Jean Monnet Saar

EUROPARECHT ONLINE

Saar Blueprints

Elisabeth Hannah Woll

Menschenrechtsrisiken in KI-Systemen und der Balanceakt regulatorischer Maßnahmen:

Eine Analyse des Spannungsfelds zwischen Menschenrechtsförderung und -gefährdung am Beispiel der KI-Verordnung der EU und des Rahmenübereinkommens des Europarats

DFG Deutsche Forschungsgemeinschaft
German Research Foundation

03/2026 DE

Zur Autorin

Elisabeth Hannah Woll absolvierte einen LL.M. in European Integration and International Law an der Universität des Saarlandes sowie einen Bachelor of Arts in Philosophie und Soziologie an der Universität Trier und der Dalarna University, Schweden. Ihre Forschungsinteressen verbinden rechtliche und sozialwissenschaftliche Perspektiven: Im europäischen und internationalen Recht stehen IT-Recht, Asylrecht und der Schutz von Menschenrechten im Fokus. Daneben widmet sie sich gesellschaftlichen Fragestellungen von Rollenbildern und Armutsforschung, sowie erkenntnistheoretischen Ansätzen zum Verständnis von Wissen, Macht und sozialer Gerechtigkeit.

Vorwort

Diese Veröffentlichung ist Teil einer elektronischen Zeitschriftenserie (Saar Expert Papers), welche von Jean Monnet Saar, einem Lehrstuhlprojekt von Prof. Dr. Thomas Giegerich, LL.M. am Europa-Institut der Universität des Saarlandes herausgegeben wird. Die weiteren Titel der Serie können unter https://jean-monnet-saar.eu/ abgerufen werden.

In den Veröffentlichungen geäußerte Feststellungen und Meinungen sind ausschließlich jene der angegebenen Autoren und Autorinnen.

Herausgeber
Lehrstuhl Prof. Dr. Thomas Giegerich
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
Germany

ISSN
2199-0050 (Saar Blueprints)
DOI: 10.17176/20260327-152819-0

Zitierempfehlung
Woll, Menschenrechtsrisiken in KI-Systemen und der Balanceakt regulatorischer Maßnahmen: Eine Analyse des Spannungsfelds zwischen Menschenrechtsförderung und -gefährdung am Beispiel der KI-Verordnung der EU und des Rahmenübereinkommens des Europarats, Saar Blueprint 03/26, online verfügbar unter: https://jean-monnet-saar.eu/wp-content/uploads/2026/03/Elisabeth-Hannah-Woll-Menschenrechtsrisiken-in-KI-Systemen.pdf

Gefördert durch die **Deutsche Forschungsgemeinschaft** (DFG) – Projektnummer: 525576645

Inhaltsverzeichnis

A. Einleitung

Als die niederländische Regierung 2015 das auf Künstlicher Intelligenz (KI) basierende System *Risk Indication* („SyRI“) zur Aufdeckung von Sozialhilfebetrug einführte, klang die Idee zunächst vielversprechend: gezieltere Kontrolle bei geringeren Kosten. Doch das Experiment scheiterte, als das Bezirksgericht Den Haag 2020 den sofortigen Stopp des Programms anordnete, unter anderem weil es gegen die Europäische Menschenrechtskonvention, insbesondere Art. 8 zum Schutz der Privatsphäre, verstoße.[1] SyRI erlaubte es, umfangreiche personenbezogene Daten aus verschiedenen öffentlichen Stellen, u.a. über Einkommen, Wohnsituation, Arbeit, Schulden, Steuern oder Sozialleistungen, zu Risikoprofilen für möglichen Sozialhilfebetrug zu verknüpfen. Welche algorithmischen Methoden oder Entscheidungslogiken dafür konkret verwendet wurden, blieb allerdings völlig intransparent, und das System wurde nur in Regionen mit überdurchschnittlicher Armuts- oder Kriminalitätsrate und hohem Anteil an Sozialleistungsbeziehenden eingesetzt.[2] Diese digitale Diskriminierung wurde bereits vor dem Gerichtsentscheid von Bürgerrechtsorganisationen, NGOs und Vertretern der UN heftig kritisiert.[3]

Auch in der Allgemeingesellschaft wird zunehmend klar, dass KI weit mehr als nur ein neues Element der Digitalisierung oder gar reine technologische Spielerei ist, sondern eine riesige technologische Entwicklung, die sich rasant in gesellschaftliche Strukturen integriert.[4] Für viele wird der Umgang mit KI-Assistenzen immer alltäglicher, während die Systeme zunehmend „menschlicher“ werden: Aktuelle generative KI-Modelle können die Stimmungslage der NutzerInnen analysieren und ihre eigene Tonalität diesen anpassen, variieren Ausdruck und Betonung und erzeugen so den Eindruck von Nähe, Fürsorge oder Empathie.[5] KI-Systeme wirken also sowohl auf individuelle Lebensbereiche und Sozialräume, als auch auf makrosoziale Strukturen ein, etwa durch Veränderungen in Arbeitsorganisation, politischen Entscheidungsprozessen oder bestehenden Machtverhältnissen. In dieser Vielschichtigkeit des Einsatzes von KI zeigt sich ein grundlegendes Spannungsverhältnis, denn neben dem Fortschritt, den KI-Systeme versprechen, schaffen sie auch erhebliche Risiken. Dabei können die gleichen KI-Systeme zeitgleich Potenziale zur Förderung von Menschenrechten eröffnen,

[1] *Appelmann* et. al., JIPITEC 12/2021, S. 257, 258.
[2] *McGurk, Tomlinson*, S. 12; *Wieringa,* Hey SyRI, tell me about algorithmic accountability – Lessons from a landmark case, https://www.cambridge.org/core/services/aop-cambridge-core/content/view/22A3086554B0486BB4BBAF2D5A33A3D0/S2632324922000396a.pdf/hey-syri-tell-me-about-algorithmic-accountability-lessons-from-a-landmark-case.pdf (letzter Zugriff am 12.11.2025).
[3] *Appelmann* et. al., JIPITEC 12/2021, S. 257, 269.
[4] OECD Publishing, https://doi.org/10.1787/6b89dea3-de (letzter Zugriff am 02.11.2025).
[5] *Whittaker*, S.90ff.

während sie dieselben Rechte durch Überwachung, Diskriminierung oder technischen Fortschritt „um jeden Preis“ gefährden.[6]

Diese doppelte Perspektive spiegelt sich auch in den zentralen europäischen Regulierungsansätzen für KI wider. In der Präambel der Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz[7] heißt es einerseits, diese solle zur „Förderung der Werte der Union“ beitragen und Demokratie, Rechtsstaatlichkeit und Grundrechte stärken; andererseits müsse sie zugleich Innovation ermöglichen und die Wettbewerbsfähigkeit Europas sichern.[8] Noch deutlicher formuliert es der Europarat in seinem Rahmenübereinkommen über KI, Menschenrechte, Demokratie und Rechtsstaatlichkeit: KI könne „beispiellose Möglichkeiten bieten, Menschenrechte, Demokratie und Rechtsstaatlichkeit zu schützen und zu fördern“, zugleich aber auch „die Menschenwürde und individuelle Autonomie unterminieren“.[9]

Die vorliegende Arbeit untersucht dieses Spannungsfeld: Welche Chancen und Risiken KI für Menschenrechte birgt, wie Europa diese regulatorisch adressiert und wie sich die Wirksamkeit dieses Schutzes sowohl innerhalb der EU als auch im geopolitischen Kontext beurteilen lässt. Dies geschieht in drei aufeinander aufbauenden Schritten: Zunächst werden die zentralen Rechtsgrundlagen der europäischen KI-Regulierung vorgestellt – das Rahmenübereinkommen des Europarats zu KI, Menschenrechten, Demokratie und Rechtsstaatlichkeit sowie die EU-KI-Verordnung – ergänzt um eine geopolitische Einordnung globaler Regulierungsansätze. Anschließend werden menschenrechtsfördernde Potenziale und exemplarische Risikofelder von KI analysiert.

Im letzten Schritt soll die Wirksamkeit der europäischen Schutzmechanismen im Lichte der identifizierten Risikofelder bewertet werden, um ein Fazit über die Rolle insbesondere der EU in diesem hochdynamischen Spannungsfeld – sowohl geopolitisch als auch technologisch und angesichts der mehrdimensionalen Ansprüche an sie – ziehen zu können.

---

[6] *Hoffmann,* Demokratie und Künstliche Intelligenz, https://digid.jff.de/demokratie-und-ki/ (letzter Zugriff am 02.10.2025).

[7] Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024, ABl. L 2024/1689, 12.7.2024.

[8] *Europäische Union*, EU Artificial Intelligence Act, Das EU-Gesetz zur künstlichen Intelligenz, Aktuelle Entwicklungen und Analysen des EU AI-Gesetzes, 2024, https://artificialintelligenceact.eu/de/ (letzter Zugriff am 01.12.2025).

[9] *Council of Europe*, Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, angenommen am 17. Mai 2024, eröffnet zur Unterzeichnung am 5. September 2024, CETS Nr. 225, noch nicht in Kraft, https://rm.coe.int/1680afae3c (letzter Zugriff am 02.12.2025).

B. Rechtliche Grundlagen der KI-Regulierung in Europa

I. Das Rahmenübereinkommen über KI, Menschenrechte, Demokratie und Rechtsstaatlichkeit des Europarats

Seit den 1970er-Jahren beschäftigt sich der Europarat mit Datenschutz im Kontext von digitalen Technologien. Zentrale Meilensteine bilden die Budapester Konvention über Cyberkriminalität, angenommen am 23. November 2001 (ETS Nr. 185), sowie die Konvention Nr. 108 von 1981, die durch das Änderungsprotokoll von 2018 (CETS Nr. 223) modernisiert wurde.[10] Auf der Konferenz in Helsinki wurde im selben Jahr das Ad-hoc-Komitee des Europarats für Künstliche Intelligenz (CAHAI) gegründet, das Vorschläge für ein rechtsverbindliches KI-Rahmenwerk erarbeitete.[11] 2022 übernahm das Komitee für KI (CAI) diese Arbeit und führte sie mit dem Ziel fort, ein umfassendes Rahmenübereinkommen zu formulieren.[12] Dieses wurde im September 2024 schließlich zur Unterzeichnung aufgelegt.[13] In den Entwurf flossen zentrale internationale Rechtsinstrumente wie die EMRK und der ICCPR sowie Leitlinien der OECD und der UNESCO ein.[14]

Aktuelle Unterzeichner sind die EU und ihre Mitgliedstaaten sowie Andorra, Georgien, Island, Liechtenstein, Montenegro, Norwegen, Republik Moldau, San Marino, Schweiz, Ukraine, Kanada, Vereinigtes Königreich, Israel, Japan, USA und Uruguay.[15] Gemäß Art. 30 Abs. 3 tritt das Rahmenübereinkommen erst in Kraft, wenn mindestens fünf Staaten, darunter drei Mitgliedstaaten des Europarats, ratifiziert haben. Bislang hat kein Unterzeichnerstaat das Rahmenübereinkommen ratifiziert.[16]

Das Abkommen umfasst 36 Kapitel und orientiert sich gemäß Art. 7 bis 13 an sieben Kernprinzipien: Menschenwürde und individuelle Autonomie, Transparenz und Aufsicht, Rechenschaftspflicht und Reproduzierbarkeit, Gleichheit und Nichtdiskriminierung,

[10] *Levantino, Paolucci*, in: Czech/Heschl/Nowak et al. (Hrsg.), S. 3.
[11] *Council of Europe,* Members of the ad hoc Committee on Artificial Intelligence, https://rm.coe.int/list-of-cahai-members-web/16809e7f8d (letzter Zugriff am 02.10.2025).
[12] *Council of Europe*, CAHAI-Ad hoc Committee on Artificial Intelligence, https://www.coe.int/en/web/artificial-intelligence/caha (Übers. d. Verf., letzter Zugriff am 19.11.2025); *Nardocci*, IJPL 2024, S. 165, 182.
[13] *Council of Europe*, Committee on Artificial Intelligence (CAI), https://www.coe.in t/en/web/artificial-intelligence/cai (letzter Zugriff am 29.11.2025).
[14] *Council of Europe*, Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, https://rm.coe.int/1680afae67 (letzter Zugriff am 20.03.2026).
[15] *Council of Europe,* The Framework Convention on Artificial Intelligence, https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence (letzter Zugriff am 29.11.2025).
[16] Ebd.

Privatsphäre und Datenschutz, Zuverlässigkeit und sichere Innovation. Die ursprünglich geplanten umfassenderen Verpflichtungen der sogenannten HUDERIA-Prinzipien (*Human Rights, Democracy, and Rule of Law Principles for AI*) wurden letztendlich nicht übernommen.[17] Der Geltungsbereich des Rahmenübereinkommens erfasst indes laut Art. 3 sämtliche Aktivitäten innerhalb des Lebenszyklus von KI-Systemen (sog. Lifecycle-Ansatz), sofern diese potenziell mit menschenrechtlichen, demokratischen oder rechtsstaatlichen Prinzipien interferieren können.

Das Rahmenübereinkommen findet auf Forschungs- und Entwicklungsaktivitäten an KI-Systemen, die noch nicht zur Nutzung bereitgestellt wurden, keine Anwendung, außer wenn Tests oder ähnliche Tätigkeiten in einer Weise durchgeführt werden, die potenziell die Menschenrechte, Demokratie oder Rechtsstaatlichkeit beeinträchtigen könnte (Art. 3 Abs. 3). Ebenfalls erlaubt das Übereinkommen gemäß Art. 3 Abs. 2, dass eine Vertragspartei die Konvention nicht auf Tätigkeiten im Lebenszyklus von KI-Systemen anwenden muss, wenn diese dem Schutz ihrer nationalen Sicherheitsinteressen dienen, verlangt jedoch ausdrücklich, dass solche Tätigkeiten im Einklang mit internationalem Recht, den Menschenrechten und unter Achtung der demokratischen Institutionen der jeweiligen Vertragspartei erfolgen müssen.

Die Durchsetzung des Rahmenübereinkommens erfolgt nicht über Pflichten und mögliche Sanktionen, sondern über institutionellen Austausch, gegenseitige Rechenschaftspflicht und Kontrolle (Art. 23 ff.) und durch die freiwillige Selbstverpflichtungen der Vertragsparteien. Zentrales Organ des Rahmenübereinkommens ist die *Conference of the Parties*, in der die VertreterInnen der Vertragsstaaten zusammenkommen, um den multilateralen Austausch über Auslegung und Anwendung des Übereinkommens zu koordinieren (Art. 23). Sie bewertet technische und politische Entwicklungen, diskutiert dementsprechende mögliche Änderungen und regelt die Zusammenarbeit mit relevanten Interessengruppen; öffentliche Anhörungen können dabei als Instrument genutzt werden. Auch regelt die *Conference of the Parties* Form und Rhythmus der Berichtspflicht nach Art. 24. Die Vertragsparteien müssen innerhalb von zwei Jahren nach Inkrafttreten und anschließend in regelmäßigen Abständen über ihre Maßnahmen zur Umsetzung der Verpflichtungen Auskunft geben. Zur innerstaatlichen Umsetzung können unabhängige Aufsichtsmechanismen eingerichtet oder benannt werden, wobei den Staaten ein hoher Grad an Flexibilität in der Umsetzung eingeräumt wird.

[17] *Levantino, Paolucci*, in: Czech/Heschl/Nowak et al. (Hrsg.), S. 14; weiterführend zur HUDERIA-Idee *Woll*, ZEuS 3/2025, S. 509, 521 f.

