

Dokument	ZBJV 160/2024 S. 613
Autor	Florent Thouvenin, Stephanie Volz
Titel	Ein Rechtsrahmen für den Einsatz von Künstlicher Intelligenz (KI) in der Schweiz
Seiten	613-638
Publikation	Zeitschrift des Bernischen Juristenvereins
Herausgeber	Jörg Schmid, Frédéric Krauskopf
Frühere Herausgeber	Heinz Hausheer, Sibylle Hofer
ISSN	0044-2127
Verlag	Stämpfli Verlag AG

Ein Rechtsrahmen für den Einsatz von Künstlicher Intelligenz (KI) in der Schweiz

Florent Thouvenin*
Stephanie Volz**

1. Einleitung

Algorithmische Systeme, die auf Künstlicher Intelligenz (KI) beruhen, werden in einer Vielzahl von Lebensbereichen in unterschiedlichen Anwendungen verwendet, von der medizinischen Diagnostik über die Wartung von Maschinen in der Industrie bis hin zum Generieren von Texten, Bildern, Musik und Videos.

Mit Blick auf das Potenzial und die Risiken von KI befassen sich zahlreiche Staaten mit der Frage, wie sie auf diese technischen Entwicklungen reagieren sollen. Das Spektrum der Reaktionsweisen ist bemerkenswert breit. Um nur zwei besonders relevante Beispiele zu nennen: In den USA hat Präsident Biden am 30. Oktober 2023 eine «*Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*» erlassen, die eine Reihe von Massnahmen vorsieht, um die Entwicklung und Nutzung von KI zu fördern und auf der Grundlage von acht allgemeinen Prinzipien zu normieren.¹ Die EU hat schon im April 2021 einen ersten Entwurf für eine KI-Verordnung vorgelegt und nach intensiven Arbeiten an mehreren Versionen am 12. Juli 2024 die «Verordnung zur Festlegung

ZBJV 160/2024 S. 613, 614

* Prof. Dr. iur., Rechtsanwalt, Inhaber des Lehrstuhls für Informations- und Kommunikationsrecht und Vorsitzender des Leitungsausschusses des Center for Information Technology, Society, and Law (ITSL) der Universität Zürich.

** Dr. iur., Rechtsanwältin, Wissenschaftliche Geschäftsführerin des Center for Information Technology, Society, and Law (ITSL) der Universität Zürich. Die Autoren danken MLaw Deborah De Col, Rechtsanwältin, für die wertvolle Unterstützung bei der Recherche und beim Erstellen des Fussnotenapparats.

¹ Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, abrufbar unter: <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence> (abgerufen am 8. Oktober 2024).

harmonisierter Vorschriften für künstliche Intelligenz» (KI-VO)² verabschiedet. Diese regelt den Einsatz von KI, verbietet gewisse Verwendungen und sieht eine Reihe von Anforderungen für Hochrisiko-KI-Systeme vor. Auf internationaler Ebene hat der Europarat am 17. Mai 2024 das «Übereinkommen über künstliche Intelligenz (KI)» (KI-Konvention) verabschiedet.³ Die KI-Konvention schafft einen rechtsverbindlichen Rahmen für den Einsatz von KI und fokussiert dabei auf die Gewährleistung von Menschenrechten, Demokratie und Rechtsstaatlichkeit. Sie enthält allgemeine Grundsätze zu Transparenz, Robustheit, Nichtdiskriminierung und Schutz der Privatsphäre und lässt den Mitgliedstaaten damit viel Raum bei der Umsetzung im nationalen Recht.

Die Schweiz hat zwar schon recht früh, aber eher zurückhaltend reagiert. Im Dezember 2019 hat die interdepartementale Arbeitsgruppe «Künstliche Intelligenz» (IDAG KI) einen umfassenden Bericht zu den Herausforderungen von KI vorgelegt, in dem sie im Wesentlichen zum Schluss kam, dass kein unmittelbarer Handlungsbedarf bestehe.⁴ 2020 hat der Bund zudem Leitlinien für den Umgang mit KI in der Bundesverwaltung erlassen.⁵ Seit der Publikation des Berichts der IDAG KI haben verschiedene Parlamentarierinnen und Parlamentarier in zahlreichen Motionen, Interpellationen und Postulaten gefordert,

ZBJV 160/2024 S. 613, 615

dass der Bund KI regulieren oder Normen für bestimmte Herausforderungen erlassen solle.⁶ Vor diesem Hintergrund hat der Bundesrat am 22. November 2023 das Eidgenössische Departement für Umwelt, Verkehr, Energie und Kommunikation (UVEK) beauftragt, bis Ende 2024 einen Bericht über mögliche Regulierungsansätze für KI vorzulegen.⁷ Mit diesem Bericht will der Bundesrat eine Grundlage schaffen, damit er 2025 einen konkreten Auftrag für eine «Regulierungsvorlage» erteilen und die Zuständigkeiten regeln kann.

Wie der Schweizer Gesetzgeber auf die mit KI verbundenen Herausforderungen reagieren soll, wird derzeit in Wirtschaft und Wissenschaft, in der Zivilgesellschaft und in den Bundesbehörden intensiv diskutiert. Der Spielraum ist *sehr* gross, wie der Vergleich mit der EU, den USA und weiteren Ländern zeigt.⁸ Für die Schweiz stehen zwei Optionen im Vordergrund: ein autonomer Nachvollzug der KI-VO der EU oder die Anwendung sowie die punktuelle Revision und Ergänzung der bestehenden Gesetze.

2. Die KI-Verordnung der EU *in a nutshell*

Die KI-VO der EU reguliert KI als Technologie, indem sie eine Reihe von Vorgaben enthält, die beim Einsatz von KI-Systemen eingehalten werden müssen.⁹ Zu diesem Zweck hat die EU den Begriff des KI-Systems definiert. Nach Art. 3 Nr. 1 KI-VO handelt es sich dabei um «ein maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach

2 Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz).

3 Committee on Artificial Intelligence (CAI), Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law vom 17. Mai 2024, abrufbar unter: <https://rm.coe.int/1680afae3c> (abgerufen am 8. Oktober 2024).

4 Herausforderungen der künstlichen Intelligenz, Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat vom 13. Dezember 2019, abrufbar unter: https://www.sbfi.admin.ch/dam/sbfi/de/dokumente/2019/12/bericht_idag_ki.pdf.download.pdf/bericht_idag_ki_d.pdf (abgerufen am 8. Oktober 2024).

5 Leitlinien «Künstliche Intelligenz» für den Bund, Orientierungsrahmen für den Umgang mit künstlicher Intelligenz in der Bundesverwaltung, abrufbar unter: <https://www.sbfi.admin.ch/sbfi/de/home/bfi-politik/bfi-2021-2024/transversale-themen/digitalisierung-bfi/kuenstliche-intelligenz.html> (abgerufen am 8. Oktober 2024).

6 Für eine Übersicht zu den politischen Vorstössen auf Bundesebene siehe David Vasella, AI Act: FAQ zur KI-Verordnung, 67 f., abrufbar unter: <https://datenrecht.ch/wp-content/uploads/240922-FAQ-AIA-DE.pdf> (abgerufen am 8. Oktober 2024).

7 Medienmitteilung vom 22. November 2023 «Bundesrat prüft Regulierungsansätze für Künstliche Intelligenz», abrufbar unter: <https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-98791.html> (abgerufen am 8. Oktober 2024).

8 Siehe dazu bspw. die Übersicht auf dem «Digital Policy Alert», abrufbar unter: <https://digitalpolicyalert.org/threads/regulating-artificial-intelligence> (abgerufen am 8. Oktober 2024).

9 Siehe ausführlich zum Ganzen David Rosenthal, Der EU AI Act – Verordnung über künstliche Intelligenz, Jusletter vom 5. August 2024; Rolf H. Weber/Simon Henseler, Regulierung von Algorithmen in der EU und in der Schweiz, [EuZ 2020, 28–42](#), 28 ff.

ZBJV 160/2024 S. 613, 616

seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können». Die Definition von KI-Systemen wurde im Verlauf der Arbeiten an der KI-VO mehrfach grundlegend überarbeitet, nicht zuletzt als Reaktion auf die fortschreitende technische Entwicklung. Mittlerweile scheint sich international ein gewisser Konsens herausgebildet zu haben. Jedenfalls entspricht die Definition der KI-VO weitgehend derjenigen der OECD und, mit gewissen Einschränkungen, auch derjenigen der US-amerikanischen *Executive Order*.¹⁰

Die KI-VO lässt sich materiell als *Produktsicherheitsrecht* verstehen.¹¹ Sie beruht auf einem *risikobasierten Ansatz*, sieht also für verschiedene Arten und Anwendungsbereiche von KI unterschiedliche Regeln vor. KI-Systeme mit untragbarem Risiko sind grundsätzlich verboten. Darunter fallen bspw. der Einsatz von KI für *social scoring* und *predictive policing*, die Verwendung von KI-Systemen für die Emotionserkennung am Arbeitsplatz oder in Bildungseinrichtungen und die in Echtzeit erfolgende biometrische Fernidentifizierung von Menschen im öffentlichen Raum (Art. 5 KI-VO).

