dr. David Sommer (Digitale Gesellschaft).

4.1. KI-Regulierung in der Schweiz – Dringendste Probleme

[14] Zum Einstieg forderten Braun Binder und Thouvenin alle Teilnehmenden auf, die dringendsten Probleme beim Einsatz von KI und bei der rechtlichen Regelung in der Schweiz zu identifizieren. Bechtold bevorzugt Regulierungslösungen, die sich in den internationalen Rechtsrahmen einfügen und die Schweiz als «Digital Hub» vorantreiben. Djonova fügte an, der Schweizer Ansatz sei historisch gewachsen prinzipienbasiert. Als weitere Herausforderung nannte sie die Haftung; die neuen Ansätze der EU[11] regulierten eine Technologie, was innovationsfeindlich sei. Sommer äusserte sich als Informatiker zu den technischen Aspekten. Aus Sicht der Digitalen Gesellschaft gehe es bei der Diskussion nicht um KI im Allgemeinen, sondern um ADM-Systeme, die sich negativ auf das Leben von Menschen auswirkten. Wie von Volz betont, seien Skalierung und Feedback Loops ein Problem. Ding eruierte drei Kernprobleme: Erstens sei die Unterscheidung zwischen KI, herkömmlicher Software und dem Menschen als Anwenderin bzw. Anwender wichtig. Zweitens gäbe es viele Einsatzbereiche, bei denen keine Personendaten genutzt würden, weshalb die datenschutzrechtlichen Regelungen allein nicht reichen. Drittens sollten nicht nur die Herausforderungen neuer Technologien, sondern auch deren Chancen diskutiert werden. Mazenauer appellierte an die Befähigung der Menschen, autonome Entscheidungen zu treffen. Dies sei sinnvoller, als Manipulation als abstrakten Begriff zu regulieren. Für Schönberger und Gygax ist ein vernünftiges Gleichgewicht zwischen Grundrechtsschutz und Innovation zu finden. Schönberger sieht die Diskriminierungsthematik unabhängig von KI: es sei inakzeptabel, dass die Schweiz – im Gegensatz zum Ausland – im Privatrecht keine entsprechenden Normen kenne. Gygax betonte die Wichtigkeit des Einbezugs unterschiedlicher Interessengruppen, wie sie das KI-Forschungsprojekt zum Ziel habe. Schneider erinnerte daran, dass Manipulation nicht nur im Konsumbereich bestehe, sondern auch den Medienbereich betreffe. Thouvenin ergänzte, dass die Regulierung in diesem Bereich bereits weiter fortgeschritten und der Bereich der Medien ebenfalls Teil des KI-Forschungsprojekts sei.

4.2. Gesichtserkennungsverbot

[15] Die erste Frage aus dem Publikum betraf Verbote von Gesichtserkennungssoftware. Schneider verwies auf das von der EU-Kommission geplante Verbot. Die Europaratskonvention wolle jedoch möglichst keine konkreten Anwendungen benennen. Denn ob eine KI-Anwendung Gutes oder Schlechtes bewirken könne, hänge massgeblich vom Kontext ihres Einsatzes ab. So hätten Gesichts- oder Gefühlserkennungssysteme im Gesundheitsbereich andere Zwecke als in der Strafverfolgung. Braun Binder ergänzte, dass der staatliche Einsatz von Gesichtserkennungs-

11 COM(2022) 496 – Vorschlag für eine Richtlinie des Europäischen Parlaments und des Rates zur Anpassung der Vorschriften über außervertragliche zivilrechtliche Haftung an künstliche Intelligenz (Richtlinie über KI-Haftung) vom 28. September 2022 und COM(2022) 495 – Proposal for a directive of the European Parliament and of the Council on liability for defective products, ebenfalls vom 28. September 2022.