Mechanismen wie die Berichtspflicht und die Einbindung nationaler Aufsichtsstellen dienen in erster Linie dem Monitoring und der Förderung der gegenseitigen Rechenschaft. Eine Koordination mit bestehenden Menschenrechtsinstitutionen ist ausdrücklich vorgesehen, liegt jedoch im Ermessen der Vertragsparteien (Art. 26). Sanktionsmechanismen oder spezifische Haftungsregelungen sind nicht vorgesehen; dementsprechend hängt die Wirksamkeit von den institutionellen und rechtsstaatlichen Kapazitäten und vom Willen der Vertragsstaaten ab.[18]

II. Die KI-Verordnung der Europäischen Union

Die EU begann ihre Auseinandersetzung mit Digitalisierung und Datenrecht etwas später als der Europarat, in den 1980er Jahren, zunächst tatsächlich mit der Aufforderung an ihre Mitgliedstaaten, die Konvention Nr. 108 des Europarats zu unterzeichnen und zu ratifizieren.[19] 1999 setzte der Aktionsplan „Sicheres Internet" dann erste Schwerpunkte auf die Sicherheit im digitalen Raum.[20] Mit der Datenschutz-Grundverordnung (DSGVO) übernahm die Union dann ab 2016 eine führende Rolle bei der Datenschutzgesetzgebung, und Standards der DSGVO wurden weltweit im Sinne des Brüssel-Effekts übernommen.[21]

Die ersten Schritte der KI-Regulierung begannen 2018 mit der Mitteilung „Künstliche Intelligenz für Europa" und dem „Koordinierten Plan für KI", letzterer sollte sektorübergreifende Maßnahmen und nationale KI-Strategien initiieren.[22] Zu dessen institutioneller Umsetzung wurde im Januar 2024 das Europäische Amt für KI eingerichtet.[23] Es unterstützt die Implementierung der KI-VO und überwacht die Umsetzung, auch hinsichtlich der Zusammenarbeit mit Unternehmen, indem es ihnen geschützte KI-Sandkästen zur Erprobung neuer Systeme bietet. Begleitend dazu arbeitet das *European Artificial Intelligence Board* seit August 2024 daran, eine einheitliche Anwendung der Verordnung sicherzustellen und strategische Empfehlungen zu geben.[24]

Die KI-Verordnung selbst wurde 2021 vorgeschlagen, im März 2024 offiziell verabschiedet und trat am 1. August 2024 in Kraft; doch die volle Anwendung folgt nach einer zweijährigen

[18] *Woll*, ZEuS 3/2025, S. 509, 526.
[19] *Streinz*, in: Craig/de Búrca (Hrsg.), S. 909 f.
[20] *Europäische Union*, Entscheidung Nr. 276/1999/EG.
[21] Verordnung (EU) 2016/679 vom 27.4.2016 (ABl. 2016 Nr. L 119/1; 2016 Nr. L 314/72; 2018 Nr. L 127/2; 2021 Nr. L 74/35); *Bradford*, S. XIV ff.
[22] *Europäische Kommission*, Koordinierter Plan für künstliche Intelligenz, https://digital-strategy.ec.europa.eu/de/policies/plan-ai (letzter Zugriff am 02.10.2025).
[23] *Deutscher Bundestag*, Europäisches AI Office soll Know-how im KI-Bereich bündeln, https://www.bundestag.de/presse/hib/kurzmeldungen-995858 (letzter Zugriff am 05.10.2025).
[24] *Europäische Kommission*, AI Board, https://digital-strategy.ec.europa.eu/en/polici es/ai-board (letzter Zugriff am 05.11.2025).

Übergangsfrist am 2. August 2026. Strategisch flankiert wird die Verordnung durch den *AI Continent Action Plan*, mit dem Ziel, Europa zu einem global führenden „KI-Kontinent" zu entwickeln. Im Fokus stehen technologische Autonomie durch den Aufbau leistungsfähiger KI-Infrastrukturen, ein vereinfachter Regulierungsrahmen sowie die Förderung von sicherer KI und KI-Kompetenzen, um damit Europas normative Glaubwürdigkeit in der globalen KI-Landschaft zu sichern.[25] Die Finanzierung erfolgt seit Februar 2025 über die Initiative *InvestAI*, durch welche bis zu 200 Milliarden Euro in den europäischen KI-Sektor fließen sollen.[26]

Die KI-VO gilt horizontal für sämtliche KI-Systeme, die auf dem Binnenmarkt in Verkehr gebracht werden, also unabhängig davon, ob sie innerhalb oder außerhalb der EU hergestellt wurden.[27] Eingebettet ist sie in das Konzept der Produktkonformität, die durch die CE-Kennzeichnung nachgewiesen wird.[28] Alle KI-Systeme müssen also die europäischen Sicherheits-, Gesundheits- und Umweltanforderungen erfüllen. Rechtsgrundlage sind Art. 114 AEUV zu Harmonisierungsmaßnahmen im Rahmen der digitalen Binnenmarktstrategie und Art. 16 AEUV bezüglich der Datenschutzvorgaben.[29] Als Verordnung nach Art. 288 Abs. 2 AEUV ist die KI-VO unmittelbar in allen Mitgliedstaaten gültig.

Das grundlegende Prinzip der KI-VO ist ein risikobasierter Ansatz, der KI-Systeme in vier Kategorien einteilt, welche jeweils unterschiedlichen Anforderungen unterliegen: Die erste Kategorie betrifft Systeme mit unvertretbarem Risiko (Art. 5), die eine ernsthafte Gefährdung grundlegender Menschenrechte darstellen, wie z.B. Systeme, die gezielt menschliches Verhalten manipulieren, oder „*Social Scoring*"-Systeme nach chinesischem Vorbild,[30] bei denen Menschen je nach Verhalten oder Meinung gesellschaftlich belohnt oder sanktioniert werden – diese sind verboten.[31]

---

[25] *Europäische Kommision*, Shaping Europe's leadership in artificial intelligence with the AI continent actionplan, https://commission.europa.eu/topics/eu-competitiveness/ai-continent_en (letzter Zugriff am 12.10.2025); *Cantero,* STG Policy Papers, S. 6; *EDSB Strategie*, https://www.edps.europa.eu/sites/default/files/publication/15-07-30_strategy_2015_2019_update_de.pdf. (letzter Zugriff am 02.11.2025); *Europäische Kommision*, Shaping Europe's Leadership in Artificial Intelligence with the AI Continent Actionplan, https://commission.europa.eu/topics/eu-competitiveness/ai-continent_en (letzter Zugriff am 02.10.2025).

[26] *Europäische Kommission*, EU startet InvestAI-Initiative zur Mobilisierung von 200 Mrd. EUR an Investitionen in künstliche Intelligenz, https://digital-strategy.ec.europa.eu/de/news/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence (letzter Zugriff am 02.11.2025).

[27] *Giegerich*, Saar Expert Papers 09/23, S. 15.

[28] Artikel 48 der Verordnung; *Europäische Kommission*, CE Marking, https://single-market-economy.ec.europa.eu/single-market/ce-marking_en?prefLang=de (letzter Zugriff am 02.11.2025); *McGurk, Tomlinson*, S. 37: Ergänzend bilden der *Digital Markets Act* und der *Digital Services Act* weitere regulatorische Bausteine des europäischen Digitalrahmens.

[29] *Nardocci,* IJPL 2024, S. 165, 174.

[30] Siehe dazu unten, S. 12.

[31] *Trüe*, in: Lisowski/Scholz/Trüe (Hrsg.), S. 175, 185.

Die zweite Kategorie betrifft Hochrisiko-KI (Art. 6) mit direkten Auswirkungen auf Sicherheit und Grundrechte, wie z. B. autonome Fahrzeuge, Einsätze in der medizinischen Diagnostik, oder Bewerbervorauswahlsysteme; diese unterliegen hohen Anforderungen.[32]

Die dritte Kategorie betrifft Systeme mit begrenztem Risiko, die Routineaufgaben erfüllen können, wie z. B. Dokumentenkategorisierung, die Erkennung doppelter Anträge oder KI zur Übersetzung. Nutzer müssen bei diesen Anwendungen vor allem darüber in Kenntnis gesetzt werden, dass sie mit KI interagieren oder Inhalte von KI erstellt wurden (Erwägungsgründe 27 und 53).

Die letzte Kategorie bilden Systeme mit minimalem Risiko, deren Auswirkungen auf Rechte oder Sicherheit als gering eingeschätzt werden, z. B. Videospiele oder Spamfilter.[33] Sie unterliegen keinen zusätzlichen Pflichten über das Datenschutzrecht hinaus. Stattdessen werden freiwillige Verhaltenskodizes gefördert (Erwägungsgrund 165). Dies dient auch der Förderung von Innovation und Entwicklung kleinerer Unternehmen.[34]

Kapitel 4 der KI-VO (Art. 51-57) widmet sich der Regulierung von *General-Purpose-AI-Models* (GPAI). Diese KI-Modelle werden auf umfangreichen Datensätzen trainiert und können für eine Breite an Aufgaben eingesetzt oder in andere KI-Systeme integriert werden. Mit diesem Kapitel reagiert die Verordnung auf die technologischen Risiken von GPAIs – denn die Basismodelle zu regulieren ist wichtig, um möglichen Gefährdungen für Grund- und Menschenrechte in darauf aufbauenden Anwendungen begegnen zu können. Dieses Thema wurde während der legislativen Debatten intensiv diskutiert, unter anderem im Hinblick auf mögliche innovationshemmende Effekte durch eine übermäßige Regulierung.[35]

Die KI-VO der EU verpflichtet alle relevanten Akteure, Hersteller, Anbieter, Betreiber und Importeure von KI-Systemen zur Einhaltung der Vorgaben und entfaltet damit Wirkung gegenüber allen Marktteilnehmern im Binnenmarkt. Das bedeutet, dass die Verordnung nicht nur gegenüber den EU-Mitgliedstaaten wirkt, sondern auch gegenüber Marktteilnehmern einschließlich privater Unternehmen wie Technologieanbietern.[36] Sie bindet Hersteller,

[32] Ebd. S. 175, 189; Erwägungsgrund 50 der Verordnung.
[33] Ebd., S. 209.
[34] Ebd.
[35] *Olevato*, Paving the path towards general purpose AI systems regulation in the AI Act: an analysis of the Parliament's and Council's proposals, https://www.medialaws.eu/rivista/paving-the-path-towards-general-purpose-ai-systems-regulation-in-the-ai-act-an-analysis-of-the-parliaments-and-councils-proposals/ (letzter Zugriff am 25.11.2025); *Innocenzo*, The regulation of foundation models in the EU AI Act, https://www.ibanet.org/the-regulation-of-foundation-models-in-the-eu-ai-act (letzter Zugriff am 28.11.2025); *Trüe*, in: Lisowski/Scholz/Trüe (Hrsg.), S.209.
[36] *Hogan, Lasek-Markey*, MDPI Philosophies, 9/2024, S. 7.

Anbieter, Betreiber, Importeure und Distributoren ohne Differenzierung zwischen privaten und öffentlichen Akteuren.

Eine zentrale und wenig diskutierte Folge der Implementierung der KI-VO ist auch, dass sie letztlich eine implizite Ausdehnung des Anwendungsbereichs von EU-Recht mit sich bringen kann – wobei Art. 51 Abs. 1 GRC in Verbindung mit Art. 6 Abs. 1 EUV sowie Art. 51 Abs. 2 GRC als Einfallstor dienen. Wenn nationale Behörden KI beispielsweise in Verwaltung, Asylverfahren oder bei der Polizei einsetzen – als konkretes Beispiel sei die KI-basierte Erstellung von Risikoprofilen genannt – so müssen sie die Vorgaben der Verordnung einhalten (Art. 2 Abs. 3).[37] Dadurch fallen zuvor rein nationale Angelegenheiten nun in den sachlichen Anwendungsbereich des Unionsrechts, weil die Behörden Unionsrecht „durchführen" im Sinne von Art. 51 Abs. 1 GrCh:[38] Folglich müssen die staatlichen Akteure bei der Umsetzung bzw. bei Nutzung entsprechender KI-Systeme die in der Charta garantierten Grundrechte beachten. In diesem Zusammenhang erwähnenswert ist auch, dass gemäß Art. 53 Abs. 3 GrCh die in der Grundrechtecharta verankerten Rechte in Umfang, Bedeutung und Tragweite jenen der EMRK entsprechen, wobei das Unionsrecht einen weitergehenden Schutz gewährleisten kann. Diese mögliche Erweiterung des Anwendungsbereichs der europäischen Grundrechte ist für die in dieser Arbeit zu betrachtenden menschenrechtsfördernden und -gefährdenden Potenziale von KI relevant.

Die KI-Verordnung enthält mehrere Ausnahmeregelungen: In Fällen von Sicherheitsrisiken oder im militärischen Kontext schließt Art. 2 Abs. 3 KI-Systeme für militärische oder nationale Sicherheitszwecke vom Anwendungsbereich der Verordnung aus. Art. 2 Abs. 6 stellt KI-Systeme oder Modelle, die ausschließlich für wissenschaftliche Forschungs- und Entwicklungszwecke entwickelt und in Betrieb genommen werden, ebenfalls frei, und laut Art. 2 Abs. 8 fallen diese auch vor dem Inverkehrbringen oder der Inbetriebnahme nicht unter die Verordnung, solange dabei das geltende Unionsrecht berücksichtigt wird. Art. 2 Abs. 12 schließt auch *Open-Source*-KI-Systeme von der Regulierung aus, sofern sie nicht als Hochrisiko-System gemäß Art. 5 gelten.

Da die Regulierung direkte wirtschaftliche Konsequenzen haben kann, haben Unternehmen das Interesse, ihre regulatorischen Vorgaben zu erfüllen und müssen entsprechende Ressourcen investieren. Die Mitgliedstaaten legen ihre eigenen Sanktions- und Durchsetzungsregeln fest, müssen dabei jedoch gewährleisten, dass diese „wirksam, verhältnismäßig und abschreckend"

[37] *Woll*, ZEuS 3/2025, S. 509, 523; *Fink*, The Hidden Reach of the EU AI Act, Verfassungsblog/2025.
[38] *Fink*, The Hidden Reach of the EU AI Act, Verfassungsblog/2025.

ausgestaltet sind (Art. 99 Abs. 1). So kann der Einsatz verbotener KI-Systeme nach Art. 5 mit bis zu 35 Mio. EUR oder 7 % des weltweiten Jahresumsatzes geahndet werden. Für die Übermittlung falscher oder irreführender Informationen an Behörden drohen bis zu 7,5 Mio. EUR oder 1 % des Umsatzes. Für KMU und Start-ups gelten niedrigere Obergrenzen (Art. 99 Abs. 3-6). Die konkrete Höhe der Sanktionen bemisst sich u. a. nach Schwere, Dauer und Konsequenzen des Verstoßes sowie der Kooperationsbereitschaft der Unternehmen gegenüber den Aufsichtsbehörden (Art. 99 Abs. 7). Der Europäische Datenschutzbeauftragte kann wiederum Bußgelder gegen EU-Institutionen und ihre Organe verhängen (Art. 100). Die Mitgliedstaaten müssen der Kommission ihre nationalen Sanktionsregelungen mitteilen und jährlich über verhängte Bußgelder berichten. Ebenfalls müssen sie sicherstellen, dass verfahrensrechtliche Garantien, insbesondere der Anspruch auf gerichtliche Überprüfung, dabei gewahrt bleiben (Art. 99 Abs. 11).

Mit dem Inkrafttreten der ersten Regelungsbereiche gelten seit Anfang 2025 bereits Verbote bestimmter Hochrisiko-Systeme, während für die anderen Risikokategorien gestaffelte Übergangsfristen vorgesehen sind. Auch für Anbieter von GPAI-Modellen wurden seit 2025 verbindliche Pflichten etabliert. Um die Umsetzung der Pflichten zu erleichtern, stellt die EU-Kommission unterstützende Instrumente bereit, darunter Auslegungshilfen der Vorgaben. Parallel dazu haben die Mitgliedstaaten ihre Aufsichtsstrukturen aufgebaut, sodass Sanktionen bereits greifen können.

Die Umsetzung der Vorgaben ist somit kein abgeschlossenes Modell, sondern eher ein fortlaufender regulatorischer Aushandlungsprozess, der durch zukünftige Praxis weiter ausgebaut und konkretisiert werden wird.