Den Kern der KI-VO bildet die Regelung von *Hochrisiko-KI-Systemen*. Der Begriff wird weit verstanden und erfasst bspw. KI-Systeme, die für die Bewertung von schulischen Leistungen, für das Filtern von Stellenbewerbungen, für die Prüfung der Kreditwürdigkeit oder als Sicherheitsbauteile von Produkten verwendet werden (Art. 6 i.V.m. Anhang II KI-VO). Für Hochrisiko-KI-Systeme gelten verschiedene Vorgaben, insbesondere zum Risikomanagement, zur Datenqualität und zur Daten-Governance, zur Protokollierung der Vorgänge und zum Erstellen technischer Dokumentationen (Art. 8 ff. KI-VO). Zudem müssen diese KI-Systeme während ihrer Verwendung von einem Menschen wirksam beaufsichtigt werden können (Art. 14).

ZBJV 160/2024 S. 613, 617

KI-VO) und ein angemessenes Mass an Robustheit, Genauigkeit und Cybersicherheit gewährleisten (Art. 15 KI-VO). Um sicherzustellen, dass diese Vorgaben eingehalten werden, müssen die Hersteller von Hochrisiko-KI-Systemen eine Konformitätsbewertung vornehmen (Art. 40 ff. KI-VO). Die Konformität wird dabei vermutet, wenn die KI-Systeme die Vorgaben der harmonisierten technischen Normen einhalten, die aufgrund von Normungsaufträgen der EU-Kommission erarbeitet werden (Art. 40 Abs. 1 KI-VO).

Für *KI-Systeme mit begrenzten Risiken* (bspw. Chatbots) sieht die KI-VO lediglich gewisse Transparenzpflichten vor; bspw. muss für natürliche Personen erkennbar sein, wenn sie mit einem Chatbot interagieren, und KI-generierte Audio-, Bild-, Video- und Textinhalte müssen als solche gekennzeichnet werden (Art. 50 KI-VO). Darüber hinaus enthält die KI-VO Vorschriften für KI-Modelle mit allgemeinem Verwendungszweck (*General Purpose AI [GPAI]*) (Art. 51 ff. KI-VO).

Die Durchsetzung der KI-VO erfolgt über einen eindrücklichen Behördenapparat: Jeder Mitgliedstaat der EU muss eine *notifizierende Behörde* und mindestens eine *Marktüberwachungsbehörde* einrichten, welche die Anwendung der Vorschriften beaufsichtigt und den Markt überwacht (Art. 70 KI-VO). In der EU-Kommission wird ein *Büro für Künstliche Intelligenz* (Art. 64 KI-VO) gebildet, das für die Durchführung der KI-VO auf EU-Ebene zuständig ist (Art. 75 und Art. 88 KI-VO). Zudem werden drei *Beratungsgremien* geschaffen, welche die einheitliche Anwendung und Durchsetzung der KI-VO durch wissenschaftliche und technische Beratung gewährleisten soll (Art. 65 ff. KI-VO). Die Durchsetzung soll ausserdem durch abschreckende *Sanktionen* sichergestellt werden: Beim Verstoss gegen verschiedene Vorgaben für Hochrisiko-KI-Systeme und bei der Verletzung der Transparenzvorschriften für KI-Systeme mit begrenztem Risiko können Anbieter, Importeure, Händler und Betreiber mit einer Geldbusse von bis zu 15 Mio. Euro oder 3% des gesamten weltweiten Jahresumsatzes bestraft werden, je nachdem, welcher Betrag höher ist (Art. 99 Abs. 4 KI-VO). Werden verbotene KI-Systeme in Verkehr gebracht, betrieben oder verwendet, drohen gar Geldbussen von bis zu 35 Mio. Euro oder 7% des gesamten weltweiten Jahresumsatzes, wobei auch hier der jeweils höhere Betrag massgebend ist.

ZBJV 160/2024 S. 613, 618

¹⁰ OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, 2024; in Sec. 3 (b) Executive Order (Fn. 1) fehlt namentlich das Merkmal, dass KI-Systeme einen unterschiedlichen Grad an Autonomie und Anpassungsfähigkeit aufweisen können.

¹¹ David Roth-Isigkeit, Der neue Rechtsrahmen für Künstliche Intelligenz in der Europäischen Union: Regelungsgegenstand, Lücken und Problemstellungen der neuen KI-Regulierung, KIR 2024, 15–20, 16.

3. Autonomer Nachvollzug der KI-VO?

3.1 Argumente pro und kontra

Für einen autonomen Nachvollzug der KI-VO spricht auf den ersten Blick, dass sich damit ein einfacher Zugang zum EU-Binnenmarkt erreichen liesse, namentlich durch die gegenseitige Anerkennung von Konformitätsbewertungen. Diese ist im «Abkommen zwischen der Schweizerischen Eidgenossenschaft und der Europäischen Gemeinschaft über die gegenseitige Anerkennung von Konformitätsbewertungen» vom 21. Juni 1999 (Mutual Recognition Agreement, MRA) geregelt. Das MRA Schweiz–EU sieht vor, dass schweizerische Konformitätsbewertungen von der EU anerkannt werden, wenn die schweizerischen Anforderungen als gleichwertig beurteilt werden. Dieser Mechanismus setzt allerdings voraus, dass der Geltungsbereich des MRA auf KI-Systeme erweitert wird. Nachdem das MRA Schweiz–EU aus politischen Gründen nicht mehr für alle Produktbereiche aktualisiert wurde, erscheint eine Erweiterung seines Geltungsbereichs auf KI-Systeme derzeit wenig wahrscheinlich. Die vage Hoffnung auf einen vereinfachten Zugang zum EU-Binnenmarkt ist deshalb kein überzeugendes Argument für einen autonomen Nachvollzug der KI-VO.

Gegen einen autonomen Nachvollzug der KI-VO sprechen mehrere Gründe. Das gewichtigste Argument ist die *fehlende Technologieneutralität*. Mit der KI-VO reguliert die EU KI als Technologie. Der EU-Gesetzgeber stellt sich zwar auf den Standpunkt, die Regelung sei technologieneutral, weil sie nicht zwischen verschiedenen Arten von KI unterscheide.¹² Das widerspricht allerdings dem gängigen Verständnis von «Technologieneutralität», nach dem der Begriff den Umstand bezeichnet, dass es bei der Anwendung und Auslegung einer Rechtsnorm keine Rolle spielt, ob und gegebenenfalls welche Technologie verwendet wird.¹³ Besonders problematisch ist die Regulie-

ZBJV 160/2024 S. 613, 619

rung einer Technologie, wenn diese laufend, allenfalls rasch oder gar grundlegend, weiterentwickelt wird. Gerade dies ist bei KI der Fall. An welchem Punkt der Entwicklungslinie wir uns derzeit befinden, kann zwar niemand mit Sicherheit sagen. Die meisten Experten gehen aber davon aus, dass die Entwicklung von KI noch am Anfang steht.¹⁴ Bei einer Regulierung in einem relativ frühen Zeitpunkt der Entwicklung besteht die Gefahr, dass sie die relevanten Probleme nicht erfasst, die weitere Entwicklung der Technologie behindert oder unvorhersehbare und unerwünschte Effekte erzielt. Das wurde schon während der Arbeiten an der KI-VO deutlich, als der bereits weit fortgeschrittene Entwurf infolge der Markteinführung von ChatGPT im November 2022 stark überarbeitet und ergänzt werden musste.

Gegen einen autonomen Nachvollzug der KI-VO sprechen auch der *grosse Aufwand und die hohen Kosten für die Compliance* und die *nachteiligen Auswirkungen auf Innovation und Marktstruktur*. Die KI-VO verbietet gewisse Praktiken im KI-Bereich (Art. 5 KI-VO) und enthält umfassende Vorgaben für Hochrisiko-KI-Systeme (Art. 8 ff. KI-VO). Das Einhalten dieser Vorgaben ist für die Anbieter und Betreiber von KI-Systemen mit grossem Aufwand und hohen Kosten verbunden. Zudem sind die Vorgaben teilweise unklar und es ergeben sich Widersprüche zu anderen Vorgaben der rasch anwachsenden und immer dichter werdenden «Digitalregulierung» der EU.¹⁵ Werden die Vorgaben der KI-VO nicht eingehalten, drohen Sanktionen in Form hoher Geldbussen. Zwar soll für KMU und Start-ups nicht der jeweils höhere, sondern der jeweils niedrigere Betrag der Bussen gelten. Dennoch sind

¹² Begründung 5.2.1. KI-VO; siehe dazu Angela Müller, Der Artificial Intelligence Act der EU: Ein risikobasierter Ansatz zur Regulierung von Künstlicher Intelligenz – mit Auswirkungen auf die Schweiz, EuZ 2022, A 1–A 25, A 5.

¹³ Zum Begriff der Technologieneutralität siehe Gabriele Gagliani, Cybersecurity, Technological Neutrality, and International Trade Law, Journal of International Economic Law 2020, 723–745, 731; Sherzod Shadikhodjaev, Technological Neutrality and Regulation of Digital Trade: How Far Can We Go?, EJIL 2021, 1221–1248, 1222.