software nur erlaubt sei, wenn eine eindeutige, formell-gesetzliche Rechtsgrundlage bestehe.[12] Sommer betonte hingegen, dass Gesichtserkennung in der Schweiz bereits ohne Rechtsgrundlage durch den Staat (namentlich die Strafverfolgungsbehörden) eingesetzt wurde. Djonova und Schönberger sprachen sich für eine differenzierte Diskussion über die Definition der Gesichtserkennung bzw. Massenüberwachung aus. Gemäss Schönberger gehe es um Echtzeitüberwachung des öffentlichen Raums und nicht um die Entsperrung von Smartphone-Bildschirmen. Erstere solle sodann auch nicht für Strafverfolgungszwecke gestattet sein, da der Einsatz unverhältnismässig sei. Gemäss Thouvenin seien proaktive gesellschaftliche Entscheidungen nötig, um festzulegen, welche Technologien in welchem Kontext eingesetzt werden sollten. Sommer fügte an, dass es nicht nur um Gesichtserkennung, sondern um generelle Überwachung anhand biometrischer Attribute gehe (z. B. Tastatur-Tippmuster oder der Gang einer Person). Sommer und Mazenauer waren sich einig, dass es sich aufgrund der bereits genannten Skalierung um eine spezifische KI-Problematik handle. Damit die Gesellschaft der «selbstverschuldeten Unmündigkeit» entkommen könne, müsse sie aufgeklärt werden und Datenkompetenzen erlernen.

4.3. Risikobasierter Ansatz – auch für die KI-Regulierung?

[16] Die Frage nach dem zu wählenden Ansatz einer KI-Regulierung wurde kontrovers diskutiert. Gemäss Bechtold und Schönberger sei ein risikobasierter Ansatz konzeptionell wichtig für die Schweiz. So könne Innovation unter Wahrung ethischer Standards und der Grundrechte vorangetrieben werden. Zentral sei eine Risiko-Nutzen-Abwägung. Gemäss Schneider sei der risikobasierte Ansatz jedoch nur ein Element. Dieser müsse kontextbasiert, gleichzeitig aber juristisch durchsetzbar und nachvollziehbar sein. Als Ausgangspunkt sei zu definieren, für welche Personen wann überhaupt ein Risiko bestehe. Die Konvention des Europarats bilde ein Instrument, diesen Kontext zu definieren, indem das gesellschaftliche Risikoempfinden rechtlich verankert werde. Dies bedürfe einer gesellschaftlichen Debatte. Erst durch die demokratische Teilnahme werde die rechtliche Einhegung und somit die tatsächliche Anwendung und Durchsetzung einer neuen Technologie möglich. Bechtold betonte, dass Impact Assessment Tools wie HUDERIA bereits in verschiedensten Industrien bestünden, z.B. im Versicherungsbereich. Das Risiko bemesse sich an verschiedenen Faktoren (u.a., ob tatsächlich autonome Entscheidungen gefällt oder Empfehlungen abgegeben werden sowie an der Datensensitivität, der algorithmischen Komplexität oder der Nachvollziehbarkeit). Diese Analyse münde in die Frage nach einem verdeckten Diskriminierungspotenzial des Systems. Ausschlaggebend sei schliesslich der Anwendungskontext. Bechtold betonte, dass unproblematische Systeme (z.B. simple Chatbots) problemlos eingesetzt werden können sollten. Schönberger machte darauf aufmerksam, dass gegebenenfalls ein Anspruch auf den Einsatz von KI-Systemen bestünde, denn KI könne auch zur Aufdeckung von Diskriminierungen beitragen und Gleichberechtigung fördern.

12 Im Rahmen des KI-Forschungsprojekts wurde dies bereits untersucht, siehe Nadja Braun Binder/Eliane Kunz/Liliane Obrecht, Maschinelle Gesichtserkennung im öffentlichen Raum, in: sui generis 2022, S. 53 ff., open access abrufbar unter https://sui-generis.ch/article/view/sg.204/2576.