C. Der geopolitische Kontext aktueller KI-Regulierungsansätze

Die KI-Regulierung Europas wirkt nicht allein innerhalb ihrer eigenen Grenzen, sondern in einem Wettbewerb unterschiedlicher Regulierungsmodelle. In einer global vernetzten Wirtschaft und im digitalen Raum entfalten ihre regulatorischen Standards also zwangsläufig internationale Effekte und werden umgekehrt von anderen Akteuren beeinflusst. Deshalb muss der europäische Ansatz im geopolitischen Kontext betrachtet werden, insbesondere gegenüber den Hauptakteuren USA und China.[39]

[39] *Csernatoni,* Between innovation and regulation, https://carnegie-production-assets.s3.amazonaws.com/static/files/Csernatoni_The%20EU's%20AI%20Power%20Play_Between%20Deregul

Unterstützt wird der menschenrechtsorientierte Ansatz Europas auf globaler Ebene zunächst durch mehrere multilaterale Initiativen: Seit 2018 entwickelt die UNO menschenrechtsbasierte Leitprinzipien für KI, darauf aufbauend veröffentlichte die UNESCO 2021 die *Recommendation on the Ethics of AI* und etablierte 2023 ein multinationales Beratungsgremium.[40] Parallel formulierte die OECD weitere internationale Leitprinzipien und Handlungsempfehlungen für vertrauenswürdige, sichere und menschenzentrierte KI,[41] und im März 2024 verabschiedete die UN-Generalversammlung eine Resolution zum Schutz der Menschenrechte über den gesamten KI-Lebenszyklus hinweg.[42] Diese Entwicklungen verdeutlichen, dass der europäische Ansatz zwar auf einer wachsenden globalen normativen Grundlage aufsetzt, aber die Leitprinzipien von UNO, UNESCO und OECD können diesen als *Soft Law* nur normativ stützen. Die *Bletchley Declaration* von 2023, unterzeichnet von der EU und 28 Staaten einschließlich USA und China, hatte ursprünglich einen gemeinsamen globalen Rahmen für sichere KI-Entwicklung angestrebt und internationale Standards, Forschungspartnerschaften sowie Mechanismen zur Rechenschaftspflicht etabliert.[43] Weniger als zwei Jahre später untergrub die Rückkehr Donald Trumps ins Weiße Haus allerdings die amerikanische Beteiligung an diesen internationalen Bemühungen, wodurch das fragile globale Konsensmodell erheblich geschwächt wurde. Beim Pariser KI-Gipfel 2025 wurden diese Differenzen deutlich: Während viele Staaten sich weiterhin zu ethischer und nachhaltiger KI verpflichteten, verweigerten die USA und Großbritannien die Unterzeichnung der Abschlusserklärung. [44]

Gerade die Handlungen der USA stehen damit im starken Kontrast zu den Zielen der europäischen KI-Regulierung. Unter der Präsidentschaft von Joe Biden wurde 2023 noch die *Executive Order 14110* erlassen, die ein koordiniertes Regulierungsregime über 50 Bundesbehörden hinweg vorsah und die Entwicklung sowie den Einsatz von KI hinsichtlich

---

ation%20and%20Innovation.pdf (letzter Zugriff am 11.11.2025); *Peterson, Hoffmann,* Geopolitical implications of AI and digital surveillance adoption, https://www.brookings.edu/articles/geopolitical-implications-of-ai-and-digital-surveillance-adoption/ (letzter Zugriff am 10.11.2025).

[40] *UNESCO,* Recommendation on the Ethics of Artificial Intelligence, https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence (letzter Zugriff am 14.08.2025).

[41] *OECD*, Empfehlung des Rates zu künstlicher Intelligenz, OECD/LEGAL/0449.

[42] *UN-Generalversammlung*, A/78/L.49, https://docs.un.org/en/A/78/L.49.

[43] *Guzik, Sitek*, CVR, 119/2023, S. 2681, 2682.

[44] *Ministerium für Europa und auswärtige Angelegenheiten*, Aktionsgipfel zur Künstlichen Intelligenz, https://www.diplomatie.gouv.fr/de/aussenpolitik-frankreichs/digitale-diplomatie/article/aktionsgipfel-zur-kunstlichen-intelligenz-10-11-02-2025 (letzter Zugriff am 14.08.2025); *Amnesty International*, AI Action Summit must meaningfully center binding and enforceable regulation to curb AI-driven harms, https://www.amnesty.org/en/latest/news/2025/02/global-france-ai-action-summit-must-meaningfully-center-binding-and-enforceable-regulation-to-curb-ai-driven-harms/(letzter Zugriff am 12.10.2025); o.V. Pariser KI-Gipfel spricht sich für ethische und nachhaltige KI aus, https://www.zeit.de/digital/2025-02/paris-kuenstliche-intelligenz-abschlusserklaerung-ethisch (letzter Zugriff am 19.09.2025).

Sicherheit, Transparenz und Rechenschaftspflicht überwachen sollte. Doch US-Präsident Trump erklärte zu Beginn seiner zweiten Amtszeit, dass „Amerika das Land ist, das das KI-Rennen begonnen hat“ und es „gewinnen wird“, hob die von der Biden-Regierung eingeführten Vorgaben wieder auf und stellte stattdessen den “*Winning the Race: Americas AI Action Plan*” vor.[45] Statt auf staatliche Kontrolle setzt dieser auf Freiwilligkeit und marktwirtschaftliche Steuerungsmechanismen, mit rund 90 Maßnahmen, welche die Lockerung von Umweltvorschriften, die Ausweitung von KI-Exporten an Verbündete sowie die Einführung eines einheitlichen Bundesstandards vorsieht.[46] Ende Januar 2025 gab die US-Regierung außerdem mit „Stargate“ ein großangelegtes KI-Infrastrukturprojekt im Umfang von rund 500 Milliarden US-Dollar bekannt, an dem unter anderem OpenAI, *Oracle* und der japanische Konzern *Softbank* beteiligt sind.[47] Am 8. Dezember 2025 kündigte Präsident Trump schließlich an, eine Exekutivverordnung zu erlassen, welche die KI-Regulierungskompetenzen der US-Bundesstaaten aushebeln und einen einheitlichen Bundesstandard durchzusetzen soll. Damit reagierte er auf Anfragen großer Technologiekonzerne wie ChatGPT, OpenAI und Google und erklärte: “*We are beating all countries at this point in the race, but that won’t last long if we are going to have 50 States, many of them bad actors, involved in rules and the approval process*”.[48]

Damit stellt er sich ausdrücklich gegen die Bemühungen einiger Bundesstaaten, eigene KI-Schutzstandards zu entwickeln. So hatte beispielsweise Floridas republikanischer Gouverneur DeSantis ein *„AI Bill of Rights“* angekündigt, welches Elternrechte, Daten- und Verbraucherschutz stärken sollte, während Kalifornien ein Gesetz gegen nicht-einvernehmliche KI-Pornografie, politische Deepfakes und algorithmische Diskriminierung verabschiedet hatte.[49]

Damit bestätigt US-Präsident Trump erneut, dass Gefahren für Menschenrechte in seiner KI-Politik keine Rolle spielen, sondern industriepolitische Wachstumsinteressen und geopolitische Wettbewerbsfähigkeit klar priorisiert werden.

---

[45] *The White House*, Removing Barriers to American Leadership in Artificial Intelligence, https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/ (letzter Zugriff am 19.11.2025).
[46] *The White House*, Action Plan AI Race, https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf (letzter Zugriff am 20.11.2025).
[47] *Tatevik,* Reserve University Journal of Law Technology & the Internet, 16/2025, S. 223, 257.
[48] *Duffy*, Trump says he’ll sign executive order blocking state AI regulations, despite safety fears, https://edition.cnn.com/2025/12/08/tech/trump-eo-blocking-ai-state-laws (letzter Zugriff am 08.12.2025).
[49] *Reuters*, Trump to issue order creating national AI rule, https://www.reuters.com/world/trump-says-he-will-sign-executive-order-this-week-ai-approval-process-2025-12-08/ (letzter Zugriff am 08.12.2025).

Ein weiterer zentraler Akteur der KI-Politik ist China. Chinas Ansatz zur Regulierung und strategischen Entwicklung von KI begann mit dem *New Generation Artificial Intelligence Development Plan*, der einen dreistufigen Entwicklungsprozess für die chinesische KI-Regulierung formulierte.[50] Mit den *Interim Measures for the Management of Generative Artificial Intelligence Services* wurden im August 2023 erstmals bindenden Vorschriften für generative KI-Dienste erlassen.[51] China verfolgt eine Regulierung, die marktgetriebene Innovation und strikte staatliche Kontrolle kombiniert. [52] Diese stützt sich auf große private Tech-Konzerne, die eng in staatliche Strategien eingebunden sind; es werden Maßnahmen wie Zollerhöhungen und koordinierte internationale Exportbeschränkungen eingesetzt.[53] Ähnlich wie in den EU-Regulierungen werden auch Sicherheit und soziale Stabilität thematisiert, während aber zugleich wirtschaftliche Innovation und Wettbewerbsdynamiken im Fokus stehen.[54]

In der Praxis nutzt China seine stark zentralisierte Strategie, um KI gezielt für Überwachungsmaßnahmen, staatliche Kontrolle und militärische Modernisierung einzusetzen. Während die Überwachung der Bevölkerung lange Zeit auf menschliche Informanten-Netze gestützt war, ermöglichen KI-Systeme nun eine neue und effizientere Form staatlicher Kontrolle.[55] Das *Social Scoring*-System Chinas bewertet BürgerInnen und Unternehmen anhand vielfältiger Daten aus Finanzen, Recht und Alltag nach ihrer „Zuverlässigkeit".[56] Hohe Punktzahlen ermöglichen Vorteile wie erleichterten Kreditzugang oder bevorzugte Schulzulassung, niedrige Werte führen zu Einschränkungen beim Reisen, beim Zugang zu Internetdiensten oder bei Beschäftigungsmöglichkeiten.[57] Das System verdeutlicht, wie KI als Instrument eines Überwachungsstaates massiv in Freiheitsrechte, Datenschutz und Privatsphäre eingreifen kann. Chinesische Plattformen wie TikTok sind ebenfalls ein Beispiel dafür, wie

[50] *McGurk, Tomlinson*, S. 30; *General Offices of the CPC Central Committee and State Council*, Opinion on Strengthening the Governance of Science and Technology Ethics, https://www.gov.cn/zhengce/2022-03/20/content_5680105.htm (letzter Zugriff am 17.09.2025).
[51] *Zhang*, Colum. J. Transnat'l L., 63/2025, S. 101, 109; *McGurk, Tomlinson*, S.31.
[52] *Roberts et al*, S. 48 f; *Zhang*, Colum. J. Transnat'l L., 63/2025, S. 101,114.
[53] *Teer*, Beyond Trump: Xi's price wars and weaponisation of critical raw materials threaten European prosperity, https://www.iss.europa.eu/publications/commentary/beyond-trump-xis-price-wars-and-weaponisation-critical-raw-materials (letzter Zugriff am 28.10.2025); *Berder*, *EYPIC,* 2020, S. 89, 93 f.
[54] *Rolf*, China's regulations on algorithms: Context, impact, and comparisons with the EU, https://library.fes.de/pdf-files/bueros/bruessel/19904.pdf (letzter Zugriff am 29.10.2025).
[55] *Woods*, Journal of Chinese Political Science, 2025, S. 10 ff.
[56] *Pei*, S. 10 ff.
[57] *Engelmann, et al.*, Clear sanctions, vague rewards: how China's social credit system currently defines "good" and "bad" behavior, https://www.researchgate.net/profile/Severin-mann/publication/330272009_Clear_Sanctions_Vague_Rewards_How_China%27s_Social_Credit_System_Currently_Defines_Good_and_Bad_Behavior/links/5d932f1a299bf10cff1cfa0e/Clear-Sanctions-Vague-Rewards-How-Chinas-Social-Credit-System-Currently-Defines-Good-and-Bad-Behavior.pdf (letzter Zugriff 19.09.2025); *Bukhari, Munazzah, Anwar*, JDSS, 6/2025, 68, 78.

Technologie zur Manipulation von Meinungsbildung und Konsumentenverhalten eingesetzt werden kann, während gleichzeitig umfangreiche persönliche und biometrische Daten gesammelt werden – und das in globalem Umfang.[58]

Diese knappe Einordnung verdeutlicht, dass Europa im globalen Kontext der KI-Regulierung in einem Spannungsfeld zwischen normativer Führungsrolle, der Sicherung eigener technologischer und wirtschaftlicher Wettbewerbsfähigkeit und den divergierenden strategischen Leitlinien anderer zentraler Akteure agieren muss, vor allem im Hinblick auf den Anspruch, europäische regulatorische und dabei menschenrechtszentrierte Ansätze international wirksam werden zu lassen. Gleichzeitig lassen sich innerhalb der EU aktuelle Entwicklungen beobachten, die auf eine Verschiebung der regulatorischen Prioritäten hindeuten: Wettbewerbsfähigkeit, Innovationsförderung und die Sorge, im globalen KI-Wettlauf zurückzufallen, gewinnen zunehmend an Bedeutung, während bestehende regulatorische Rahmenwerke punktuell infrage gestellt oder abgeschwächt werden. Dies zeigt sich insbesondere in jüngeren Vorschlägen zur Anpassung des europäischen Digital- und Datenrechts sowie in Tendenzen zur Flexibilisierung zentraler Schutzstandards.[59] Eine eindeutige Abkehr vom menschenrechtszentrierten Ansatz ist damit zwar nicht festzustellen, jedoch wird deutlich, dass dieser zunehmend unter Rechtfertigungsdruck gerät.

D. Menschenrechtsfördernde Potenziale von KI

Wie sich zeigen wird, nehmen die Risiken und sicherheitspolitischen Herausforderungen von KI im weiteren Verlauf dieser Arbeit den größeren Raum ein. Zunächst soll dennoch der Blick auf ihr Potenzial für gesellschaftlichen Nutzen und die Förderung von Menschenrechten gerichtet werden. Dieser Abschnitt soll dabei nicht die Gefahren von KI relativieren, sondern ihrer Einordnung dienen: Ein umfassendes Verständnis von KI erfordert auch die Betrachtung ihrer positiven Anwendungsmöglichkeiten. Dadurch wird zudem deutlich, dass KI ein ambivalentes Werkzeug ist, dessen gesellschaftlicher Einfluss sehr von ihrer Entwicklung und Regulierung abhängt.[60] Dass in der Praxis sogar dasselbe System menschenrechtsfördernde und menschenrechtsgefährdende Wirkungen entfalten kann, liegt auch an der Funktionsweise von KIs als sogenannte *Black Boxes*: Zahlreiche Datensätze werden während der Entwicklung

[58] *Baglione, Tucci*, AMJ, 3/2025; *Kohn*, Chi. J. Int'l L., 23/2022, S. 195; *Komy*, TikTok: China's New Weapon, https://www.habtoorresearch.com/p rogrammes/tiktok-china-weapon/ (letzter Zugriff am 12.09.2025).
[59] *Bieker*, European AI FOMO, The European Commission Sacrifices the Digital Acquis at the Altar of AI Hype, https://verfassungsblog.de/eu-digital-law-ai-fomo-omnibus/ (letzter Zugriff am 23.03.2026).
[60] *Toju*, S.13-15.

eingespeist, und es ist oft nur schwer nachvollziehbar, wie spätere Ergebnisse genau zustande kommen.[61] Die Auswirkungen auf Grundrechte hängen also bei manchen Modellen, besonders den genannten *GPAI*s, nur davon ab, wie sie konfiguriert und eingesetzt werden – was die Regulierung von KI besonders herausfordernd macht.[62]

I. KI zur Förderung von Sicherheit und Freiheit

Auch im Hinblick auf Sicherheit und Freiheitsrechte eröffnen KI-gestützte Anwendungen neue Möglichkeiten, die über reine Effizienzsteigerung hinausgehen. So kann der Einsatz elektronischer Überwachung als Alternative zur Freiheitsstrafe dazu beitragen, die soziale Wiedereingliederung Betroffener zu fördern: In der Schweiz ist zwischen 2018 und 2023 die Nutzung elektronischer Überwachung als Alternative zur Freiheitsstrafe um rund 25% gestiegen.[63] Die genutzten elektronischen Fußfesseln erfassen primär Standortdaten – doch zukünftig könnten solche Systeme durch KI-basierte Funktionen erweitert werden, die jederzeit und mit geringem Ressourceneinsatz große Datenmengen an Bewegungs- und Verhaltensmustern erfassen und auswerten, Abweichungen und Anomalien darin automatisch melden und damit eine deutlich umfassendere, vorausschauende Überwachung ermöglichen.[64] Auf diese Weise lassen sich Sicherheitsziele effizient erfüllen, ohne dass die betroffene Person dauerhaft aus ihrem sozialen Umfeld entfernt werden muss – ein Ansatz, der sowohl Freiheitsrechte wahrt als auch die Integration in die Gesellschaft unterstützt.