¹⁴ Nestor Maslej et al., The AI Index 2023 Annual Report, Stanford University 2023, 73, abrufbar unter: https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf; Yoshua Bengio et al., International Scientific Report on the Safety of Advanced AI, DSIT Interim Report 2024, 29, abrufbar unter: https://assets.publishing.service.gov.uk/media/6655982fdc15efdddf1a842f/international_scientific_report_on_the_safety_of_advanced_ai_interim_report.pdf (beide abgerufen am 8. Oktober 2024).

¹⁵ Ein grundlegender Widerspruch besteht bspw. zwischen dem Verbot automatisierter Einzelentscheidungen gemäss Art. 22 DSGVO und dem Ansatz der KI-VO, die solche Entscheidungen (mit Ausnahme der Verbote in Art. 5 KI-VO) zulässt und lediglich gewisse Vorgaben für Hochrisiko-KI-Systeme vorsieht; siehe dazu Florent Thouvenin/Alfred Früh/Simon Henseler, Article 22 GDPR on Automated Individual Decision-Making: Prohibition or Data Subject Right?, EDPL 2022, 183–198, 192 f. Eingehend zur Problematik Alexander Golland, KI und KI-Verordnung aus datenschutzrechtlicher Sicht, EuZW 2024, 846–854, 850.

ZBJV 160/2024 S. 613, 620

die Sanktionen nicht nur, aber besonders für kleinere Anbieter derart bedrohlich, dass sie diese davon abhalten können, ihre KI-Systeme in der EU anzubieten oder zu verwenden. Im Ergebnis dürften die strengen Vorgaben für Hochrisiko-KI-Systeme und die hohen Geldbussen die Marktposition der grossen Anbieter (*Big Tech*) stärken und den Markteintritt von kleineren Unternehmen erschweren, weil die grossen Anbieter in der Lage sind, den erforderlichen Aufwand zu betreiben, um die Vorgaben der KI-VO einzuhalten, und sich gegen allfällige Sanktionen zu wehren oder die Bussen (als Teil des Geschäftsmodells) zu zahlen. Dieser Effekt lässt sich seit einigen Jahren bei der Datenschutzgrundverordnung (DSGVO) beobachten. Neu ist, dass nun selbst grosse Anbieter ankündigen, gewisse Dienste in der EU wegen der regulatorischen Hürden und der Unvorhersehbarkeit neuer Regulierungen erst gar nicht anzubieten.¹⁶

Als *Produktsicherheitsrecht* enthält die KI-VO Vorgaben, die gewährleisten sollen, dass sich die mit dem Einsatz von KI-Systemen verbundenen Risiken nicht verwirklichen. Im Gegensatz zu anderen Regelungen des Produktsicherheitsrechts enthält die KI-VO allerdings keine sektor- oder produktspezifischen Vorgaben (bspw. für Kinderspielzeug oder Medizinprodukte), sondern allgemeine Regeln, welche die Sicherheit verschiedenartiger Produkte beim Einsatz einer bestimmten Technologie gewährleisten sollen. Die horizontalen Vorgaben der KI-VO kommen dabei zu den bestehenden vertikalen (also: sektor- oder produktspezifischen) Regelungen der Produktsicherheit noch hinzu. Damit besteht die Gefahr von Doppelungen und Widersprüchen. Beides liesse sich vermeiden, wenn statt einer neuen horizontalen Regelung die bestehenden vertikalen Regelungen überprüft und bei Bedarf so angepasst würden, dass ihre Vorgaben die Sicherheit der Produkte auch bei der Verwendung von KI gewährleisten. Ihrem produktsicherheitsrechtlichen Ansatz entsprechend benennt die KI-VO zudem die Schäden nicht, deren Eintritt ihre Vorgaben verhindern sollen, und sie enthält – abgesehen von den nach Art. 5 KI-VO verbotenen Praktiken – auch keine Normen, die bestimmte Handlungen verbieten oder Betrof-

ZBJV 160/2024 S. 613, 621

fenen Anspruch auf Ersatz eines erlittenen Schadens gewähren. Die KI-VO vermag damit für sich allein den Eintritt von Schäden oder anderen rechtlich relevanten Nachteilen weder zu verhindern noch zu kompensieren. Vielmehr ist es ohne Weiteres möglich, alle Vorgaben der KI-VO einzuhalten und Betroffene beim Einsatz von KI-Systemen dennoch zu diskriminieren, zu manipulieren oder zu schädigen.

Damit ist klar, dass ein autonomer Nachvollzug der KI-VO allein nicht genügen würde. Vielmehr braucht es zusätzlich Anpassungen des Schweizer Rechts, um bspw. Diskriminierung durch KI-Systeme zu verhindern und sicherzustellen, dass Betroffene einen Schaden, der ihnen durch den Einsatz eines KI-Systems entstanden ist, erfolgreich geltend machen können.

3.2 Gründe der EU für den Erlass der KI-VO

Die kurze Analyse der wichtigsten Vor- und Nachteile eines autonomen Nachvollzugs der KI-VO wirft die Frage auf, weshalb die EU eine KI-VO erlassen hat. Dafür gibt es durchaus Gründe, die teils in der Kompetenzordnung und teils in der Regulierungstradition der EU und ihrer führenden Mitgliedstaaten liegen, teils aber auch politisch motiviert sind.

Die KI-VO stützt sich in erster Linie auf die Kompetenz zum Erlass von Vorschriften für die Errichtung und das Funktionieren des *Binnenmarktes* (Art. 114 AEUV) und zum Erlass von Vorschriften im Bereich des Datenschutzes (Art. 16 AEUV). Zentrales Ziel der KI-VO ist denn auch das Schaffen eines einheitlichen Rechtsrahmens für KI in der ganzen Union (Art. 1 Abs. 1 KI-VO). Mit dem Erlass der KI-VO hat die EU auch auf die Bestrebungen einzelner Mitgliedstaaten reagiert, nationale Vorschriften zur Regulierung von KI zu erlassen (ErwGr. 3 KI-VO). Da KI in einer Vielzahl von Anwendungsfeldern zum Einsatz kommt, in denen die EU keine Regelungskompetenz hat (z.B. im Bereich der Bildung oder der Strafverfolgung), eröffnet der Ansatz der Regulierung von KI als Technologie der EU die Möglichkeit, auch in Bereichen Einfluss zu nehmen, in denen sie eigentlich keine Regelungskompetenz hat.

Der Erlass der KI-VO hat auch politische Gründe: Die Präsidentin der Kommission, Ursula von der Leyen, hat in ihrer Antrittsrede im Europäischen Parlament am 27. November 2019 ausgeführt, dass

ZBJV 160/2024 S. 613, 622

¹⁶ Bspw. hat Facebook angekündigt, das multimodale KI-Modell LLaMa wegen «unvorhersehbaren regulatorischen Entwicklungen» in der EU nicht einzuführen, siehe dazu: <https://www.theguardian.com/technology/article/2024/jul/18/meta-release-advanced-ai-multimodal-llama-model-eu-facebook-owner> (abgerufen am 8. Oktober 2024).

die EU Schlüsseltechnologien wie KI beherrschen und «den Rahmen für die Welt» setzen müsse, wie sie das bereits mit der DSGVO im Datenschutzrecht getan habe.¹⁷ Damit war das Versprechen abgegeben, KI ebenso umfassend zu regulieren wie die Bearbeitung von Personendaten. Entsprechend war es politisch äusserst wichtig, dass noch vor Ablauf der Amtszeit der Kommission im Juli 2024 eine Regulierung von KI verabschiedet werden konnte. Für grundlegende Fragen nach Sinn und Zweck eines solchen Regulierungsansatzes blieb weder Zeit noch politischer Spielraum. Auch den wohl wichtigsten Grund für den Erlass der KI-VO hat die Präsidentin der Kommission schon in ihrer Antrittsrede klargestellt: Die EU will bei KI einen globalen Standard setzen, an dem sich alle anderen Staaten orientieren müssen (sogenannter *Brussels effect*).¹⁸ Dieses Ziel lässt sich nur durch eine Regulierung von KI als Technologie erreichen, also durch Vorgaben, die eingehalten werden müssen, wenn KI-Systeme in der EU in Verkehr gebracht oder betrieben (Art. 2 Abs. 1 Bst. a KI-VO) oder die Ausgaben dieser Systeme in der EU verwendet werden (Art. 2 Abs. 1 Bst. c KI-VO). Eine Regelung, die nur die Auswirkungen der Verwendung einer Technologie in der EU erfassen würde, hätte keinen solchen Effekt und könnte damit keinen globalen Standard setzen.

3.3 Erkenntnisse

Die Gründe, die in der EU zum Erlass der KI-VO geführt haben, sind für die Schweiz nicht relevant. Die kurze Analyse macht deutlich, dass eine Regulierung von KI als Technologie wenig Sinn macht – jedenfalls nicht für die Schweiz.