4.4. Konfliktpotenzial EU – Europarat?

[17] Schneider beantwortete die Frage nach einem potenziellen Konflikt zwischen der EU-Kommission und dem Europarat, bspw. betreffend des Individualrechtsschutzes. Die Mandate würden sich grundsätzlich unterscheiden. Der Europarat habe den Auftrag, Demokratie, Menschenrechte und Rechtsstaatlichkeit zu schützen. Während diese Bereiche traditionell in der Kompetenz der EU-Mitgliedstaaten lägen, komme der EU-Kommission die Aufgabe zu, den gemeinsamen Wirtschaftsraum zu regulieren. Allerdings sei in der EU eine Tendenz festzustellen, dass zunehmend über Marktregulierungsinstrumente auch menschenrechtliche und demokratische Aspekte reguliert werden. Denn dass alle Mitgliedstaaten die Werte teilten, sei grundlegend für einen gemeinsamen Wirtschaftsraum. Die Herausforderung läge also weniger zwischen EU und Europarat als zwischen der EU und ihren Mitgliedstaaten.

4.5. Abschliessende Statements

[18] Einige Diskussionsteilnehmende legten am Ende der Veranstaltung die für sie wichtigsten Aspekte einer KI-Regulierung in der Schweiz dar. Für Sommer ist ausschlaggebend, dass die Diskussionen nicht nur um Regulierung geführt werden, sondern dass auch die entsprechenden Entwicklerinnen und Entwickler der Machine Learning-Communities bei der Schaffung und Bereitstellung von Best-Practices und automatisierten Überprüfungsautomatisierungen unterstützt werden. Ding stimmte zu und betonte die Wichtigkeit der praktischen Umsetzbarkeit. Eine widersprechende Regelung zur EU sei keinesfalls erstrebenswert. Für Djonova ist die Transparenzdiskussion wichtig; sie müsse in jedem Fall adressatengerecht erfolgen. Schneider resümierte: Der internationale Einfluss, namentlich der EU-KI-VOE und eines kommenden US KI-Rechtsrahmens[13], müssten bei der Erarbeitung einer KI-Regulierung in der Schweiz so oder so berücksichtigt werden. Die Konvention des Europarates könne helfen, Interoperabilität der Systeme zu erreichen. Die Schweiz verfolge die Regulierungen und deren künftige Umsetzung in der EU und anderen Staaten aufmerksam, analysiere deren Auswirkungen auf die Schweiz und diskutiere mögliche eigene Regulierungsoptionen.

5. Würdigung und Ausblick

[19] Die Veranstaltung hat die Aktualität und Wichtigkeit der Diskussion um den Einsatz von KI in der Schweiz sowie deren rechtliche Erfassung gezeigt. Die Schweiz tut gut daran, die internationalen Regulierungsbestrebungen für eigene Regelungen zu berücksichtigen. Weiterhin noch offene Fragen bestehen insbesondere bezüglich der Haftung beim Einsatz algorithmischer Systeme. Ausserdem dürfen die technischen Aspekte und die Praktikabilität einer Lösung nicht ausser Acht gelassen werden. Konkrete Umsetzungsperspektiven zeigen sich für die öffentliche Verwaltung mit der Transparenz durch ADM-Register. Im Privatrecht besteht Handlungsbedarf bei der Diskriminierung. Dagegen bieten bestehende Generalklauseln eine direkte rechtliche Handhabe gegen Manipulation.

13 Blueprint for an AI Bill of Rights of The White House vom 4. Oktober 2022, weiterführend https://www.whitehouse.gov/ostp/ai-bill-of-rights/.

Fabienne Graf, MLaw, LL.M. (Duke), wissenschaftliche Mitarbeiterin am ITSL der Universität Zürich. Doktorandin an der Universität Luzern sowie der Humboldt-Universität zu Berlin.

Liliane Obrecht, MLaw, wissenschaftliche Mitarbeiterin und Doktorandin bei Prof. Dr. Nadja Braun Binder, Universität Basel.

Soraya Weiner, MLaw, wissenschaftliche Mitarbeiterin am ITSL der Universität Zürich und Doktorandin bei Prof. Dr. Florent Thouvenin, Universität Zürich.

Interessenbindung: Die Autorinnen sind wissenschaftliche Mitarbeiterinnen im Forschungsprojekt «Nachvollziehbare Algorithmen: ein Rechtsrahmen für Künstliche Intelligenz». Sie danken allen Referierenden und Teilnehmenden des Panels für die konstruktiven Rückmeldungen zum Text.