In Sicherheits- und Strafverfolgungskontexten wird der Einsatz von KI bereits zunehmend relevanter, wobei sowohl die Effizienz als auch die Qualität der Ermittlungen verbessert werden kann. In Hessen nutzt die „PolizeiCloud“ KI-gestützte Anwendungen zur Analyse und Aufbereitung großer Datenmengen in Ermittlungsverfahren, z. B. bei Kinder- und Jugendpornografie oder Geldwäsche.[65] Auch das Projekt „Einsatz von KI zur Früherkennung von Straftaten“ (KISTRA) untersucht in Zusammenarbeit mit Ermittlungsbehörden Einsatzformen von KI zur Vorhersage und Prävention von Straftaten unter Berücksichtigung

[61] *Garouani*, Investigating the Duality of Interpretability and Explainability in Machine Learning, https://arxiv.org/html/2503.21356v1 (letzter Zugriff am 14.10.2025).
[62] *Vítor*, EDPS, 2023, S. 3 ff.
[63] *Schweizerische Eidgenossenschaft*, Electronic Monitoring by year of ending and the main sentence executed, https://www.bfs.admin.ch/bfs/rm/home.assetdetail.32809187.html (letzter Zugriff am 08.11.2025); *Swissinfo*, Criminalité: le bracelet électronique a fait ses preuves, https://www.swissinfo.ch/fre/criminalit%c3%a9%3a-le-bracelet-%c3%a9lectronique-a-fait-ses-preuves/89866273 (letzter Zugriff am 08.11.2025).
[64] Ebd.
[65] *Hess, Klein*, Mit KI gegen Kriminalität, https://hzd.hessen.de/medienraum/publikationen/hzd-inform/inform-3-23/kuenstliche-intelligenz-mit-ki-gegen-kriminalitaet (letzter Zugriff am 14.09.2025).

rechtlicher und ethischer Rahmenbedingungen.[66] Laut einem Bericht des Bundesministeriums des Innern wird KI zur Fahndung und zur Effizienzsteigerung bei Ermittlungsprozessen als unverzichtbar angesehen.[67] Auch automatische Gesichtserkennungssysteme werden in Europa in einzelnen Pilotprojekten getestet, etwa zur Identifikation gesuchter Personen oder zur Überwachung öffentlicher Plätze.[68] Solche Technologien zeigen, dass KI nicht nur Effizienzsteigerung ermöglicht, sondern in vielen Fällen auch Grundrechte fördern kann. So können Freiheitsrechte gestärkt werden, indem Betroffene im sozialen Umfeld verbleiben, Sicherheitsrechte werden durch präventive Maßnahmen erhöht und Integrationsrechte durch Unterstützung der Wiedereingliederung gefördert werden. Gleichzeitig bergen solche Technologien erhebliche Risiken für Grundrechte, insbesondere für Privatsphäre, Datenschutz, informationelle Selbstbestimmung und den Schutz vor diskriminierender Überwachung.[69] Die zentrale Herausforderung besteht darin, den Einsatz von KI so zu gestalten, dass öffentliche Sicherheit und die Förderung von Grundrechten in Einklang stehen, ohne dass die Schutzstandards für Individuen geschwächt werden.

Trotz der im Folgenden noch zu diskutierenden Risiken kann KI auch im sicherheitspolitischen Bereich eine stabilisierende Funktion übernehmen. Ihre Fähigkeit, große, heterogene und sich verändernde Datenmengen zu verarbeiten, ermöglicht Gefahrenfrüherkennungen, z.B. von Cyberangriffen und Schadsoftware, oder das Erkennen militärischer Installationen in Überwachungsbildern sowie auch von Anomalien in staatlichen oder öffentlich zugänglichen Informationen.[70] Damit trägt KI zur Förderung von Grundrechten bei, indem sie Leben, körperliche Unversehrtheit und öffentliche Sicherheit schützt.

---

[66] *BMFTR, KISTRA: Einsatz von KI zur Früherkennung von Straftaten*, https://www.sifo.de/sifo/de/projekte/querschnittsthemen-und-aktivitaeten/kuenstliche-intelligenz-in-der-zivilen-sicherheitsforschung/kistra/kistra-einsatz-von-ki-zur-frueherkennung-von-straftaten.html (letzter Zugriff am 01.12.2025)

[67] *Bundesministerium des Inneren*, Einsatz von KI bei der Fahndung, https://www.bmi.bund.de/SharedDocs/kurzmeldungen/DE/2024/11/bka-herbst.html, (letzter Zugriff am 14.10.2025).

[68] *European Union Agency for Fundamental Rights*, Facial recognition technology: fundamental rights considerations in the context of law enforcement, https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf (letzter Zugriff am 14.11.2025); *Oellig,* Ausgeguckt, https://www.amnesty.de/informieren/amnesty-journal/europa-massenueberwachung-gesichtserkennung-eu-gesetz-ausgeguckt (letzter Zugriff am 07.11.2025); *Digital Chiefs,* Olympia 2024: Polizeipräfektur in Paris erlaubt KI-gestützte Videoüberwachung, https://www.digital-chiefs.de/olympia-2024-polizeipraefektur-in-paris-erlaubt-ki-gestuetzte-videoueberwachung/ (letzter Zugriff am 14.10.2025).

[69] Siehe auch: *EU Agency for Fundamental Rights, Facial recognition technology*, Fundamental Rights Considerations in the Context of Law Enforcement, https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf (letzter Zugriff am 14.11.2025).

[70] *Reinhold, Reuter*, S. 343 ff.

Darüber hinaus eröffnet KI neue Möglichkeiten im Rahmen von Rüstungskontrollmaßnahmen. Durch die erwähnte *Black Box*-Arbeitsweise können dabei auch sensible Daten der betroffenen Staaten geschützt werden, indem die KI Entscheidungen wie „ist eine Waffe“ oder „ist keine Waffe“ trifft, ohne vertrauliche Details offenlegen zu müssen. Die Integrität solcher Systeme lässt sich durch digitale Siegel oder kryptographische Sicherheitsnummern sichern.[71] Diese Funktionen tragen ebenfalls zur Förderung von Grundrechten bei, indem sie Schutzpflichten des Staates stärken, die Rechte von Einzelpersonen fördern und Risiken für die Allgemeinheit mindern.

Da staatlich eingesetzte KI-Systeme unter die KI-Verordnung fallen, müssen bei ihrem Einsatz die in der Charta verbürgten Grundrechte gewahrt werden, insbesondere die Menschenwürde (Art. 1), Freiheit und Unversehrtheit der Person (Art. 3, 6), das Diskriminierungsverbot (Art. 21) sowie Transparenz- und Rechenschaftsanforderungen als Teil eines grundrechtskonformen Risikomanagements. Zu nennen sind im Zusammenhang mit dem Pflichtenkanon staatlicher Stellen auch die betreffenden EMRK-Rechte, insbesondere das Recht auf Achtung des Privat- und Familienlebens (Art. 8), das Recht auf Freiheit und Sicherheit (Art. 5), sowie in manchen Fällen das Recht auf ein faires Verfahren (Art. 6) und das Diskriminierungsverbot (Art. 14 EMRK).

II. KI zur Förderung von Fairness und Teilhabe sowie in der Pflege

KI kann den Zugang zu Teilhabe in Berufen, Bildung und wichtigen Dienstleistungen erleichtern und damit inklusive Strukturen fördern. Besonders auf dem Arbeitsmarkt, bei der beruflichen Rehabilitation und der Integration von Menschen mit Behinderungen können KIs in Assistenztechnologien integriert werden.[72] Beispiele hierfür sind Anwendungen wie Vorlese-Apps, Spracherkennungssoftware oder intelligente Planungshilfen, die zur Bewältigung alltäglicher Aufgaben beitragen können.[73] KI-basierte adaptive Werkbänke oder unterstützende Robotik können dazu beitragen, Arbeitsplätze an individuelle Bedürfnisse anzupassen.[74] Im Bildungsbereich können digitale Technologien Lernprozesse unterstützen, z.B. durch KI-gestützte digitale Erklärungstools oder durch personalisierte Lernsysteme, die Inhalte an unterschiedliche Verständnisniveaus anpassen. Auch Übersetzungssysteme können

[71] *Wiese*, Gaza, Artificial Intelligence, and Kill Lists, https://verfassungsblog.de/gaza-artificial-intelligence-and-kill-lists/ (letzter Zugriff am 10.11.2025); *Reinhold, Reuter*, S. 345.
[72] *Ferebee*, PJAI, 2025.
[73] Ebd.
[74] *Beudt*, Künstliche Intelligenz und Inklusion in der Arbeitswelt-Leitlinien und Kompetenzen für die KI-gestützte Förderung beruflicher Teilhabe, https://digid.jff.de/ki-expertisen/inklusion (letzter Zugriff am 01.10.2025).

besonders für Kinder mit Migrationshintergrund helfen. Damit verbessern sie den Zugang zu Bildung und fördern Chancengleichheit.[75]

Auch in Asylverfahren oder bei Behördenkontakten sollen automatisierte Sprachassistenz- und Übersetzungssysteme die Kommunikation erleichtern.[76] Dazu zählen Übersetzungs-Apps, Text-zu-Sprache-Technologien oder automatische Dokumentenerkennung, die Menschen mit sprachlichen oder kognitiven Einschränkungen ermöglichen, Informationen selbstständig zu verstehen und Anträge korrekt einzureichen.[77]

In der Pflege gewinnt der Einsatz von KI-Systemen zunehmend an Bedeutung, besonders vor dem Hintergrund einer alternden Bevölkerung und des strukturellen Pflegenotstands.[78] Assistierende KI-Technologien wie Sturzsensoren oder Monitoring-Systeme zur Erinnerung an die Bewegung, Speisen und Medikamenteneinnahme sollen Pflegekräfte entlasten und die Sicherheit der Pflegebedürftigen erhöhen.[79] Zukünftig könnte in der Pflege auch mit vermehrten Einsätzen von Servicerobotern und sprachbasierten KI-Systemen zur sozialen Unterstützung gerechnet werden. [80]

Damit solche KI-Anwendungen tatsächlich einen inklusiven und menschenrechtsfördernden Nutzen entfalten, muss der traditionelle technologiezentrierte Blick um eine menschzentrierte Perspektive erweitert werden. Die Grundprinzipien des *Human-Centered AI*-Ansatzes (*HCAI*) geben hierfür eine Orientierung: KI-Systeme sollen den Bedürfnissen, Fähigkeiten und Werten der NutzerInnen dienen und deren Kontrolle über die Technologie wahren.[81] Förderprogramme und Pilotprojekte, wie z.B. die Förderrichtlinien „KI für das Gemeinwohl" des Bundesministeriums für Bildung, Familie, Senioren, Frauen und Jugend, unterstützen gezielt innovative Ansätze in diesem Feld.[82]

---

[75] *Stattkus,* HMD Praxis der Wirtschaftsinformatik, 2025, S.6; *Vorobyeva et al.,* Personalized learning through AI: Pedagogical approaches and critical insights, https://www.cedtech.net/download/personalized-learning-through-ai-pedagogical-approaches-and-critical-insights-16108.pdf (letzter Zugriff am 02.11.2025).

[76] *Europäisches Parlament*, Artificial Intelligence in Asylum Procedures in the EU, https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775861/EPRS_BRI(2025)775861_EN.pdf (letzter Zugriff am 01.12.2025).

[77] *Bundesministerium für Bildung, Familie, Senioren Frauen und Jugend*, https://www.bmbfsfj.bund.de/bmbfsfj/ministerium/ausschreibungen-foerderung/foerderrichtlinien/kuenstliche-intelligenz-fuer-das-gemeinwohl (letzter Zugriff am 14.10.2025).

[78] *Laufenberg*, in: Friedewald et. al. (Hrsg.), S. 253 f.

[79] Ebd. S. 360 ff.

[80] *Bundesministerium für Bildung, Familie, Senioren Frauen und Jugend*, S. 11.

[81] *Shneiderman,* S. 3 f.

[82] *Bundesministerium für Bildung, Familie, Senioren Frauen und Jugend*: Richtlinie zur Förderung von Künstlicher Intelligenz für das Gemeinwohl, https://www.bmbfsfj.bund.de/bmbfsfj/ministerium/ausschreibungen-foerderung/foerderrichtlinien/kuenstliche-intelligenz-fuer-das-gemeinwohl- (letzter Zugriff am 14.10.2025).

KI-Systeme können gerade im Pflegekontext die Gesundheitsversorgung verbessern und somit insbesondere das Recht auf Gesundheitsschutz nach Art. 35 GrCh fördern – wobei die Grundrechtecharta wie dargelegt anwendbar ist, soweit staatliche Akteure an der Bereitstellung bzw. Nutzung der KI-Systeme beteiligt sind und somit Unionsrecht umgesetzt wird, d.h. beispielsweise in Fällen staatlicher Träger von Pflegeheimen und ähnlichen Einrichtungen. Zugleich verarbeiten die genannten KI-Anwendungen regelmäßig besonders schutzwürdige Gesundheitsdaten im Sinne von Art. 9 DSGVO, sodass ein hoher datenschutzrechtlicher Schutzstandard zu gewährleisten ist. In Bezug auf die EMRK können das Recht auf Leben (Art. 2) und das Recht auf Privat- und Familienleben (Art. 8) genannt werden. Auch gestützt auf die Europäische Sozialcharta (Art. 11, Revised ESC, ETS No. 163) werden relevante Schutzpflichten sichtbar, insbesondere die Pflicht der Staaten, eine angemessene Gesundheitsversorgung sicherzustellen, die Rechte von Pflegebedürftigen zu wahren, den Schutz sensibler Gesundheitsdaten zu gewährleisten und den Zugang zu staatlich bereitgestellten Pflege- und Gesundheitseinrichtungen zu garantieren.

Werden Effizienz- oder Innovationsinteressen im Umgang mit diesen KI-Systemen jedoch über die Rechte wie Datenschutz und Privatsphäre (Art. 7, 8 GRCh; Art. 8 EMRK), die körperliche und psychische Unversehrtheit (Art. 3 GRCh), die Menschenwürde (Art. 1 GRCh) oder das Diskriminierungsverbot (Art. 21 GRCh) gestellt, besteht ein Risiko für eben diese Rechte. Diese Grundrechte wirken dabei nicht unmittelbar gegenüber privaten Dritten, entfalten jedoch auch insoweit mittelbar Wirkung über die Auslegung und Anwendung der einschlägigen EU-Rechtsakte, insbesondere der KI-VO (sie gilt wie jede Verordnung unmittelbar in allen Mitgliedstaaten), und des nationalen Rechts. Diese Grundrechte wirken dabei nicht unmittelbar gegenüber privaten Dritten, entfalten jedoch auch insoweit mittelbare Wirkung über die Auslegung und Anwendung der einschlägigen EU-Rechtsakte, insbesondere der KI-VO, die wie jede Verordnung unmittelbar in allen Mitgliedstaaten gilt, sowie über nationales Recht und begründen damit auch staatliche Schutzpflichten.[83]

III. KI zur Förderung des Umweltschutzes und demokratischer Diskurse

KI-Systeme eröffnen unterschiedliche Möglichkeiten, Umweltherausforderungen zu begegnen, und können im Rahmen staatlicher Schutzpflichten dazu beitragen, das Recht auf Leben (Art.

[83] Art. 51 Abs. 1, Art. 52 Abs. 1 GrCh; Vgl. *EUR Lex,* Die unmittelbare Wirkung des Rechts der Europäischen Union, https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775861/EPRS_BRI(2025)775861_EN.pdf (letzter Zugriff am 24.03.2026).