ZBJV 160/2024 S. 613, 623

Die Vorgaben der KI-VO sind zwar auch für Schweizer Unternehmen relevant, die KI-Systeme in der EU anbieten oder betreiben wollen oder deren KI-Systeme Ausgaben produzieren, die in der EU verwendet werden. Diese Unternehmen werden sich aber ohnehin und unabhängig von der Ausgestaltung des Schweizer Rechts an die Vorgaben der KI-VO halten müssen. Sollte der Schweizer Gesetzgeber mit der Entwicklung eines Rechtsrahmens für KI den Schweizer Unternehmen einen Standortvorteil gegenüber Unternehmen aus der EU verschaffen wollen, so lässt sich dies jedenfalls nicht durch einen autonomen Nachvollzug der KI-VO, sondern nur durch eine eigenständige Lösung erreichen. Diese sollte Innovation ermöglichen und fördern und gleichzeitig sicherstellen, dass niemand durch den Einsatz von KI-Systemen einen Schaden oder einen anderen rechtlich relevanten Nachteil erleidet. Dabei ist darauf zu achten, dass die Vorgaben des Schweizer Rechts denjenigen anderer Rechtsordnungen nicht widersprechen, dem EU-Recht ebenso wenig wie dem Recht anderer relevanter Märkte.

4. Anwendung und punktuelle Ergänzung des Schweizer Rechts

4.1 Grundsatz

Anstelle einer Regulierung von KI als Technologie sollte der Schweizer Gesetzgeber einen Ansatz verfolgen, der die konkreten Schäden und die anderweitigen rechtlich relevanten Nachteile erfasst, die beim Einsatz von KI-Systemen eintreten können. Im Vordergrund stehen dabei die Gefahr von Diskriminierung und Manipulation sowie Herausforderungen im Bereich des Datenschutzes, bei der Transparenz und bei der Haftung.¹⁹

Diese Herausforderungen sind als solche nicht neu, beim Einsatz von KI ergeben sich aber teilweise neue Probleme. Entsprechend sieht das Schweizer Recht bereits eine Vielzahl von (horizontalen) Normen

ZBJV 160/2024 S. 613, 624

vor, die angewendet werden können, um den Eintritt von Schäden oder anderweitigen Nachteilen zu verhindern oder einen allfälligen Schaden zu ersetzen, bspw. das allgemeine Persönlichkeitsrecht ([Art. 28 ZGB](#)), die Generalklausel des Bundesgesetzes gegen den unlauteren Wettbewerb ([Art. 2 UWG](#)), die

¹⁷ Rede der gewählten Kommissionspräsidentin von der Leyen im Europäischen Parlament anlässlich der Debatte zur Vorstellung des Kollegiums der Kommissionsmitglieder und seines Programms, 27. November 2019, abrufbar unter: https://ec.europa.eu/commission/presscorner/detail/de/speech_19_6408 (abgerufen am 8. Oktober 2024).

¹⁸ Marco Almada/Anca Radu, The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy, German Law Journal 2024, 1–18, 2; abrufbar unter: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4592006 siehe auch Charlotte Siegmund/Markus Anderljung, The Brussels Effect and Artificial Intelligence: How EU regulation will impact the global AI market, Centre for the Governance of AI Report 2022, 1–97, 22, abrufbar unter: <https://www.governance.ai/research-paper/brussels-effect-ai> (beide abgerufen am 8. Oktober 2024).

¹⁹ Siehe dazu Nadja Braun Binder/Thomas Burri/Melinda Florina Lohmann/Monika Simmler/Florent Thouvenin/Kerstin Noëlle Vokinger, Künstliche Intelligenz: Handlungsbedarf im Schweizer Recht, Jusletter vom 28. Juni 2021, Rn. 8 ff.

Bestimmungen des allgemeinen Deliktsrechts ([Art. 41 ff. OR](#)) und die Vorgaben des Datenschutzgesetzes ([DSG](#)). Diese Normen müssen allerdings mit Blick auf die Herausforderungen beim Einsatz von KI ausgelegt und angewendet sowie punktuell ergänzt werden. Neben diesen allgemeinen Regeln sind beim Einsatz von KI-Systemen auch die jeweiligen sektorspezifischen (vertikalen) Vorgaben einzuhalten, bspw. bei Medizinprodukten oder automatisierten Fahrzeugen. Diese Vorgaben müssen überprüft und gegebenenfalls angepasst werden, um sicherzustellen, dass die mit dem Einsatz von KI verbundenen Herausforderungen erfasst werden können.

Bei der Anwendung und den punktuellen Ergänzungen des Schweizer Rechts stellt sich eine Vielzahl von Fragen. Aus Platzgründen konzentriert sich dieser Beitrag auf die Problemfelder Diskriminierung, Datenschutz und Transparenz, bei denen der grösste Handlungsbedarf besteht. Die Ausführungen beschränken sich dabei auf die wichtigsten Aspekte. Für eine vertiefte Analyse und für die Frage der Manipulation verweisen wir auf eine Reihe von White Papers, die wir im Juni 2024 publiziert haben.²⁰

4.2 Diskriminierung

4.2.1 Problemstellung

a) Rechtliche Perspektive

Entscheidungen von KI-Systemen können Menschen diskriminieren. Eine rechtlich relevante Diskriminierung liegt vor, wenn Menschen aufgrund eines geschützten Merkmals (bspw. Alter, Geschlecht, Herkunft oder religiöse Überzeugung) qualifiziert ungleich behandelt und dadurch benachteiligt werden, ohne dass dafür ein

ZBJV 160/2024 S. 613, 625

sachlicher Grund vorliegt.²¹ Die Ungleichbehandlung muss zudem unverhältnismässig sein.²² Stützt sich eine Diskriminierung direkt auf ein geschütztes Merkmal, handelt es sich um eine direkte Diskriminierung. Eine indirekte Diskriminierung liegt nach dem Bundesgericht vor, wenn eine an sich neutrale Regelung, d.h. eine Regelung, die nicht an ein geschütztes Merkmal anknüpft, in ihren tatsächlichen Auswirkungen Angehörige einer geschützten Gruppe besonders benachteiligt.²³ Relevant war diese Form der Diskriminierung bisher allerdings nur in Fällen, in denen ein Geschlecht durch eine Regelung gegenüber einem anderen benachteiligt wurde, ohne dass dies sachlich begründet gewesen wäre, namentlich bei der schlechteren Entlohnung von sogenannten Frauenberufen.²⁴

b) Technische und tatsächliche Perspektive

Im Zusammenhang mit KI-Systemen sind zwei Formen von Diskriminierung besonders relevant: Im Vordergrund steht eine mögliche Diskriminierung durch *proxies* oder Stellvertretervariablen. Eine *Proxy*-Diskriminierung liegt vor, wenn eine Entscheidung von scheinbar neutralen Daten abhängt, im Ergebnis aber Personen mit geschützten Merkmalen diskriminiert werden, weil der *proxy* (oder die *proxies*) mit einem geschützten Merkmal mehr oder weniger stark korreliert.²⁵ Die Verwendung von KI-Systemen kann zudem zu einer intersektionalen Diskriminierung führen. Eine solche liegt vor, wenn

ZBJV 160/2024 S. 613, 626

²⁰ Die White Papers sind abrufbar unter: <https://www.itsl.uzh.ch/de/Wissenstransfer/Publikationen.html> (abgerufen am 8. Oktober 2024).

²¹ [BGE 149 I 248 E. 7.2 S. 276 f.](#); [BGE 147 I 73 E. 6.1 S. 81 f.](#); [BGE 139 I 169 E. 7.2.1 ff. S. 174 f.](#); Rainer J. Schweizer/Kim Fankhauser, in: Ehrenzeller/Egli/Hettich/Hongler/Schindler/Schmid/Schweizer (Hrsg.), Die schweizerische Bundesverfassung, St. Galler Kommentar, 4. Aufl., Zürich 2023 (zit. [BV](#) SGK-Autor/in), Art. 8 N 60.

²² [BGE 147 I 1 E. 5.2 S. 11 f.](#); [BGE 138 I 217 E. 3.3.5 S. 222 f.](#); [BGE 135 I 49 E. 6.1 S. 58 f.](#); Bernhard Waldmann, in: Waldmann/Belser/Epiney (Hrsg.), Bundesverfassung, Basler Kommentar, Basel 2015, Art. 8 N 87.

²³ [BGE 149 I 248 E. 7.2 S. 276 f.](#); [BGE 141 I 241 E. 4.3.2 S. 250 f.](#); [BGE 135 I 49 E. 4.1 S. 53 f.](#); dazu ausführlich: Bernhard Waldmann, Das Diskriminierungsverbot von [Art. 8 Abs. 2 BV](#) als besonderer Gleichheitssatz, Bern 2003, 338 ff.

²⁴ [BGE 124 II 409 E. 7 S. 424 f.](#) Für eine ausführliche Übersicht zur Rechtsprechung betreffend «Frauenberufe» siehe [BGE 141 II 411 E. 6.1.2 S. 419 f.](#)

²⁵ Sarah Legner, KI-Verordnung und algorithmische Diskriminierung, RD 2024, 393–460, 429; Gabriele Buchholtz/Martin Scheffel-Kain, Algorithmen und Proxy Discrimination in der Verwaltung: Vorschläge zur Wahrung digitaler Gleichheit, NVwZ 2022, 585–664, 613.

eine Person durch das Zusammenspiel mehrerer geschützter Merkmale benachteiligt wird.²⁶ Beim Einsatz von KI-Systemen ist es oft schwierig nachzuvollziehen, welches Merkmal für eine Entscheidung ausschlaggebend war, weil solche Systeme in aller Regel nicht auf ein einzelnes Merkmal, sondern auf eine Kombination von Merkmalen abstellen.

Ursachen für Diskriminierung finden sich in allen Phasen des Lebenszyklus von KI-Systemen, von der Entwicklung bis zur Anwendung.²⁷ Oft liegt die Ursache in den Daten, mit denen KI-Systeme trainiert werden. Enthalten die Trainingsdaten einen *bias*, wird auch das mit diesen Daten trainierte KI-System regelmässig diskriminierende Entscheidungen treffen. Ein häufiges Problem ist der *representation bias*, der entsteht, wenn nicht alle Personengruppen in einem Datensatz ausgewogen repräsentiert sind. Namentlich Minderheiten sind in Datensätzen oft untervertreten.²⁸ Diese ungleiche Verteilung führt dazu, dass KI-Systeme vor allem Muster aus stärker vertretenen Gruppen erlernen, während die Merkmale unterrepräsentierter Gruppen vernachlässigt werden.²⁹ Die Ursache für Diskriminierung kann aber auch in der Anwendung liegen, etwa wenn KI-Systeme in einem Kontext verwendet werden, für den sie nicht vorgesehen sind (*transfer context bias*).³⁰

ZBJV 160/2024 S. 613, 627

Diskriminierende Entscheidungen sind kein neues Phänomen, sie wurden bislang aber meist von Menschen getroffen. Beim Einsatz von KI-Systemen bestehen jedoch besondere Herausforderungen. KI-Systeme sind darauf ausgerichtet, eine Vielzahl von Entscheidungen zu treffen. Beim Einsatz solcher Systeme kommt es deshalb regelmässig zu einer Skalierung. Wenn ein KI-System diskriminierende Entscheidungen fällt, wirkt sich das regelmässig auf eine Vielzahl von Menschen aus.³¹ Durch Rückkoppelungseffekte (*feedback loops*) können zudem bestehende Diskriminierungen in KI-Systemen fortgesetzt oder gar verstärkt werden.³² Das ist namentlich der Fall, wenn der Output eines KI-Systems für das Training dieses oder eines anderen KI-Systems verwendet wird.

Zugleich bieten KI-Systeme die Chance, Diskriminierung zu bekämpfen. Zum einen kann die Skalierung dazu führen, dass sich Fälle von Diskriminierung erkennen lassen, die als Einzelfälle schwer festzustellen wären. Zum andern kann Diskriminierung durch KI-Systeme mithilfe von Tests statistisch besser nachgewiesen und durch eine Anpassung dieser Systeme auch einfacher korrigiert werden als bei menschlichen Entscheidungen.

4.2.2 Lösungsansätze

Diskriminierung durch KI-Systeme lässt sich rechtlich auf verschiedene Weise erfassen. Zwei Punkte sind aber stets zentral: Zum einen muss eine Regelung technologieneutral ausgestaltet werden, damit sie Diskriminierung durch Menschen ebenso erfassen kann wie Diskriminierung durch KI-Systeme oder andere (künftige) Technologien. Zum anderen sollte eine Regelung auch (und gerade) Diskriminierung durch Private erfassen. Anders als in den meisten umliegen-

²⁶ Zum Begriff der intersektionalen (mehrfachen) Diskriminierung siehe [BV](#) SGK-Schweizer/Fankhauser (Fn. 21), Art. 8 N 63; Sandra Hotz, Auf dem Weg zur Gleichstellung der Geschlechter?, [SJZ 2020, 3–18](#), 16 f.

²⁷ Mario Martini, Regulating Algorithms: How to Demystify the Alchemy of Code?, in: Ebers/Navas (Hrsg.), *Algorithms and Law*, Cambridge 2020, 100–134, 105 f.

²⁸ Johannes Kevekordes, KI im Überblick, in: Hoeren/Sieber/Holznagel (Hrsg.), *Handbuch Multimedia-Recht*, 60. Aufl., München 2024, N 38; Mario Martini, *Blackbox Algorithmus – Grundfragen einer Regulierung künstlicher Intelligenz*, Berlin 2019, 13.

²⁹ Xuanqi Gao/Chao Shen/Weipeng Jiang/Chenhao Lin/Qian Li/Qian Wang/Qi Li/Xiaohong Guan, Fairness in machine learning: definition, testing, debugging, and application, *Science China Information Sciences* 2024, 1–21, 6 ff.; siehe auch Janneke Gerards/Raphaële Xenidis, *Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law*, European Commission, European network of legal experts in gender equality and non-discrimination, Europäische Kommission, Brüssel 2020, 47.

³⁰ Moin Nadeem/Anna Bethke/Siva Reddy, StereoSet: Measuring stereotypical bias in pretrained language models, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* 2021, 5356–5371, 5356 f.; siehe auch Xavier Ferrer/Tom van Nuenen/José M. Such/Mark Coté/Natalia Criado, Bias and Discrimination in AI: A Cross-Disciplinary Perspective, *IEEE Technology and Society Magazine* 2020, 72–80, 72.

³¹ Siehe Anya E. R. Prince/Daniel Schwarcz, Proxy Discrimination in the Age of Artificial Intelligence and Big Data, *Iowa Law Review* 2020, 1257–1318, 1261 f.

³² Paul F. Langer/Jan C. Weyerer, Diskriminierungen und Verzerrungen durch Künstliche Intelligenz, Entstehung und Wirkung im gesellschaftlichen Kontext, in: Oswald/Borucki (Hrsg.), *Demokratietheorie im Zeitalter der Frühdigitalisierung*, Wiesbaden 2020, 223 f.; Carsten Orwat, *Diskriminierungsrisiken durch Verwendung von Algorithmen*, Antidiskriminierungsstelle des Bundes, Berlin 2019, 83, 89 f.

ZBJV 160/2024 S. 613, 628

den Ländern ist in der Schweiz nur Diskriminierung durch den Staat in allgemeiner Weise verboten ([Art. 8 Abs. 2 BV](#)). Für Private bestehen nur punktuelle Regeln, namentlich bei der Gleichstellung der Geschlechter und bei der Behindertengleichstellung. Oft sind es aber Private, die KI-Systeme für Entscheidungen einsetzen, bspw. bei der Gewährung von Krediten. Die Beschränkung des Diskriminierungsverbots auf staatliche Akteure ist deshalb zu eng. Vor diesem Hintergrund sind drei Lösungsansätze denkbar.

a) Gleichbehandlungsgesetz

Der umfassendste Ansatz besteht in der Schaffung eines allgemeinen Gleichbehandlungsgesetzes, das auf den bestehenden Regelungen im Gleichstellungsgesetz ([GIG](#)) und im Behindertengleichstellungsgesetz ([BehiG](#)) aufbauen könnte. Ein solches Gesetz müsste nicht nur ein allgemeines Diskriminierungsverbot enthalten, sondern auch Rechtfertigungsgründe und eine Erheblichkeitsschwelle vorsehen. Sofern politisch erwünscht, könnte der Anwendungsbereich auf besonders relevante Themenfelder wie Arbeit oder Konsum beschränkt werden. Die Durchsetzung liesse sich mit prozessualen Massnahmen wie Beweislast erleichterungen und klägerfreundlichen Regelungen zur Verteilung von Verfahrenskosten erleichtern.³³

b) Grundsatz im Datenschutzrecht

Alternativ könnte ein neuer Grundsatz der Datenbearbeitung im [DSG](#) eingeführt werden, der vorsieht, dass die Bearbeitung von Personendaten nicht zu einer Diskriminierung der betroffenen Person führen darf. Da eine Entscheidung über eine Person in aller Regel mit der Bearbeitung von Daten über diese Person verbunden ist, liesse sich Diskriminierung mit diesem Ansatz weitgehend erfassen. Wie bei jedem Verstoss gegen einen Grundsatz der Datenbearbeitung durch Private wäre auch hier eine Rechtfertigung möglich ([Art. 31 DSG](#)).

Für das Erfassen von Diskriminierung über das Datenschutzrecht müsste nicht zwingend ein neuer Grundsatz im [DSG](#) vorgesehen werden. Da das [DSG](#) die Grundrechte und die Persönlichkeit der be-

ZBJV 160/2024 S. 613, 629

troffenen Personen schützt ([Art. 1 DSG](#)), das Diskriminierungsverbot grundrechtlicher Natur ist und Diskriminierungen als Persönlichkeitsverletzungen qualifiziert werden können,³⁴ liesse sich Diskriminierung aufgrund der Bearbeitung von Personendaten schon heute als Verstoss gegen den datenschutzrechtlichen Grundsatz von Treu und Glauben ([Art. 6 Abs. 2 DSG](#)) erfassen. Angesichts der Bedeutung des Problems und des Bedarfs nach Rechtssicherheit wäre es allerdings sinnvoll, den Grundsatz der Nichtdiskriminierung ausdrücklich im [DSG](#) zu normieren.