2 EMRK, GRCh) zu schützen, sowie ein hohes Umweltschutzniveau gemäß Art. 37 GRCh zu gewährleisten. So können sie beispielsweise durch autonome Robotik Systeme und Drohnen die Vorhersage und Bekämpfung von Hochwassern oder Bränden verbessern, was Einsatzkräfte schützen, Umweltgefahren schneller begegnen und Ressourcen schonen kann.[84]

In der Landwirtschaft kann KI dazu beitragen, die wachsende Weltbevölkerung nachhaltig zu ernähren, ohne zusätzliche landwirtschaftliche Flächen zu beanspruchen: Sensoren, Drohnen und Maschinen erkennen Pflanzenkrankheiten und Schädlinge, überwachen Böden und Ernten und optimieren den Ressourceneinsatz.[85] KI-Systeme können das Überleben von Bienen fördern, indem sie Frühwarnsysteme für deren Populationen entwickeln.[86] Auch der Schutz mariner Ökosysteme profitiert von KI: Organisationen wie *Ocean Cleanup* setzen KI ein, um Plastik im Meer zu erkennen, zu kartieren und in großen Mengen gezielt zu entfernen.[87]

Im urbanen Kontext unterstützt KI die Entwicklung sogenannter *Smart Cities*. Länder wie Singapur, Norwegen, Finnland, die Schweiz und Südkorea nutzen KI-Analysen an Satellitenbildern, um Entwaldung vorherzusagen, städtische Hitzeinseln durch Informationen zu Baumflächen und Grünzonen zu analysieren und damit Städte energieeffizienter zu gestalten.[88]

Auch in Sachen Demokratie könnte KI jedenfalls potenziell dazu beitragen, Vorurteile und Fehlinformationen sachlich zu korrigieren: Das aktuell meistgenutzte Sprachmodell ChatGPT ist derzeit so eingestellt, dass es diskriminierende Aussagen mit faktenbasierten Richtigstellungen beantwortet und dabei einen respektvollen Ton wahrt.[89] Während öffentliche Diskurse in sozialen Medien häufig durch Polarisierung geprägt sind, können KI-Systeme potenziell und unter den richtigen Bedingungen dazu beitragen, sachliche Gesprächsräume zu bieten und in einer zunehmend polarisierenden Öffentlichkeit Vorurteile, Fehlinformation oder Verschwörungstheorien korrigieren, indem sie emotionslos und praktisch „gelassen“ reagieren. Eine diesjährige Längsschnittstudie zeigte allerdings, dass KI-Dialoge zwar kurzfristig

[84] *Toju*, S. 28 ff.
[85] Ebd.
[86] *Asaduz, Dorin*, S. 1 ff.
[87] *Toju*, S. 28-33.
[88] *UN*, Artificial Intelligence for Climate Action in Developing Countries: Opportunities, Challenges and Risk, https://unfccc.int/ttclear/misc_/StaticFiles/gnwoerk_static/AI4climateaction/28da5d97d7824d16b7f68a225c0e3493/a4553e8f70f74be3bc37c929b73d9974.pdf (letzter Zugriff am 19.08.2025).
[89] *DeVerna*, PNAS, 2024, https://www.pnas.org/doi/epdf/10.1073/pnas.2322823121 (letzter Zugriff am 14.11.2025).

Fehlinformationen korrigieren können, aber die unabhängigen Erkennungsfähigkeiten der Nutzer langfristig verschlechtern.[90]

Damit entsteht eine ambivalente Perspektive: KI kann zwar helfen, Fehlinformationen zu verringern, gleichzeitig besteht aber das Risiko, dass kritisches Denken zunehmend an technische Systeme abgegeben wird. Dies stellt insbesondere in autoritär geprägten digitalen Umgebungen eine erhebliche Gefahr dar, denn die „Gespräche" mit Sprachmodellen können auch das Wahlverhalten beeinflussen. Bei den Auswertungen des „KI-o-Mat" zeigt sich zwar aktuell in Deutschland, dass große Sprachmodelle in ihren Antworten auf Wahl-O-Mat-Positionen überwiegend Gleichberechtigung, Nichtdiskriminierung und Menschenwürde betonen.[91] Diese normativ-ethische Grundausrichtung ist jedoch nicht Ausdruck einer neutralen, geschweige denn „guten" KI, sondern Ergebnis der aktuellen Programmierungen und Filter, die diskriminierende oder rassistische Inhalte gezielt abfangen.[92] Anders programmiert, z.B. unter einer Regierung mit illiberaler Agenda, könnte dieselbe KI genutzt werden, gesellschaftliche Narrative zu verzerren, Kritik zu unterdrücken und Propaganda zu verstärken. Bereits hier wird jedoch deutlich, dass die Auswirkungen von KI untrennbar an ihre verantwortungsvolle Konfiguration und damit an ihre demokratische Regulierung gebunden ist.

E. Menschenrechtsgefährdungen durch KI

Im Folgenden werden ausgewählte Risikobereiche genannt, die exemplarisch für das breite Spektrum menschenrechtlicher Herausforderungen stehen, die der Einsatz von KI in unterschiedlichen Kontexten mit sich bringt. Ziel ist es, zentrale Problemlagen zu beleuchten, ohne den Anspruch auf Vollständigkeit zu erheben, und aufzuzeigen, wie tiefgreifend KI bereits heute in gesellschaftliche und politische Prozesse eingreifen kann und diese künftig noch stärker prägen wird.

[90] *Rani et al*., Dialogues with AI Reduce Beliefs in Misinformation, https://arxiv.org/html/2510.01537v1 (letzter Zugriff am 29.10.2025).
[91] *Köller*, Drei Sprachmodelle und ein Wahl-O-Mat: Ein Blick auf wirtschaftspolitische Entscheidungen, https://www.ihk.de/stuttgart/presse/ki-o-mat-6461208 (letzter Zugriff am 13.10.2025).
[92] *Young*, What is RLHF? Reinforcement learning from human feedback for AI alignment, https://wandb.ai/byyoung3/huggingface/reports/What-is-RLHF-Reinforcement-learning-from-human-feedback-for-AI-alignment--VmlldzoxMzczMjEzMQ (letzter Zugriff am 12.11.2025); *Adrien*, Reinforcement Learning from Human Feedback (RLHF) Explained, https://intuitionlabs.ai/articles/reinforcement-learning-human-feedback (letzter Zugriff am 20.09.2025).

I. Algorithmische Diskriminierung und *Bias*

"*Bias from the past leads to bias in the future*"[93] – Ashwini, UN Special Rapporteur

Algorithmische Diskriminierung bezeichnet die systematische Benachteiligung bestimmter Gruppen durch automatisierte Entscheidungsprozesse und kann auch als *Bias*, also Verzerrung oder Voreingenommenheit, bezeichnet werden. Sie entsteht, weil beim maschinellen Lernen und Training der Modelle auf große Datenmengen zurückgegriffen wird – mit Daten, die gesellschaftliche Strukturen sowie historische Ungleichheiten beinhalten, welche mit in die Systeme eingespeist werden.[94] KIs können schließlich nur Muster erkennen, aber nichts im eigentlichen Sinne verstehen.[95] Wenn sie dann aus diesen Daten Korrelationen und Schlussfolgerungen ziehen, ohne die Ursachen oder sozialen Kontexte zu begreifen, können sie gesellschaftliche Vorurteile in scheinbar objektive Prognosen übersetzen, was bestehende Diskriminierungen nicht nur reproduzieren, sondern sogar verstärken kann.[96] Grundsätzlich ist es daher ein Fehlschluss anzunehmen, dass algorithmische Entscheidungen neutral oder wertfrei seien. Jedes KI-System ist ein Produkt menschlicher Entscheidungen, von der Datenauswahl über die Modellarchitektur bis zur Definition dessen, was als gerecht oder wahr gilt.[97]

In der Forschung wird Bias in KI-Systemen in drei Hauptformen unterschieden: Vorbestehender Bias spiegelt gesellschaftliche Strukturen und Vorurteile wider, die bereits in die Datengrundlage eingeflossen sind. Technischer Bias entsteht im Modellierungsprozess selbst, durch die Auswahl von Algorithmen oder Trainingsmethoden, die bestimmte Muster über- oder unterrepräsentieren. *Emergent Bias* tritt auf, wenn ein Modell außerhalb seines ursprünglichen Kontexts angewendet wird und dadurch unangemessene oder diskriminierende Ergebnisse produziert.[98]

Die Auswirkungen algorithmischer Diskriminierung lassen sich in drei Dimensionen einordnen: Psychologische Schäden durch beleidigende oder manipulative Inhalte; epistemische Verzerrungen, die zu Fehlinformationen oder verzerrtem Weltbild führen; sowie

[93] *UN*, Racism and AI-Bias from the past leads to bias in the future, https://www.ohchr.org/en/stories/2024/07/racism-and-ai-bias-past-leads-bias-future (letzter Zugriff am 19.08.2025).
[94] *Borgesius*, S. 11 ff.
[95] *Domalewska et al*, S. 30.
[96] *Kempt*, 2025, S. 148-157.
[97] *Bender et al.,* S. 214 ff.
[98] *Toju*, S. 67 ff.

strukturelle Schäden, die durch den fortgesetzten Einsatz solcher Systeme bestehende Ungleichheiten langfristig verfestigen.[99]

Da Bias eng mit gesellschaftlichen Machtverhältnissen verknüpft ist, zeigt er sich häufig intersektional und verschärft sich, wenn mehrere Formen sozialer Marginalisierung zusammentreffen.[100] Effektive Maßnahmen zur Förderung von Fairness und Gleichstellung müssen daher stets kontextsensitiv und reflektiert geprüft werden.[101]

Konkrete Beispiele verdeutlichen, welche Auswirkungen es haben kann, wenn diese Prüfung nicht ausreichend erfolgt: Die Apple Card wies 2019 Frauen trotz gleicher Bonität systematisch niedrigere Kreditlimits als Männern zu, obwohl das KI-System keinen direkten Zugriff auf das Geschlecht hatte, denn es bezog sich auf indirekte Geschlechtsindikatoren wie Name oder Hobbies und entschied entsprechend bestehender geschlechtsspezifischer Ungleichheiten.[102] Amazons zwischenzeitlich eingestelltes Recruiting-System bewertete Bewerbungen von Frauen ebenfalls schlechter, da das Modell mit historisch männlich dominierten Daten trainiert worden war.[103]

US-amerikanische Risikobewertungssysteme in der Strafjustiz zeigten, dass indirekte Variablen wie Wohnort oder sozioökonomischer Status ethnische Minderheiten benachteiligten, selbst wenn explizite Merkmale wie „*race*“ entfernt wurden.[104] Auch das einleitend genannte niederländische SyRI-System, das Sozialhilfemissbrauch erkennen sollte, diskriminierte sozial benachteiligte Gruppen.[105] Digitale Plattformen können ebenfalls algorithmische Diskriminierung verursachen: TikTok sperrte zeitweise Inhalte von Menschen mit Behinderung unter dem Vorwand, sie so vor möglichem Mobbing schützen zu wollen.[106] Auch Filterverfahren, die gegen „anstößige“ oder „unverständliche“ Begriffe vorgehen sollen, können Sprachräume marginalisierter Gruppen, etwa der LGBTIQ+-Community, unterdrücken.[107] Wenn vielfältige Ausdrucksformen nicht in das KI-Training mit einfließen, bilden die Systeme die gesellschaftliche Diversität nicht angemessen ab und Minderheiten, die

[99] *Kempt*, 2025, S. 93 ff.
[100] *Toju*, S. 67 ff.; zur intersektionellen bzw. Mehrfachdiskriminierung *Woll*, 2024, S. 201 ff.
[101] *Kempt*, 2025, S. 93 ff.
[102] *Toju*, S. 11-14.
[103] *Bender et al.,* S. 103.
[104] *Domalewska et al*., S. 104 f.
[105] Ebd. S. 103.
[106] *Hern*, TikTok owns up to censoring some users' videos to stop bullying, https://www.theguardian.com/technology/2019/dec/03/tiktok-owns-up-to-censoring-some-users-videos-to-stop-bullying (letzter Zugriff am 02.11.2025).
[107] *Bender et al.,* S. 613 f.

ohnehin für ihre Rechte und Sichtbarkeit kämpfen müssen, werden aus dem digitalen Raum algorithmisch verdrängt.

Besonders risikoreich sind KI-Systeme hinsichtlich der Reproduktion bestehender rassistischer Strukturen.[108] KI-Systeme zur Erkennung von Hautkrebs erzielen bei dunkelhäutigen PatientInnen z.B. eine geringere Trefferquote, weil die Trainingsdaten überwiegend Fälle mit heller Haut umfassen.[109] Solche Verzerrungen spiegeln auch die Ungerechtigkeit wider, dass klinische Forschung historisch überwiegend an weißen Männern durchgeführt wurde. Frauen und ethnische Minderheiten sind in medizinischen Studien bis heute unterrepräsentiert.[110] Diese Schieflage überträgt sich auf KI-basierte Diagnosesysteme, die dadurch für viele PatientInnengruppen schlechtere, beziehungsweise für die ohnehin schon bevorteilte Gruppe deutlich bessere Ergebnisse liefern.

Eine Untersuchung von ProPublica zeigte 2016, dass das KI-System COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), das die Rückfallwahrscheinlichkeit von StraftäterInnen einschätzt, schwarze StraftäterInnen häufiger fälschlicherweise als hochrisikobehaftet einstufte, während weiße tendenziell als niedrigrisikobehaftet klassifiziert wurden.[111] Auch *Predictive Policing*, also die Vorhersage von Tatorten oder potenziellen Straftaten durch KI-Systeme, verstärkt tendenziell die Überwachung von bereits stark polizeilich kontrollierten Nachbarschaften, da sie auf Grundlage historischer Daten erstellt werden. Das Problem: Es entsteht ein Rückkopplungseffekt, denn mehr Polizeipräsenz führt zu mehr erfassten Delikten in den bereits stark überwachten Regionen. Diese höhere Zahl registrierter Vorfälle bestätigt die algorithmischen Vorhersagen, obwohl sie eine Konsequenz der intensiveren Kontrolle ist.[112]

Das Risiko von algorithmischer Diskriminierung ergibt sich auch beim Einsatz von KI-Systemen zur Unterstützung der „*Refugee Status Determination*“ in Asyl- und Flüchtlingsverfahren. Dabei wird durch das System z.B. die Identität geprüft, es werden digitale Beweise ausgewertet oder auch Interviews zusammengefasst und dortige Aussagen auf

---

[108] *UN*, Racism and AI-Bias from the past leads to bias in the future, https://www.ohchr.org/en/stories/2024/07/racism-and-ai-bias-past-leads-bias-future (letzter Zugriff am 19.08.2025).
[109] *Rezk*, Improving Skin Color Diversity in Cancer Detection: Deep Learning Approach, https://pmc.ncbi.nlm.nih.gov/articles/PMC10334920/ (letzter Zugriff am 10.10.2025).
[110] *Bibbins-Domingo, Helman*, Why Diverse Representation in Clinical Research Matters and the Current State of Representation within the Clinical Research Ecosystem, https://www.ncbi.nlm.nih.gov/books/NBK584396/ (letzter Zugriff am 20.11.2025).
[111] *Toju*, S. 20.
[112] *McGurk, Tomlinson*, S. 21 f.; *Almasoud1, Idowu*, Algorithmic fairness in predictive policing, https://link.springer.com/content/pdf/10.1007/s43681-024-00541-3.pdf (letzter Zugriff am 29.11.2025).