Diese Lösung hätte verschiedene Vorteile: Da nur eine punktuelle Erweiterung des geltenden Rechts notwendig wäre, liesse sie sich gesetzgeberisch relativ einfach umsetzen. Zudem sieht das [DSG](#) griffige Durchsetzungsmechanismen und mit dem EDÖB auch eine etablierte Aufsichtsbehörde vor.

c) Diskriminierung als Persönlichkeitsverletzung

Eine dritte Möglichkeit besteht darin, Diskriminierungen als Verletzung des allgemeinen Persönlichkeitsrechts ([Art. 28 ZGB](#)) zu erfassen. Das [ZGB](#) definiert den Begriff der Persönlichkeit nicht, Lehre und Rechtsprechung haben aber im Verlauf der Zeit verschiedene rechtlich geschützte Aspekte der Persönlichkeit herausgearbeitet. Dazu zählen insbesondere die physische und die psychische Integrität, die Geheim- und die Privatsphäre sowie die Ehre.³⁵ Diese Aufzählung ist nicht abschliessend und unterliegt dem gesellschaftlichen Wandel.³⁶ Werden Entscheidungen über eine Person auf Grundlage eines geschützten Merkmals getroffen, kann dies ohne Weiteres als Verletzung der Persönlichkeit erfasst werden,

³³ Ebenso Isabelle Wildhaber/Melinda Florina Lohmann/Gabriel Kasper, Diskriminierung durch Algorithmen – Überlegungen zum schweizerischen Recht am Beispiel prädiktiver Analytik am Arbeitsplatz, ZSR 2019, 459–489, 470 f.

³⁴ Siehe dazu sogleich, 4.2.2 c).

³⁵ Statt vieler: Andreas Meili, Zivilgesetzbuch I, Basler Kommentar, 7. Aufl., Basel 2022 (zit. [ZGB](#) I BSK-Autor/in), Art. 28 N 16 ff. m. w. H.

³⁶ Andrea Büchler, Zivilgesetzbuch, Orell Füssli Kommentar, 4. Aufl., Zürich 2021, Art. 28 N 2.

zumal die geschützten Merkmale (bspw. Herkunft, Geschlecht, Sprache oder religiöse Überzeugung) die Persönlichkeit in aller Regel (mit-)definieren.³⁷

ZBJV 160/2024 S. 613, 630

4.3 Datenschutz

4.3.1 Problemstellung

Bei der Entwicklung und Verwendung von KI-Systemen werden regelmässig grosse Mengen an Daten bearbeitet. In vielen Fällen handelt es sich dabei um Personendaten, also um Angaben, die sich auf eine bestimmte oder bestimmbar natürliche Person beziehen ([Art. 5 lit. a DSG](#)). Die Bearbeitung von Personendaten ist umfassend durch das Datenschutzgesetz ([DSG](#)) geregelt. Die Vorgaben des [DSG](#) sind technologieneutral gefasst und gelten damit auch für Datenbearbeitungen durch KI-Systeme. Die mit dem Einsatz solcher Systeme verbundenen Herausforderungen für den Datenschutz lassen sich deshalb durch eine geeignete Auslegung und Anwendung und durch punktuelle Ergänzungen des [DSG](#) erfassen.

Datenschutzrechtliche Fragen stellen sich während des gesamten Lebenszyklus von KI-Systemen, von der Entwicklung über das Anbieten bis zur Verwendung. Besonders relevant ist die Frage, ob die Verwendung von Personendaten für das Training von KI-Systemen nach dem [DSG](#) zulässig ist und welche Vorgaben bei Entscheidungen, die von KI-Systemen gefällt werden, zu beachten sind.

4.3.2 Lösungsansätze

a) Training von KI-Systemen

Das Bearbeiten von Personendaten beim Training von KI-Systemen steht regelmässig in einem Spannungsverhältnis zu den Grundsätzen der Datenbearbeitung, insbesondere zum Grundsatz der Zweckbindung.³⁸ Nach diesem dürfen Personendaten nur zu einem bestimmten und für die betroffene Person erkennbaren Zweck beschafft und nur in einer Weise, die mit diesem Zweck vereinbar ist, bearbeitet werden ([Art. 6 Abs. 3 DSG](#)). Da die beim Training von KI-Systemen verwendeten Personendaten in aller Regel in einem anderen Kontext und für andere Zwecke beschafft wurden, stellt sich die

ZBJV 160/2024 S. 613, 631

Frage, ob eine solche Nutzung als Zweckänderung zu qualifizieren ist. Diese Frage kann nur im Einzelfall beantwortet werden. In manchen Fällen wird man einen Verstoß gegen den Grundsatz der Zweckbindung zwar verneinen können, bspw. wenn Röntgenbilder von Knochenbrüchen verwendet werden, um ein KI-System zu trainieren, das Knochenbrüche erkennen soll.³⁹ In vielen Fällen wird aber ein Verstoß vorliegen, etwa wenn Social-Media-Daten verwendet werden, um ein Gesichtserkennungssystem zu trainieren.

Ein Verstoß gegen einen Grundsatz der Datenbearbeitung kann gerechtfertigt werden, wenn die betroffene Person in die Bearbeitung einwilligt, ein überwiegendes privates oder öffentliches Interesse vorliegt oder eine gesetzliche Grundlage besteht ([Art. 31 Abs. 1 DSG](#)). Eine Rechtfertigung durch Einwilligung wird wegen der grossen Zahl der betroffenen Personen und der verwendeten Quellen (bspw. Social-Media-Posts) oft nicht möglich sein. Im Vordergrund steht deshalb die Rechtfertigung aufgrund überwiegender Interessen.⁴⁰ Das [DSG](#) nennt eine Reihe von Konstellationen, in denen ein überwiegendes Interesse des Verantwortlichen in Betracht kommt ([Art. 31 Abs. 2 DSG](#)), unter anderem die Datenbearbeitung für nicht personenbezogene Zwecke wie Forschung, Planung und Statistik ([Art. 31 Abs. 2 lit. e DSG](#)). Beim Training eines KI-Systems werden Personendaten nicht bearbeitet, um Erkenntnisse über die betroffenen Personen zu gewinnen oder Folgen für diese Personen abzuleiten, sondern um ein System zu entwickeln, das eine bestimmte Aufgabe (oder eine Vielzahl unbestimmter Aufgaben) erfüllen kann. Die Bearbeitung ist damit nicht personenbezogen und lässt sich grundsätzlich durch das überwiegende Interesse des Verantwortlichen rechtfertigen.

³⁷ Heinz Hausheer/Regina E. Aebi-Müller, Das Personenrecht des Schweizerischen Zivilgesetzbuches, 5. Aufl., Bern 2020, N 12.06; sinngemäss [ZGB](#) I BSK-Meili (Fn. 35), Art. 28 N 38 f.

³⁸ Lena Leffer/Maximilian Leicht, Datenschutzrechtliche Herausforderungen beim Einsatz von Trainingsdaten für KI-Systeme, Jusletter IT vom 24. Februar 2022, Rn. 92.

³⁹ Siehe dazu Samuel Klaus, KI trifft Datenschutz – Risiken und Lösungsansätze, in: Epiney/Rovelli (Hrsg.), Künstliche Intelligenz und Datenschutz / L'intelligence artificielle et protection des données, Zürich 2021, 81–95, 82.

⁴⁰ Daniel W. Seiler/Marcel Griesinger, Spannungsfeld Künstliche Intelligenz (KI) und Datenschutzrecht, Jusletter vom 25. September 2023, Rn. 10 ff.

Die Rechtfertigung der Bearbeitung von Personendaten für nicht personenbezogene Zwecke ist allerdings an weitere Voraussetzungen geknüpft: (i) Der Verantwortliche muss die Daten anonymisieren, sobald es der Zweck erlaubt. Ist eine Anonymisierung unmöglich oder erfordert sie einen unverhältnismässigen Aufwand, muss er angemessene Massnahmen ergreifen, um die Bestimmbarkeit zu

ZBJV 160/2024 S. 613, 632

verhindern ([Art. 31 Abs. 2 lit. e Ziff. 1 DSG](#)). (ii) Werden besonders schützenswerte Personendaten bearbeitet, darf der Verantwortliche diese Dritten nur so bekannt geben, dass die betroffenen Personen nicht bestimmbar sind. Ist dies nicht möglich, muss gewährleistet sein, dass die Dritten die Daten nur zu nicht personenbezogenen Zwecken bearbeiten ([Art. 31 Abs. 2 lit. e Ziff. 2 DSG](#)) (iii) Zudem dürfen die Ergebnisse nur so veröffentlicht werden, dass die betroffenen Personen nicht bestimmbar sind ([Art. 31 Abs. 2 lit. e Ziff. 3 DSG](#)). Während sich die ersten beiden Voraussetzungen in der Regel problemlos erfüllen lassen, hängt das Einhalten der dritten Vorgabe davon ab, ob man davon ausgeht, dass die Verwendung eines KI-Systems zu einer «Veröffentlichung» der Personendaten führt, wie teilweise vertreten wird.⁴¹ Die beim Training verwendeten Daten dürften sich allerdings in den meisten Fällen nur mit grossem Aufwand aus einem trainierten KI-System ermitteln lassen (*reverse engineering*), und es dürfte nach der allgemeinen Lebenserfahrung in den wenigsten Fällen zu erwarten sein, dass jemand diesen Aufwand auf sich nimmt.⁴² Damit wird in der Regel keine Veröffentlichung von Personendaten vorliegen.

Auch wenn die Bearbeitung von Personendaten beim Training von KI-Systemen schon im Rahmen des geltenden [DSG](#) regelmässig als zulässig zu qualifizieren sein wird, wäre es sinnvoll, wenn der Gesetzgeber die Frage ausdrücklich klären würde, indem er die Liste der Regelbeispiele für das Vorliegen eines überwiegenden Interesses ([Art. 31 Abs. 2 DSG](#)) um das Bearbeiten von Personendaten für das Training von KI-Systemen ergänzt. Alternativ könnte es genügen, wenn der Eidgenössische Datenschutz- und Öffentlichkeitsbeauftragte (EDÖB) in einem Leitfaden oder Merkblatt klarstellen würde, unter welchen Voraussetzungen das Bearbeiten von Personendaten beim Training von KI-Systemen nach seiner Rechtsauffassung in der Regel zulässig ist.⁴³

ZBJV 160/2024 S. 613, 633

b) Automatisierte Einzelentscheidungen

Entscheidungen von KI-Systemen sind in vielen Konstellationen als automatisierte Einzelentscheidungen zu qualifizieren. Eine solche liegt vor, wenn eine Entscheidung ausschliesslich auf der automatisierten Bearbeitung von Personendaten beruht und die Entscheidung für die betroffene Person mit einer Rechtsfolge verbunden ist oder sie erheblich beeinträchtigt ([Art. 21 Abs. 1 DSG](#)).

Wird eine Entscheidung vollautomatisiert gefällt, also ohne menschliche Überprüfung, muss der Verantwortliche die betroffene Person informieren ([Art. 21 Abs. 1 DSG](#)) und ihr auf Antrag die Möglichkeit geben, ihren Standpunkt darzulegen ([Art. 21 Abs. 2 DSG](#)). Zudem kann die betroffene Person verlangen, dass die automatisierte Einzelentscheidung von einer natürlichen Person überprüft wird ([Art. 21 Abs. 2 DSG](#)), soweit keine Ausnahme greift (siehe dazu [Art. 21 Abs. 3 DSG](#)), und sie kann vom Verantwortlichen gestützt auf das Auskunftsrecht Informationen über die Logik verlangen, auf der die Entscheidung beruht ([Art. 25 Abs. 2 lit. f DSG](#)).

Was diese Logik bei KI-Systemen genau bedeutet, ist allerdings unklar. Anders als bei der Verwendung von klassischer Software lassen sich Entscheidungen von KI-Systemen nicht ohne Weiteres nachvollziehen und erklären, weil der Weg vom Input zum Output nicht durch explizite Programmierung vorgegeben ist, sondern vom KI-System erlernt wird.⁴⁴ Je nach Komplexität und Architektur eines KI-Systems ist dieser Weg für Menschen nicht nachvollziehbar. Die betroffenen Personen können deshalb nicht verlangen, dass ihnen der Betreiber eines KI-Systems die für dessen Entscheidung kausalen Faktoren oder die verwendeten Algorithmen bekannt gibt. Der Betreiber muss die betroffenen Personen aber so weit informieren, dass diese die Funktionsweise des Systems in den Grundzügen verstehen und den Entscheidungsvorgang im Wesentlichen nachvollziehen können. Dazu muss er bspw. über die dem KI-System vorgegebenen Ziele und

⁴¹ David Rosenthal, Datenschutz und KI: Worauf in der Praxis zu achten ist, Jusletter IT vom 26. April 2022, Rn. 48 ff.; Rolf H. Weber, Künstliche Intelligenz und Datenschutz, Jusletter IT vom 4. Juli 2024, Rn. 20.

⁴² Seiler/Griesinger (Fn. 40), Rn. 22; sinngemäss auch Philip Glass, Datenschutzrecht für künstliche Intelligenz in der öffentlichen Verwaltung, in: Widmer (Hrsg.), Datenschutz – Rechtliche Schnittstellen, Zürich 2023, 177–255, 203 f.

⁴³ Ebenso Weber (Fn. 41), Rn. 22.

⁴⁴ Rolf H. Weber, Künstliche Intelligenz: Regulatorische Überlegungen zum «Wie» und «Was», EuZ 2022, B 1–B 18, B 12.

über die Datenkategorien informieren, mit denen das System trainiert wurde. Zudem sollten die Betroffenen darüber informiert werden, welche sie betreffenden Personendaten in die Entscheidung eingeflossen sind und wie diese Daten in etwa gewichtet wurden.

ZBJV 160/2024 S. 613, 634

Die besondere Informationspflicht ([Art. 21 DSGVO](#)) und das erweiterte Auskunftsrecht ([Art. 25 Abs. 2 lit. f DSGVO](#)) bestehen nach dem Wortlaut des [DSG](#) nur bei vollautomatisierten Entscheidungen. Diese Einschränkung wird zu Recht als problematisch erachtet, weil es für die betroffenen Personen kaum einen Unterschied macht, ob eine von einem KI-System getroffene Entscheidung noch von einem Menschen abgesegnet wird.⁴⁵ Vielmehr hat sich gezeigt, dass Menschen dazu neigen, von einer Maschine vorgeschlagene Entscheidungen zu bestätigen und nicht mehr zu hinterfragen, weil die Zeit für eine Überprüfung nicht reicht oder weil eine Abweichung von der Entscheidung der Maschine einer Begründung oder gar einer Rechtfertigung bedürfte. Im Ergebnis besteht deshalb oft kein Unterschied zwischen vollautomatisierten Entscheidungen und der Verwendung von sogenannten *decision support systems*. Zu Recht wurde deshalb schon bei der Totalrevision des [DSG](#) angeregt, den engen Begriff der vollautomatisierten Entscheidung im Zusammenhang mit der Verwendung von KI-Systemen zu erweitern.⁴⁶ Diese Erweiterung könnte und sollte bei den Anpassungen des Schweizer Rechts an die Herausforderungen von KI nachgeholt werden, bspw. durch Streichen des Wortes «ausschliesslich» in [Art. 21 Abs. 1 DSG](#).⁴⁷

4.4 Transparenz

4.4.1 Problemstellung

Die Gewährleistung der Transparenz von KI-Systemen ist eine Herausforderung, weil ein zentrales Merkmal vieler KI-Anwendungen darin besteht, dass die Modelle und Systeme keinen vorgegebenen Regeln

ZBJV 160/2024 S. 613, 635

folgen, sondern über das Erkennen von Mustern in Daten selbst Regeln entwickeln, um von einem Input zu einem Output zu gelangen. KI-Systeme sind damit inhärent intransparent, die Transparenz ist also nicht gegeben, sondern muss erst geschaffen werden. Mit dieser Frage befasst sich seit einiger Zeit die Forschungsrichtung der *Explainable AI*.⁴⁸

Aus rechtlicher Sicht ist Transparenz allerdings kein Selbstzweck. Vielmehr dient Transparenz in aller Regel einem anderen Zweck, bspw. dem Schutz vor Diskriminierung oder der Durchsetzung von Rechtsansprüchen.⁴⁹ Das Mass an Transparenz, das die Rechtsordnung sicherzustellen hat, muss sich deshalb am Zweck orientieren, der mit einer Transparenz fordernden Norm erreicht werden soll. Ausgehend vom Zweck lassen sich bei KI-Systemen zwei zentrale Dimensionen von Transparenz unterscheiden: Die *Erkennbarkeit* soll sicherstellen, dass Menschen bei der Interaktion mit einem KI-System (bspw. einem Chatbot) und beim Konsum KI-generierter Inhalte (bspw. einem Video) erkennen, dass sie mit einer Maschine interagieren oder von einer Maschine erstellte Inhalte konsumieren. Die *Erklär- und Nachvollziehbarkeit* betrifft dagegen die Frage, ob und in welchem Mass es für Menschen verständlich sein muss, wie KI-Systeme Entscheidungen treffen oder Inhalte erzeugen.⁵⁰ Hier geht es um die ungleich komplexere Frage nach der Transparenz der Funktionsweise von KI-Systemen, der zugrunde liegenden

⁴⁵ Nadja Braun Binder, Automatisierte Entscheidungen: Perspektive Datenschutzrecht und öffentliche Verwaltung, [SZW 2020, 27–34](#), 33 f.; Adrian Bieri/Julian Powell, in: Bieri/Powell (Hrsg.), Datenschutzgesetz, Orell Füssli Kommentar, Zürich 2023, Art. 21 N 6.

⁴⁶ Siehe dazu Vorstoss zur Revision des [DSG](#), SPK-N, 31.1./1. 2. 2019, und Beratung der SPK-N vom 23. Mai 2019.

⁴⁷ Siehe dazu die parlamentarische Initiative 23.438 vom 15. Juni 2023 «Datenschutzgesetzgebung anpassen und um teilautomatisierte Entscheide aufgrund künstlicher Intelligenz ergänzen» abrufbar unter: https://www.parlament.ch/centers/kb/Documents/2023/Kommissionsbericht_SPK-N_23.438_2024-06-28.pdf (abgerufen am 8. Oktober 2024).