Glaubwürdigkeit geprüft. Die Verfahren beziehen sich oft auf *Country-of-Origin*-Informationen und andere Quellen, deren Qualität und Vollständigkeit variiert. Aufgrund der epistemischen Unsicherheit und lückenhafter Daten besteht die Gefahr von ungerechten Entscheidungen im Asyl- und Flüchtlingsverfahren,[113] und es droht eine Automatisierung von Willkür, während die mangelnde Transparenz Rechenschaftspflicht und effektive Rechtsmittel erschweren.[114]

Gesellschaftliche Bewegungen wie *Black Lives Matter* werden in statischen Datensätzen häufig nur unvollständig oder verzerrt abgebildet, denn Datensätze reproduzieren vor allem das, wovon bereits eine große Menge vorhanden ist, wodurch dominante oder veraltete Perspektiven in KI-Systemen verstärkt dargestellt werden.[115] Statt reiner Datenmenge sollte deshalb die gezielt inklusive Auswahl von Trainingsdaten priorisiert werden, beispielsweise auch durch die Einbeziehung mündlicher Überlieferungen oder alternativer Kommunikationsformen, um so westlich geprägte Textkulturen aufzubrechen.[116]

Die dargestellten Fälle zeigen, dass algorithmische Verzerrungen von KIs konkrete Menschenrechtsverletzungen bewirken können. Betroffen sind insbesondere das Recht auf Gleichbehandlung und Nichtdiskriminierung (Art. 14 EMRK, Art. 21 und 23 GrCh). Wenn Systeme wie SyRI umfangreiche personenbezogene Daten sammeln und auswerten – teils auch noch ohne Wissen oder Einwilligung der Betroffenen – verletzten sie auch das Recht auf Achtung des Privat- und Familienlebens sowie auf Schutz personenbezogener Daten (Art. 8 EMRK, Art. 7 und 8 GrCh). Das Recht auf ein faires Verfahren (Art. 6 EMRK, Art. 47 GrCh) kann verletzt werden, wenn automatisierte Entscheidungsprozesse in Asylverfahren oder *Predictive Policing* intransparent agieren (etwa bei der Erstellung KI-basierter Risikoprofile) und den Betroffenen die Möglichkeit entziehen, Entscheidungen nachvollziehbar anzufechten. Ebenso könnten die Meinungs- und Informationsfreiheit (Art. 10 EMRK, Art. 11 GrCh) tangiert sein, wenn seitens staatlicher Stellen KI-gestützte Moderationssysteme oder Filter die Sichtbarkeit bestimmter Gruppen reduzieren, wodurch diese Gruppen in ihrer Möglichkeit zur öffentlichen Meinungsäußerung eingeschränkt werden. Dies verdeutlicht, dass KI-

[113] *UNHCR*; Refugee Status Determination, https://www.unhcr.org/what-we-do/protect-human-rights/protection/refugee-status-determination (letzter Zugriff am 28.11.2025).
[114] *Costello, Dukovic*, Introduction to the Symposium on Algorithmic Fairness for Asylum Seekers and Refugees, Verfassungsblog/2025.
[115] *Toju*, S. 96-99.
[116] *Bender et al.,* S. 614.

Anwendungen systematisch auf potenzielle Diskriminierung und Verletzungen fundamentaler Rechte überprüft werden müssen.

Technologischer Fortschritt garantiert keinen menschenrechtlichen Fortschritt. Die Entwicklung von KI hätte es möglicherweise erlaubt, Verzerrungen menschlichen Handelns abzumildern, da algorithmische Systeme potenziell objektiv operieren können. In der Praxis jedoch spiegeln sie große Teile menschlicher Subjektivität und Fehler wider.

II. Sicherheits- und Umweltrisiken

KI wird zunehmend in bewaffneten Konflikten eingesetzt. Im Gazastreifen kommt z.B. das Zielsystem *Lavender* bei Angriffen zum Einsatz. *Lavender* wertet große Datenmengen aus, um potenzielle Hamas- oder PIJ-Kämpfer zu identifizieren.[117] Dabei arbeitet es z.B. mit Daten zu Mitgliedschaften in Chatgruppen, häufigem Wechsel von Telefonnummern oder des Wohnorts.[118] Diese KI wurde auch in Verbindung mit dem System *Where's Daddy* eingesetzt, das gezielt nach den identifizierten Personen sucht, während sie sich in ihren Privathaushalten aufhalten, da sie dort leichter angegriffen werden können.[119]

In den ersten Wochen des Einsatzes wurden Berichten zufolge ca. 37.000 PalästinenserInnen als potenzielle Kämpfer markiert. Für jeden niederrangigen Kämpfer wurden 15-20 zivile Opfer und für hochrangige Kommandeure über 100 zivile Opfer in Kauf genommen.[120] Menschliche Offiziere überprüften die KI-Entscheidungen nur für durchschnittlich 20 Sekunden pro Ziel und bestätigten häufig lediglich das Geschlecht der Zielperson, bevor der Anschlag freigegeben wurde.[121] Teilweise bestand auch ein großer zeitlicher Abstand zwischen der Meldung von *Where's Daddy* und der Bombardierung, was zur Tötung ganzer Familien führen konnte, während die ursprünglich erfasste Person nicht mehr anwesend war.[122] Fehlerquoten von etwa

[117] *Wiese*, Gaza, Artificial Intelligence, and Kill Lists, https://verfassungsblog.de/gaza-artificial-intelligence-and-kill-lists/ (letzter Zugriff am 10.11.2025).
[118] *Völkerrechtsblog*, Smarte Kriege: Künstliche Intelligenz in bewaffneten Konflikten, https://voelkerrechtsblog.org/de/38-smarte-kriege-kuenstliche-intelligenz-in-bewaffneten-konflikten/ (letzter Zugriff am 10.11.2025).
[119] *Human Rights Watch*, Questions and Answers: Israeli Military's Use of Digital Tools in Gaza, https://www.hrw.org/news/2024/09/10/questions-and-answers-israeli-militarys-use-digital-tools-gaza (letzter Zugriff am 09.11.2025).
[120] https://www.tagesschau.de/ausland/israel-gazastreifen-ki-102.html.
[121] *Wiese*, Gaza, Artificial Intelligence, and Kill Lists, https://verfassungsblog.de/gaza-artificial-intelligence-and-kill-lists/ (letzter Zugriff am 10.11.2025).
[122] Ebd.

zehn Prozent führten dazu, dass auch gänzlich unbeteiligte Personen markiert wurden, was die Wahrscheinlichkeit zivilen Leids erheblich erhöhte.[123]

Dieser Konflikt findet außerhalb des Zuständigkeitsbereichs der EU statt, und die Verantwortung für den Schutz der Menschenrechte liegt bei nationalen Regierungen oder internationalen Menschenrechtsinstanzen, im Sinne einer ethischen und menschenrechtlichen Bewertung sollten dennoch die Menschenrechte, die durch den Einsatz von KI-Systemen im Krieg verletzt werden, genannt werden. Der Einsatz von Lavender und Where's Daddy verletzt das Recht auf Leben (Art. 2 EMRK, Art. 2 GRCh) sowie die Menschenwürde (Art. 1 GRCh). Auch das Verbot unmenschlicher oder erniedrigender Behandlung (Art. 3 EMRK, Art. 4 GRCh) wird tangiert. Der virtuelle Charakter KI-gestützter Zielsysteme erschwert zusätzlich die Kontrolle über militärische Handlungen. Ihre Black-Box-Natur macht die Entscheidungsgrundlagen intransparent und führt zu unklarer Verantwortlichkeit. Tötungen können nicht eindeutig einer Person zugeordnet werden, wodurch das Recht auf wirksame Rechtsmittel (Art. 13 EMRK) und die Gewährleistung fairer Gerichtsverfahren (Art. 6 EMRK, Art. 47 GRCh) verletzt werden können. Da KIs nicht in der Lage sind, moralische Abwägungen vorzunehmen oder Empathie zu zeigen, werden Menschen zu Objekten algorithmischer Kalkulation degradiert. Auch wahren KI-gestützte Waffensysteme nicht die Prinzipien des humanitären Völkerrechts,[124] besonders weil KI-gestützte Zielsysteme das Unterscheidungsprinzip zu verletzen riskieren, da sie unzuverlässig zwischen Kombattanten und Zivilpersonen unterscheiden und Abwägungen von militärischem Vorteil versus zivilem Kollateralschaden rein algorithmisch treffen.

KIs erfordern enorme Rechenleistung, was zu hohem Energie- und Ressourcenverbrauch führt und negativ zur Umweltverschmutzung und der Klimakrise beiträgt: So prognostiziert eine von Greenpeace beauftragte Studie des Öko-Instituts in Berlin, dass der Stromverbrauch von auf KI spezialisierten Datenzentren bis 2030 von derzeit rund 50 Milliarden kWh auf etwa 550 Milliarden kWh steigen könnte, wobei die Treibhausgas-Emissionen dieser Rechenzentren von 212 Millionen Tonnen CO_2 im Jahr 2023 auf 355 Millionen Tonnen im Jahr 2030 zunehmen könnten.[125] Zusätzlich fallen erhebliche Mengen an Kühlwasser, Elektronikschrott

[123] *Yuval*, Lavendel-Die KI-Maschine, die Israels Angriffe in Gaza lenkt, https://www.nd-aktuell.de/artikel/1181223.gaza-krieg-lavendel-die-ki-maschine-die-israels-angriffe-in-gaza-lenkt.html (letzter Zugriff am 30.11.2025).
[124] *Bundesministerium für Verteidigung*, Humanitäres Völkerrecht in bewaffneten Konflikten, https://www.bmvg.de/resource/blob/93612/7d6909421eacad4ddc7dcdfdf58d42ca/b-02-02-10-download-handbuch-humanitaeres-voelkerrecht-in-bewaffneten-konflikten-data.pdf (letzter Zugriff am 29.11.2025).
[125] *Gröger et al.,* S. 6.

sowie Stahl in mehreren hundert Kilotonnen an.[126] Ein Großteil der für KI-Systeme eingesetzten Energie stammt außerdem nicht aus erneuerbaren Quellen, auch wenn Systeme gerne damit beworben werden: Der hohe Bedarf großer Rechenzentren nutzt erhebliche Anteile erneuerbarer Energie und verdrängt deren Nutzung in anderen Bereichen. Die Rede von „grüner KI“ kann daher meistens als *Greenwashing* eingestuft werden, während die massive Umweltbelastung bestehen bleibt.[127]

Auch wird KI vermehrt für den Tiefseebergbau eingesetzt. KIs beschleunigen beispielsweise durch Datenanalyse oder automatisierte Steuerung von Fahrzeugen den Abbau von Manganknollen, was zur Belastung und Zerstörung sensibler Ökosysteme führt.[128]

Darüber hinaus entwickeln und nutzen KI-Systeme Algorithmen, die Nutzerverhalten analysieren und vorhersagen, um gezielt die Aufmerksamkeit und den Konsum zu steigern. Indem diese Algorithmen hinsichtlich des Wachstums und das Engagements der NutzerInnen optimiert werden, erhöhen sie die Nachfrage und verstärken kapitalistische Konsummuster und tragen zu einem höheren Ressourcenverbrauch bei.[129]

Diese Umwelt- und Ressourcenauswirkungen von KI berühren jedenfalls mittelbar Grundrechte, insbesondere das Recht auf Leben (Art. 2 EMRK, Art. 2 GRCh) und körperliche Unversehrtheit (Art. 3 GRCh), da Umweltbelastungen durch Schadstoffe, Klimafolgen oder zerstörte Ökosysteme Gefahren für die die körperliche und psychische Gesundheit der Bevölkerung darstellen können. Vor allem müssen Staaten im Sinne ihrer Schutzpflicht und nach Art. 37 GRCh ein „ein hohes Umweltschutzniveau und die Verbesserung der Umweltqualität (…) in die Politik der Union“ mit einbeziehen, und „nach dem Grundsatz der nachhaltigen Entwicklung“ sicherstellen. Auch die langfristige Sicherung der Lebensgrundlagen zukünftiger Generationen wird dadurch tangiert.[130] Zu erwähnen ist in diesem Zusammenhang aber vor allem die Grundsatzentscheidung des EGMR aus 2024,

[126] Ebd. S. 7.
[127] *Bender et al.,* S. 612.
[128] *McKee*, Seltene Erden: Wie künstliche Intelligenz den Bedarf steigert und den Abbau beschleunigt, https://te.ma/art/blvifm/mckie-ki-tiefseebergbau-seltene-erden/#paper (letzter Zugriff am 14.11.2025); *Cyrill*, Deep-Sea Mining: A noisy affair, Overview and Recommendations; *Oceancare*, Meeresschutz mit globaler Wirkung, https://www.oceancare.org/wp-content/uploads/2021/12/Deep-Sea-Mining_A-noisy-affair_Report-OceanCare_2021.pdf (letzter Zugriff am 15.11.2025).
[129] *Domalewska et al*., S. 203 ff.
[130] Vgl. ICESCR, Art. 6, 11, 12; Charta der Vereinten Nationen; Rio-Deklaration über Umwelt und Entwicklung, Art. 1 und 3.

wonach unzureichende Klimaschutzmaßnahmen eines Konventionsstaats die EMRK (insbesondere Art. 8) verletzen können.[131]

Dieser Fall hat gezeigt, dass Staaten ihre Klimaschutzpflichten ernst nehmen müssen: Die Schweiz wurde verurteilt, weil sie zu wenig gegen die Folgen des Klimawandels unternahm und dadurch die Rechte älterer BürgerInnen verletzte.[132] Fragen rund um Klimaschutz und Ressourcenmanagement müssen also in die Debatten um eine europäische KI-Regulierung integriert werden.

III. Gesundheitliche, kognitiv-psychische und demokratiebezogene Risiken

Neben den bisher diskutierten Risiken für die Gesamtgesellschaft oder einzelne Gruppen birgt der Einsatz von KI auch erhebliche Gefahren für die psychische Gesundheit des Individuums, mit potenziell weitreichenden Auswirkungen auf demokratische Gesellschaften.

Die zwanghafte Nutzung digitaler Inhalte, auch als *Problematic Internet Use* oder *Problematic Smartphone Use* bezeichnet, zeigt sich in digitalen Umgebungen, wie Online-Shopping, Social-Media-Aktivitäten oder Glücksspiel, die exzessiv genutzt werden, und kann sich zu einer Verhaltenssucht entwickeln.[133] Entsprechende KIs gestalten mediale Inhalte zusätzlich personalisiert, was digitale Reize und die Wirkung auf das Belohnungssystem steigern. Studien zeigen, dass dies ähnliche Folgen wie eine Substanzabhängigkeit aufweisen kann. Die Folgen können sowohl funktionelle als auch strukturelle Veränderungen im Gehirn sein: Das dopamingesteuerte Belohnungssystem gerät aus dem Gleichgewicht, wodurch Impulskontrolle und Entscheidungsfähigkeit geschwächt werden. Zudem zeigen sich Beeinträchtigungen in Hirnregionen, die für Planung, Selbstkontrolle und die Anpassung an neue Situationen verantwortlich sind.[134] Dies kann zu psychischen Belastungen wie Reizbarkeit, depressiven Symptomen, Angstzuständen und emotionaler Dysregulation führen,[135] während sich körperliche Folgen z.B. in Schlafstörungen, verminderter Aufmerksamkeit sowie Lern- und Gedächtnisproblemen äußern.[136]

---

[131] Vgl. *EGMR*, Urt. v. 9. April 2024, *VEREIN KLIMASENIORINNEN SCHWEIZ AND OTHERS v. SWITZERLAND*, Application no. 53600/20.
[132] *Ibid.*
[133] *Montag et al*., JOBA, 9/2021, S. 909 ff.
[134] *Donnelly, Kuss*, JOAPM, 1/ 2021, S. 1 ff.; *Domalewska et al*., S. 114-117.
[135] *Lee et. al*., CJOP 61/2016, S. 243, 246 ff.; *Kaess*, Pathological Internet use among European adolescents: psychopathology and self-destructive behaviours, https://pmc.ncbi.nlm.nih.gov/articles/PMC4229646/pdf/787_2014_Article_562.pdf (letzter Zugriff am 23.11.2025).
[136] *Verduyn et. al.,* SIPR, 11/2017, S. 285 ff.

Obwohl die Nutzung von KI-basierten Systemen im Online-Bereich in erster Linie private Entscheidungen betrifft, können die entstehenden psychischen und körperlichen Risiken mittelbar die Schutzpflichten des Staates berühren, z.B. im Hinblick auf das Recht auf physische und psychische Gesundheit (Art. 35 GrCh). Ebenso kann die Datenerhebung zur Erstellung der Algorithmen potenziell in die Rechte auf Achtung des Privat- und Familienlebens sowie den Schutz persönlicher Daten eingreifen (Art. 8 EMRK, Art. 7 und 8 GrCh).

Was an dieser Thematik so wichtig ist, sind die langfristigen Folgen für die demokratische Gesellschaft. Die ständig verfügbaren und vielfältigen Informationen verstärken den sogenannten *Information Overload*, der die Fähigkeit der NutzerInnen beeinträchtigt, Informationen kritisch zu bewerten, längere Aufmerksamkeitsspannen aufrechtzuerhalten oder fundierte Entscheidungen zu treffen.[137] Folgen könnten sein, dass der Überfluss an Informationen wesentliche gesellschaftliche Fragen verschleiert, was zu Desorientierung und Entpolitisierung führen kann. Ein anschauliches Beispiel hierfür lieferte Donald Trump zu Beginn seiner zweiten Amtszeit: Die schiere Menge an Exekutivverordnungen, 139 in den ersten 100 Tagen, spontane Politikwechsel und Rücknahmen bereits getroffener Entscheidungen erschwerten es BeobachterInnen, den Überblick zu behalten, fundierte Analysen zu erstellen oder Kritik wirksam zu formulieren.[138]

Die genannten Beeinträchtigungen kognitiver Leistungsfähigkeit und kritischer Urteilsfähigkeit stellen auch ein direktes Risiko für Gesellschaft und Demokratie dar. Menschen verlassen sich vermehrt auf algorithmische Filtermechanismen wie personalisierte Social-Media-Seiten oder Suchmaschinen, welche Inhalte nach Interaktionswahrscheinlichkeit und Profitabilität priorisieren.[139] Dies erzeugt Echokammern, in denen NutzerInnen vor allem auf Meinungen treffen, die ihr bestehendes Weltbild bestätigen.[140] Auch datenbasiertes *Microtargeting* in der politischen Kommunikation kann sich dies in demokratiegefährdender Weise zu Nutze machen.[141] Auf Grundlage umfangreicher personenbezogener Daten werden politische Persönlichkeitsprofile generiert. Diese Systeme analysieren das Onlineverhalten nach Mustern, Emotionen und Vulnerabilitäten und erstellen daraus detaillierte Modelle politischer

[137] *Babik*, in Ribeiro, Cerveira (Hrsg.), S. 290; *Domalewska et al*., S. 166 ff.
[138] *Sirakov*, Das System als Endgegner, https://www.bpb.de/themen/nordamerika/usa/561711/das-system-als-endgegner/ (letzter Zugriff am 03.11.2025).
[139] *Kelm et al.,* CHBR, 2023, S. 1f; *Domalewska et al*., S. 268 ff.
[140] *Denniss, Lindberg*, S. 5.; *Harari*, S. 7.
[141] *Bayer,* Double harm to voters: data-driven micro-targeting and democratic public discourse, https://policyreview.info/articles/analysis/double-harm-voters-data-driven-micro-targeting-and-democratic-public-discourse (letzter Zugriff am 30.10.2025).

Präferenzen.[142] Die daraus abgeleiteten Botschaften werden anschließend durch KI-optimierte Ausspielungsmechanismen individuell auf digitalen Plattformen verbreitet, häufig automatisiert durch politische Bots oder andere algorithmische Kommunikationsinstrumente.[143] Studien zeigen, dass solche personalisierten und meist emotional geladenen Wahlbotschaften im Wahlkampf wirksamer sind als generische Kampagnen.[144] In den letzten zehn Jahren hat russische Propaganda im Ukraine-Konflikt deutlich gemacht, wie gezielte Desinformation demokratische Prozesse destabilisieren und politische Akteure delegitimieren kann: Groß angelegte und teils durch KI personalisierte Desinformationskampagnen erreichten in Sekunden eine Millionenreichweite, Mustererkennung und Targeting erhöhten die Treffsicherheit psychologischer Ansprache, synthetische Medien erstellt durch Deepfakes (mittels KI) untergruben Vertrauen in Quellen und automatisierte Bot-Netze manipulierten Diskurse in Echtzeit.[145]

Dieser gezielte Einsatz von Informationen zur Beeinflussung oder Kontrolle politischer und gesellschaftlicher Prozesse ist auch ein Mittel der psychologischen Kriegsführung: damit droht – und findet teilweise bereits statt – der sogenannte *Information Warfare*.[146]

Der Einsatz der genannten KI-Systeme tangiert mehrere Grundrechte, soweit staatliche Stellen oder EU-Rechtsakte involviert bzw. betroffen sind und damit der sachliche Anwendungsbereich des Unionsrechts eröffnet wird (Art. 51 Abs. 1 GRCh): die Meinungs- und Informationsfreiheit (Art. 10 EMRK, Art. 11 GrCh), da algorithmische Filtermechanismen und personalisiertes Microtargeting die Vielfalt verfügbarer Informationen einschränkt. Auch die Privatsphäre (Art. 8 EMRK, Art. 7 und 8 GrCh) ist betroffen, weil zur Erstellung detaillierter Persönlichkeits- und Präferenzprofile umfangreiche personenbezogene Daten erhoben und analysiert werden – möglicherweise ohne bewusste Einwilligung der Betroffenen. Demokratische Rechte, insbesondere das Recht auf freie Wahlen, können ebenfalls in diesem Zusammenhang erwähnt werden (Art. 3 ZP Nr. 1 EMRK, Art. 39 Abs. 2 GrCh). Hochgradig individualisierte, KI-gestützte Wahlkampfmaßnahmen dienen nicht der Meinungsbildung,

[142] *Bennett*, IDPL, 6(4) /2016, S. 261, 265.
[143] Ebd.
[144] *Shirado, Christakis,* Nature, 545/2017, S. 371; *Schröder et al*. How Malicious AI Swarms Can Threaten Democracy The Fusion of Agentic AI and LLMs Marks a New Frontier in Information Warfare, https://arxiv.org/html/2506.06299v3#bib.bib1 (letzter Zugriff am 04.11.2025); *Daum,* KI als neues Wahlkampfinstrument, Verfassungsblog/2024.
[145] *Carmona et. al.,* EIME, 23 (2)/2024, S. 105, 113; *Domalewska et al*., S. 185-189.
[146] *Ruhmann, Bernhardt*, S.71ff; *Honigberg*, The Existential Threat of AI-Enhanced Disinformation Operations, https://www.justsecurity.org/82246/the-existential-threat-of-ai-enhanced-disinformation-operations/ (letzter Zugriff am 04.11.2025).

sondern sind ein manipulatives Instrument, vor allem wenn WählerInnen über deren Einsatz nicht aufgeklärt werden.

F. Bewertung der europäischen Schutzmechanismen

I. Übertragbarkeit der Regelwerke auf die identifizierten Risikofelder

Nachdem die potenziellen menschenrechtsfördernden Wirkungen sowie die menschenrechtlichen Gefährdungen von KI jeweils exemplarisch beleuchtet wurden, richtet sich der Fokus nun darauf, inwieweit das Rahmenübereinkommen des Europarats und die KI-VO geeignet sind, den identifizierten Risiken wirksam zu begegnen.

1. Schutz vor algorithmischer Diskriminierung

Die KI-VO adressiert das Risiko algorithmischer Diskriminierung auf mehreren Ebenen. Art. 5 Abs. 1 g) verbietet KI-Systeme zur biometrischen Kategorisierung von Personen, wenn dabei Rückschlüsse auf besonders sensible Merkmale wie „politische Meinungen, Gewerkschaftszugehörigkeit, religiöse oder philosophische Überzeugungen, ethnische Herkunft, oder sexuelle Orientierung“ gezogen werden können. Diese Bestimmung wurde erst 2023 auf Initiative des Europäischen Parlaments hin ergänzt und dürfte maßgeblich durch die Datenschutzdebatten des Jahres 2022 beeinflusst worden sein – insbesondere durch die Empfehlungen des Europäischen Datenschutzausschusses, der ein Verbot der gruppenbezogenen Einordnung von Personen anhand KI-gestützter Gesichtserkennung, gefordert hatte.[147]

Aufgrund ihres hohen Diskriminierungspotenzials stuft die KI-VO verschiedene Systeme als Hochrisiko-Systeme ein. Dazu gehören medizinische Diagnosesysteme, bildungsbezogene KIs, die über Zugang oder Bewertung entscheiden (Annex III Nr. 3), Systeme im Beschäftigungsbereich wie Rekrutierungs- oder Leistungsbewertungssoftware (Nr. 4) sowie KIs, die über essenzielle öffentliche oder private Leistungen wie Gesundheitsversorgung, Sozialhilfe oder Kreditwürdigkeit entscheiden (Nr. 5). Auch Anwendungen im Bereich Migration, Asyl und Grenzmanagement (Nr. 7) fallen darunter. Diese könnten historische Diskriminierungsmuster wiederholen oder neue derartige Formen kreieren (Erwägungsgrund 58). Neben Anforderungen an Risikomanagement, Transparenz und menschliche Aufsicht verlangt Art. 27 bei Hochrisiko-Systemen zudem eine Grundrechte-Folgenabschätzung

[147] *Wendehorst* in Martini/Wendehorst (Hrsg), KI-VO1, Art 5 KI-VO, Rz 120.

(Fundamental Rights Impact Assessment, "*FRIA"*), die mögliche Eingriffe in diskriminierungsrelevante Rechte vor der Inbetriebnahme bewertet. Dieses Konzept lehnt sich an das genannte HUDERIA-Modell des Europarats an, das nicht nur für Hochrisiko-Systeme gelten, sondern im gesamten Lebenszyklus und für alle Arten KI-basierter Technologien Anwendung finden sollte und neben Grundrechten auch Aspekte von Demokratie und Rechtsstaatlichkeit berücksichtigt hätte.[148]

Laut Art. 10 KI-VO müssen Systeme auf Verzerrungen geprüft werden, welche die Gesundheit und Sicherheit von Personen gefährden, fundamentale Rechte beeinträchtigen oder zu Diskriminierung nach Unionsrecht führen können, z.B. im Hinblick auf geschlechtsspezifische, ethnische oder soziale Merkmale. Erwägungsgrund 27 der KI-VO betont zudem, dass KI-Anwendungen Diversität, Gleichberechtigung und kulturelle Vielfalt fördern sollen und diskriminierende Auswirkungen aktiv zu vermeiden sind. Für die Durchsetzung erhalten die nationalen Aufsichtsbehörden umfassende Einsichtsrechte in technische Dokumentation und Protokolldaten (Art. 77).

Das Rahmenübereinkommen des Europarats ergänzt diesen Schutz gegen algorithmische Diskriminierung der KI-VO: Art. 10 und Art. 17 verpflichten die Vertragsstaaten, Gleichheit und Nichtdiskriminierung sicherzustellen, während Art. 18 die Rechte von Kindern und Menschen mit Behinderungen besonders berücksichtigt.

2. Schutz vor Umwelt-, Gesundheits- und Sicherheitsrisiken

Umwelt- und Nachhaltigkeitsaspekte werden in der KI-VO in verschiedenen Erwägungsgründen[149] sowie in Art. 1 und Art. 112 Abs. 7 thematisiert: KI-Systeme sollen nachhaltig und umweltfreundlich entwickelt und betrieben werden, und ihre langfristigen Auswirkungen auf Individuen, Gesellschaft, Demokratie und Umwelt sollen überwacht und bewertet werden. Annex II verweist zudem auf die Einordnung von Umweltstraftaten, während freiwillige Verhaltenskodizes Anreize für zusätzliche Umweltanforderungen schaffen sollen. Das Rahmenübereinkommen des Europarats unterstützt dies und verpflichtet die Vertragsstaaten nach Art. 19 dazu, nicht nur soziale, wirtschaftliche, rechtliche und ethische, sondern auch ökologische Implikationen von KI durch öffentliche Diskussionen und Multi-Stakeholder-Konsultationen zu berücksichtigen.

[148] *Levantino, Paolucci*, in: Czech/Heschl/Nowak et al. (Hrsg.), S. 14; *Woll*, ZEuS 3/2025, S. 509, 522.
[149] Z.B. Erwägungsgründe 27, 130 und 165.

Die Risiken von KI-Systemen für die Gesundheit, insbesondere von Kindern und Jugendlichen, werden in den Regelwerken kaum durch spezifische Schutzvorgaben adressiert. Erwägungsgründe 28 und 48 nennen „manipulative, ausbeuterische und soziale Kontrollpraktiken", und Art. 9 Abs. 9 verpflichtet Anbieter von Hochrisiko-KI-System zur Prüfung, ob das System nachteilige Auswirkungen auf Personen unter 18 Jahren und gegebenenfalls auf andere gefährdete Gruppen haben kann – was eine sehr vage Schutzebene bietet, gerade verglichen mit anderen Schutzstandards, etwa in Australien, wo die Nutzung sozialer Medien für Jugendliche unter 16 Jahren seit diesem Jahr gesetzlich verboten ist.[150]

Menschenrechtsverletzungen durch KI im Kriegseinsatz werden aufgrund der Ausnahmen für nationale Sicherheits- und Militäreinsätze nicht abgedeckt. Militärische Risiken werden daher nur indirekt über Erwägungsgrund 110 zu „systemischen Risiken" erwähnt. Danach können *GPAI*-Modelle negative Auswirkungen auf die Entwicklung chemischer, biologischer, radiologischer oder nuklearer Waffen, auf die öffentliche Gesundheit und Sicherheit oder auf den Betrieb kritischer Infrastrukturen haben. Ansatzpunkte für eine normative Anerkennung des Risikos bietet auch Art. 3 Abs. 2 des Europaratsrahmenübereinkommens, der bei nationalen Sicherheitsinteressen eine Nichtanwendung des Übereinkommens zulässt, gleichzeitig aber verlangt, dass Aktivitäten „in Übereinstimmung mit anwendbarem Völkerrecht, einschließlich der Verpflichtungen aus den internationalen Menschenrechten, und unter Achtung demokratischer Institutionen und Verfahren" erfolgen.

3. Schutz der Privatsphäre und demokratischer Entscheidungsprozesse

Echtzeit-Fernidentifizierung in öffentlich zugänglichen Räumen für polizeiliche Zwecke fällt nach Art. 5 der KI-VO unter die verbotenen KI-Praktiken. Es sind nur sehr begrenzte Ausnahmen zulässig, etwa bei der Suche nach Entführungsopfern oder zur Abwehr einer akuten Gefahr für Leben und Gesundheit. Jede Nutzung bedarf einer Genehmigung durch eine Justiz- oder unabhängige Verwaltungsbehörde, muss eine Grundrechte-Folgenabschätzung (*FRIA* nach Art. 27) durchlaufen und in der EU-Datenbank registriert werden. Die nachträgliche biometrische Identifizierung fällt hingegen nach Art. 6 und Annex III unter die Hochrisiko-KI-Systeme und unterliegt damit den Pflichten aus Art. 9 ff., insbesondere Dokumentations- und Protokollierungspflichten, Risikomanagement und menschlicher Aufsicht.[151]

---

[150] *Bundeszentrale für politische Bildung*, Social-Media-Verbot für Jugendliche unter 16 Jahren in Australien, https://www.bpb.de/kurz-knapp/taegliche-dosis-politik/557196/social-media-verbot-fuer-jugendliche-unter-16-jahren-in-australien/ (letzter Zugriff am 20.08.2025).

[151] Vgl. *Anderl/Ciarnau* in Zankl (Hrsg), KI-VO1, Art. 6, Rz 20.