⁴⁸ Für einen Überblick des Forschungsstandes siehe: Vikas Hassija et al., Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence, *Cognitive Computation* 2024, 45–74, 53 ff.

⁴⁹ John Zerilli/Alistair Knott/James Maclaurin/Colin Gavaghan, Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?, *Philosophy and Technology* 2019, 661–683, 676; Florent Thouvenin/Alfred Früh, Automatisierte Entscheidungen: Grundfragen aus der Perspektive des Privatrechts, [SZW 2020, 3–17](#), 10.

⁵⁰ Rolf H. Weber, Juristische Sprachmodelle zwischen Transparenz und Datensicherheit, in: Schweighofer/Eder/Costantini/Schmautzer/Pfister (Hrsg.), Tagungsband des 27. Internationalen Rechtsinformatik Symposiums IRIS 2024, 19–26, 21 f.; Nadja Braun Binder, Künstliche Intelligenz und automatisierte Entscheidungen in der öffentlichen Verwaltung, [SJZ 2019, 467–476](#), 472 f.

Algorithmen und der verwendeten Trainings- und Inputdaten. Diese Dimension der Transparenz ist aus (privat-)rechtlicher Sicht vor allem bei automatisierten Einzelentscheidungen relevant.⁵¹

ZBJV 160/2024 S. 613, 636

4.4.2 Lösungsansätze

a) Informationspflicht bei Interaktion mit KI-Systemen

Die Transparenz in Form der Erkennbarkeit sollte durch Einführung einer allgemeinen Informationspflicht sichergestellt werden, welche die Betreiber von KI-Systemen verpflichtet, die betroffenen Personen zu informieren, wenn sie mit einem solchen System interagieren. Die KI-VO sieht eine solche Pflicht vor (Art. 50 Abs. 1 KI-VO). Die betroffenen Personen müssen dabei in der Regel nicht ausdrücklich informiert werden, weil es genügt, wenn sie aus den Umständen erkennen können, dass sie mit einem KI-System und nicht mit einem Menschen interagieren. Die Erkennbarkeit lässt sich damit meist unkompliziert durch *transparency by design* sicherstellen, etwa indem Chatbots zu Beginn einer Konversation darauf hinweisen, dass sie keine Menschen sind.

Bei der Interaktion mit einem Menschen bearbeiten KI-Systeme in aller Regel Personendaten, womit die Bestimmungen des [DSG](#) zur Anwendung kommen. Dass die Betroffenen über den Einsatz eines KI-Systems zu informieren sind, lässt sich damit schon aus dem datenschutzrechtlichen Grundsatz der Transparenz herleiten.⁵² Um Rechtssicherheit zu schaffen, wäre es aber sinnvoll, eine Informationspflicht bei der Interaktion mit KI-Systemen ausdrücklich im [DSG](#) vorzusehen, bspw. im Kontext der bestehenden Informationspflicht bei automatisierten Einzelentscheidungen ([Art. 21 DSG](#)).

b) Kennzeichnungspflicht für KI-generierte Inhalte

Generative KI-Systeme können in Sekunden verschiedenartige Inhalte wie Texte, Bilder, Musik oder Videos erstellen. Bereits heute ist die Qualität dieser Inhalte so gut, dass sie sich oft kaum noch von Inhalten unterscheiden lassen, die von Menschen erstellt wurden. Aus heutiger Sicht scheint von KI-generierten Inhalten eine gewisse Täuschungsgefahr auszugehen, etwa bei *deep fakes*.⁵³ Die KI-VO der EU

ZBJV 160/2024 S. 613, 637

und die KI-Konvention des Europarates sehen deshalb eine Pflicht zur Kennzeichnung von KI-generierten Inhalten vor (Art. 50 Abs. 2 und Abs. 4 KI-VO; Art. 8 KI-Konvention).

Dieser Ansatz scheint allerdings wenig sinnvoll. Für viele Menschen dürfte es in vielen Konstellationen schon heute nicht mehr darauf ankommen, ob ein Inhalt von einem KI-System oder von einem Menschen erstellt wurde. Mit der zunehmenden Verbreitung von generativer KI werden wir ohnehin nicht mehr davon ausgehen (können), dass Bilder oder Videos die Realität wiedergeben. Eine Kennzeichnungspflicht würde zudem gerade dort nicht helfen, wo Transparenz besonders wichtig wäre, insbesondere bei *deep fakes*. Eine allgemeine Kennzeichnungspflicht für KI-generierte Inhalte bringt deshalb kaum einen Nutzen. Um die Gefahr von Täuschung und Manipulation zu verringern, sollte stattdessen sichergestellt werden, dass die Bevölkerung (weiter) über die Möglichkeiten generativer KI informiert und für die damit verbundenen Risiken sensibilisiert wird.

Anstelle einer allgemeinen Kennzeichnungspflicht könnten sektorspezifische Pflichten sinnvoll sein. Das gilt namentlich für die Medien, deren Inhalte von der Bevölkerung regelmässig als besonders vertrauenswürdig eingestuft werden. Allerdings haben sich viele Medien schon selbst Regeln zum Umgang mit generativer KI gegeben. Namentlich hat der Schweizer Presserat einen Leitfaden zu KI im Journalismus erlassen, nach dem mithilfe von KI erstellte Inhalte als solche zu kennzeichnen sind.⁵⁴ Da die Selbstregulierung durch den Presserat in der Schweiz bisher gut funktioniert hat, besteht hier kein gesetzgeberischer Handlungsbedarf.

⁵¹ Siehe dazu vorne, 4.3.2 b).

⁵² So auch der EDÖB, siehe Medienmitteilung vom 9. November 2023 «Geltendes Datenschutzgesetz ist auf KI direkt anwendbar», [abrufbar unter: https://www.edoeb.admin.ch/edoeb/de/home/kurzmeldungen/2023/20231109_ki_dsg.html](https://www.edoeb.admin.ch/edoeb/de/home/kurzmeldungen/2023/20231109_ki_dsg.html) (abgerufen am 8. Oktober 2024).

⁵³ Siehe dazu David Rosenthal, Datenschutz beim Einsatz generativer künstlicher Intelligenz, Jusletter vom 6. November 2023, Rn. 56.

⁵⁴ Presserat, Leitfaden «KI im Journalismus», abrufbar unter: https://presserat.ch/journalistenkodex/ki_leitfaden/ (abgerufen am 8. Oktober 2024).

5. Fazit

Der vorliegende Beitrag zeigt auf, wie der Schweizer Gesetzgeber auf die Herausforderungen reagieren kann (und sollte), die mit dem zunehmenden Einsatz von KI-Systemen verbunden sind. Zwei Optionen stehen im Vordergrund: ein autonomer Nachvollzug der KI-VO der EU oder die Anwendung sowie die punktuelle Revision und Ergänzung der bestehenden Gesetze.

ZBJV 160/2024 S. 613, 638

Ein autonomer Nachvollzug der KI-VO ist aus mehreren Gründen nicht sinnvoll: Zum einen reguliert die EU KI als Technologie. Angesichts der raschen und grundlegenden Weiterentwicklung von KI besteht bei diesem Ansatz die Gefahr, dass er die relevanten Probleme nicht erfasst und die weitere Entwicklung und Verwendung der Technologie behindert. Zum andern ist das Einhalten der Vorgaben der KI-VO mit grossem Aufwand und hohen Kosten verbunden. Werden die Vorgaben nicht eingehalten, drohen hohe Geldbussen. Im Gegensatz zu kleineren Unternehmen können grosse Anbieter (*Big Tech*) diese Kosten stemmen und die Bussen bezahlen. Im Ergebnis wird die KI-VO deshalb den Markteintritt von kleineren Unternehmen erschweren und sich wohl nachteilig auf den (Innovations-)Wettbewerb auswirken. Zudem führt die KI-VO zu Doppelungen mit dem bestehenden Produktsicherheitsrecht und kann den Eintritt von Schäden und anderweitigen rechtlich relevanten Nachteilen nicht verhindern.

Statt KI als Technologie zu regulieren, sollte der Schweizer Gesetzgeber durch geeignete technologie neutrale Massnahmen sicherstellen, dass niemand durch den Einsatz von KI-Systemen zu Schaden kommt oder einen anderen rechtlich relevanten Nachteil erleidet. Mit diesem Ansatz sollte Innovation ermöglicht und gefördert werden. Die beispielhaften Ausführungen zu Diskriminierung, Datenschutz und Transparenz haben gezeigt, dass die mit KI verbundenen Herausforderungen durch eine sinnvolle Auslegung und Anwendung sowie punktuelle Anpassungen und Ergänzungen des geltenden Rechts erfasst werden können. Die Anpassungen und Ergänzungen sollten in einem Rahmengesetz erlassen und (soweit möglich) in bestehende Gesetze integriert werden.⁵⁵ Auf diese Weise könnte (und sollte) ein sinnvoller und zukunftsfähiger Rechtsrahmen für den Einsatz von KI in der Schweiz geschaffen werden.

⁵⁵ Ebenso: Weber (Fn. 44), passim.