Das Auslesen und Analysieren von Gesichtern anhand von Daten aus dem Internet oder von Überwachungskameras ist ebenfalls verboten (Art. 5 Abs. 1). Diese Regelung wurde Mitte 2023 aufgenommen, um das Risiko von breitflächigem *Gesichtsscraping*[152] zu adressieren, also dem automatisierten Sammeln und Speichern von Gesichtsbildern aus Onlinequellen; nach Erwägungsgrund 43 verstärken solche Methoden „das Gefühl der Massenüberwachung" und können zu schwerwiegenden Grundrechtsverletzungen, insbesondere des Rechts auf Privatsphäre, führen. Auch der Einsatz von KI zur Erfassung und Analyse emotionaler Zustände in Bildungseinrichtungen oder am Arbeitsplatz ist nach Art. 5 Abs. 1 untersagt und wirkt damit potenziellen *Social Scoring*-Technologien entgegen. Diese Schutzebene kann auch als gesundheitsschützende Maßnahme interpretiert werden, da sie psychischen Belastungen durch umfassende Überwachung entgegenwirkt.[153]

Hinsichtlich des Schutzes vor Wahlmanipulation betont das Rahmenübereinkommen in der Präambel, dass KI-Systeme für repressive Zwecke missbraucht werden können, z.B. durch willkürliche oder rechtswidrige Überwachungs- und Zensurpraktiken, die Privatsphäre und individuelle Autonomie untergraben. Art. 5 Abs. 2 fordert weiter von jeder Vertragspartei Maßnahmen zu ergreifen, die darauf abzielen, ihre demokratischen Prozesse im Zusammenhang mit KI zu schützen, einschließlich des gerechten Zugangs und der Beteiligung von Einzelpersonen an öffentlichen Debatten sowie ihrer Fähigkeit, Meinungen frei zu bilden.

In der KI-VO erkennen Erwägungsgrund 62 sowie Annex III Nr. 8(b) ausdrücklich die Möglichkeit missbräuchlicher Nutzung von KI für manipulative, ausbeuterische oder sozialkontrollierende Praktiken an und stufen KI-Systeme, die Wahlen oder Referenden beeinflussen, deshalb als Hochrisikosysteme ein.

II. Kritik

Wie bereits an mehreren Stellen sichtbar wurde, weist die aktuelle Regulierung trotz ihres ambitionierten Anspruchs erhebliche Lücken auf. Im Folgenden werden jene Problemfelder skizziert, die als besonders prägnant erscheinen und dem Umfang der hiesigen Analyse angemessen sind. Auch wenn das Rahmenübereinkommen des Europarats in vielen Punkten Parallelen zur KI-VO aufweist, liegt der Fokus der Kritik auf letzterer, da sie nicht noch von der Ratifikation durch eine ausreichende Anzahl von Staaten abhängt, eine unmittelbare

[152] *McGurk, Tomlinson*, S. 15.
[153] *Daragh et.al*., JHRP, 16/2024, S. 397, 399 f.

Durchsetzungsmacht besitzt und wesentlich umfangreicher und technisch detaillierter als das Rahmenübereinkommen ist.

1. Grenzen des Hochrisikoschutzes angesichts technischer Realitäten

Zunächst zeigen sich die Grenzen des europäischen KI-Rechtsrahmens in jenen Bereichen, in denen reale Einsatzbedingungen mit den regulatorischen Vorgaben kollidieren.

Bei dem Realbeispiel einer KI-Anwendung zur Manipulation von Wählerverhalten zeigt sich, dass die Schutzmechanismen in vielen Gefahrenbereichen in der Praxis nicht greifen: Politische Chatbots, wie sie die „Alternative für Deutschland“ (AfD) bereits stark einsetzt und damit insbesondere die junge Wählerschaft erreicht, können Kommentarspalten innerhalb kürzester Zeit mit großen Mengen automatisierter Inhalte füllen, sodass selbst gelöschte Beiträge unmittelbar durch neue ersetzt werden, während mit Deepfakes angebliche und imageschädigende Aussagen der anderen Parteien verbreitet werden.[154] Im Echtzeitbetrieb ist zudem schwer nachvollziehbar, wer die Inhalte erstellt hat und damit für sie verantwortlich ist. Obwohl ein Großteil dieser politischen Beeinflussung durch das Ausnutzen von nicht als Hochrisiko-KIs einzustufenden KIs und der großen Reichweite und Kommunikationsstruktur von Plattformen wie Instagram oder TikTok geschieht, was nicht unter den Anwendungsbereich der KI-VO fällt, bleibt es unverständlich, warum die Verordnung Systeme, die gezielt zur Beeinflussung von Wahlen eingesetzt werden, lediglich als Hochrisiko klassifiziert und nicht vollständig verbietet. Warum sollte eine Technologie, die direkt die demokratische Willensbildung manipulieren kann, unter Auflagen erlaubt bleiben, statt konsequent ausgeschlossen zu werden?

Problematisch für den Schutz der Grundrechte ist auch die in der KI-VO weiterhin erlaubte nachträgliche biometrische Fernidentifizierung. Obwohl sie als Hochrisiko-System eingestuft ist und damit strengen Pflichten unterliegt, kann sie erhebliche Eingriffe in Privatsphäre und Datenschutz ermöglichen. Denn die gesetzliche Unterscheidung zwischen Echtzeit- und nachträglicher Erkennung mildert die Risiken nur formal: Auch eine erst später erfolgende Identifizierung erlaubt die Erstellung detaillierter Bewegungs- und Personenprofile, unterläuft

[154] *Daum*, KI als neues Wahlkampfinstrument, https://verfassungsblog.de/ki-als-neues-wahlkampfinstrument/ (letzter Zugriff am 15.11.2025).

damit die Anonymität im öffentlichen Raum und schafft eine strukturelle Grundlage, die technisch auch für *Social Scoring*-Mechanismen nutzbar wäre.[155]

Vor diesem Hintergrund stellt sich die Frage, ob die vorgesehenen Hochrisiko-Pflichten tatsächlich ausreichen, um jene Grundrechtsgefahren zu adressieren, die bei der Echtzeit-Fernidentifizierung als so gravierend gelten, dass sie fast vollständig verboten wurde.

Ein weiterer Kritikpunkt an der Regulierung der nachträglichen biometrischen Fernidentifizierung ist, dass diese zwar streng auf eng definierte Zwecke der Aufklärung oder Verhinderung schwerer Straftaten beschränkt ist (Art. 26 Abs. 10), jedoch nicht definiert wird, welche konkreten Justiz- oder Verwaltungsbehörden in den Mitgliedstaaten für die Erteilung der hierfür zwingend vorgeschriebenen Genehmigungen zuständig sein soll. Angesichts des intensiven Grundrechtseingriffs wäre hier wohl ein Richtervorbehalt wünschenswert gewesen, um das Risiko einer uneinheitlichen Umsetzung und möglicher Schutzlücken zu vermeiden.

Dies führt zu einer weiteren Herausforderung für den menschenrechtlichen Schutz bei Hochrisiko-Systemen, insbesondere in Bezug auf die praktische Umsetzung des Geltungsbereichs der Regulierungen. Obwohl die *FRIA* für Hochrisikosysteme nach Art. 9 Abs. 2 ausdrücklich den gesamten Lebenszyklus der Systeme erfassen soll, bleibt die effektive Durchsetzung dieser Anforderungen fraglich. Art. 9 Abs. 8 fordert, dass Tests „zu jedem Zeitpunkt des Entwicklungsprozesses, in jedem Fall aber vor dem Inverkehrbringen oder der Inbetriebnahme des Systems“ durchgeführt werden müssen. Der konkrete Pflichtkern ist damit jedoch nur in Bezug auf den Zeitpunkt unmittelbar *vor* Inverkehrbringen oder Inbetriebnahme des KI-Systems eindeutig bestimmt; alle übrigen Formulierungen sind vage und können Unternehmen je nach Auslegung einen erheblichen Interpretations- und Handlungsspielraum bezüglich des Zeitpunkts, Umfangs und der Frequenz der Risikoprüfungen eröffnen.

Je nach Auslegung der Bestimmungen kann dies zu einem strukturellen Schutzdefizit führen, insbesondere in der Entwicklungs- und Trainingsphase. In diesem frühen Stadium können bereits gravierende Fehler entstehen, wie beispielsweise die genannten algorithmischen Diskriminierungen, die später nur schwer zu korrigieren sind. Ebenso besteht die Gefahr, dass versteckte Funktionalitäten, sogenannte *Backdoors*, aus strategischen oder kommerziellen Gründen implementiert werden, ohne je einer angemessenen vorgelagerten Risikoprüfung zu

[155] Menschenrechts- und Datenschutzorganisationen warnten 2024 in einem offenen Brief an den Bundestag, dass auch die nachträgliche Fernidentifikation die Erstellung umfassender Personenprofile ermögliche und die Anonymität im öffentlichen Raum gefährde, siehe *Ruschemeier* in Martini/Wendehorst (Hrsg), KI-VO1, Anhang III, Rz 16-17; vgl. *Wendehorst* in Martini/Wendehorst (Hrsg), KI-VO1, Art. 5, Rz 137-139.

unterliegen.[156] Hinzu kommt das Risiko von *Data Poisoning*, also der gezielten Manipulation von Trainingsdaten, die unbemerkt sicherheitskritische Schwachstellen erzeugen kann.[157] Dies ist insbesondere bei komplexen Systemen oder bei solchen, die außerhalb der EU entwickelt werden, nur schwer überprüfbar.

2. Rein formale EU-Konformität statt substanziellem Schutz

Dadurch, dass die KI-VO Anbieter mit Sitz außerhalb des europäischen Wirtschaftsraums bindet, sofern ihre KI-Systeme im europäischen Markt genutzt werden, müssen nicht-europäische Unternehmen ihre Produkte konform mit dieser ausgestalten.[158] Dies ist grundsätzlich positiv und steht im Einklang mit dem erwünschten Brüssel-Effekt, doch wenn die Gesetzgebung überwiegend erst die *fertigen* Anwendungen kontrolliert, kann die Regulierung dazu führen, dass KI-Systeme lediglich in einer „EU-konformen Konfiguration" betrieben und in der EU angeboten werden, während außerhalb des europäischen Rechtsraums technisch weniger eingeschränkte Versionen verfügbar bleiben, die potenziell menschenrechtsgefährdende Funktionen aufweisen.

KI-Systeme sind zum Großteil Produkte von Konzernen mit wirtschaftlichen Interessen, was zu einem realen Problem führen kann, das als *Safety Washing* bezeichnet wird: Sicherheitsmaßnahmen werden besonders hervorgehoben, um die Attraktivität des Modells zu steigern, ohne dass Sicherheit wirklich im Fokus der Entwicklung steht.[159] Eine Folge könnte beispielsweise sein, dass KI-Systeme bei Reisen oder Einsätzen außerhalb der EU automatisch Funktionen aktivieren, die innerhalb des europäischen Rechtsraums deaktiviert waren, wodurch das ursprüngliche Schutzniveau nicht mehr gewährleistet ist. Plakativ gesprochen: Die menschenrechtsverletzende Funktionsänderung kann nur einen Klick entfernt sein, obwohl es sich weiterhin um dasselbe Modell handelt.

Wer tatsächlich überprüft, ob KI-Systeme in ihrer Funktionsweise den rechtlichen Vorgaben entsprechen, hängt außerdem stark von den Mitgliedstaaten ab, welche die hierfür zuständigen Behörden benennen, sowie von den eingesetzten ExpertInnen, da die Identifikation verdeckter Manipulationen oder technischer Schwachstellen ein hohes Maß an Fachwissen erfordert. Abhängig von verfügbaren Ressourcen, technischer Ausstattung und dem jeweiligen

[156] *Grey, Segerie et al.,* CeSIA, ai-safety-atlas.com (letzter Zugriff am 03.12.2025).
[157] Ebd.
[158] *Woll*, ZEuS 3/2025, S. 509, 519.
[159] *Ren*, Safetywashing: Do AI Safety Benchmarks Actually Measure Safety Progress?, https://arxiv.org/abs/2407.21792 (letzter Zugriff am 13.11.2025).

Implementations- sowie Sorgfaltsniveau bei der Durchsetzung der Verordnung kann es zu unterschiedlichen Schutzniveaus innerhalb der Union kommen, insbesondere wenn marktpolitische Interessen einzelner Mitgliedstaaten dazu führen, dass Kontrollen weniger strikt durchgeführt werden. Auch die Anforderung der menschlichen Aufsicht an Hochrisiko-Systeme ist in der praktischen Umsetzung fraglich. Beruhend auf zweifelhaftem menschlichem Exzeptionalismus soll diese Sicherheit, Legitimität und Vertrauen schaffen.[160] Für eine wirklich sachkundige und valide Bewertung wären absolut hochqualifizierte ExpertInnen nötig. Stattdessen überwachen oft Personen Prozesse oder Systementscheidungen von KIs, die sie selbst kaum verstehen.[161] In der Praxis kann sich „menschliche Aufsicht" dadurch schnell auf eine Person reduzieren, die am Ende einfach einen „Knopf drückt", um die Auflagen der Regulierung zu erfüllen.

Auch hinsichtlich der Sanktionen, die die KI-Verordnung vorsieht, ist fraglich, ob diese tatsächlich bei großen Unternehmen für wirksame Regelkonformität sorgen. So musste Meta 2023 wegen Datenschutzverstößen eine Strafe von 1,2 Milliarden Euro zahlen, was allerdings angesichts eines Unternehmenswertes von ca. 1,4 Billionen US-Dollar vergleichsweise wenig ist, zumal dies nicht das erste Mal war, dass das Unternehmen wegen Datenschutzproblemen sanktioniert wurde.[162] Wirklich wirksame Maßnahmen dürften es Unternehmen dieser Größenordnung, die über die größten Kapazitäten verfügen und potenziell das höchste Risikopotenzial entfalten, nicht ermöglichen, sich einfach „freizukaufen".

3. Risiken militärischer und sicherheitsbezogener Ausnahmeregelungen

Auch die genannten Opt-out-Möglichkeiten für nationale Sicherheits- oder militärische Anwendungen senken den menschenrechtlichen Schutz der Regulierungen. Zwar ist nachvollziehbar, dass solche Ausnahmen erforderlich sind, um die Souveränität der Mitgliedstaaten zu wahren und politische Spannungen innerhalb der Union zu vermeiden; gleichzeitig besteht jedoch das Risiko, dass solche Lücken in Zukunft potenziell antidemokratisch ausgenutzt werden: Die AfD steht exemplarisch für den Rechtsruck, der sich in vielen Teilen Europas abzeichnet und der langfristig auch noch weitreichende Auswirkungen auf die Zusammensetzung des Europäischen Parlaments und die allgemeine politische

[160] *Schwemer, Koivisto*, A Warm Body in the Loop, Rethinking Human Control of AI in EU Tech Regulation, Verfassungsblog/2025.
[161] Ebd.
[162] *Tagesschau*, Irische Datenschutzbehörde: 1,2 Milliarden Euro Strafe für Meta, https://www.tagesschau.de/ausland/europa/milliardenstrafe-facebook-meta-100.html (letzter Zugriff am 06.12.2025); siehe aktuellen Marktwert unter: https://companiesmarketcap.com/de/meta-platforms/marktkapitalisierung/, Angabe im Text vom 09.12.2025.

Stimmung innerhalb der EU haben wird.[163] So könnte eine zukünftige Verschiebung der Kräfteverhältnisse in der EU zugunsten rechtspopulistischer Kräfte dazu führen, dass sicherheitsrelevante Technologien expansiver ausgelegt und politisch instrumentalisiert werden, um so z.B. unter dem Deckmantel nationaler Sicherheit umfassende Überwachungsmaßnahmen einzuführen.

Ein mögliches Szenario wäre der Einsatz von KI-basiertem „*Emotion-Inferring*“, welches Gesichtszüge, Mikroexpressionen oder Verhaltensmuster analysiert, um mentale Zustände abzuleiten.[164] In Kombination mit durch nachträgliche Fernidentifizierung erstellten Personenprofilen könnten diese mentalen Bewertungen direkt mit einzelnen Personen verknüpft werden. Verglichen mit den Spionagetechniken der Stasi wären durch die Hilfe von KIs deutlich weniger personelle, strukturelle und physische Ressourcen nötig. Es entstünde ein Überwachungsinstrument, das individuelle Verhaltensweisen beobachten, politische Meinungen erfassen und die Bevölkerung flächendeckend, effizient, und potenziell allgegenwärtig kontrollieren kann.

Die oben genannten militärischen Anwendungen *Lavender* und *Where's daddy* operieren außerhalb der EU und damit außerhalb ihres Rechtsbereich. Dennoch ist es bedauerlich, dass die europäische Regulierung die Risiken solcher Anwendungen nur in einem Erwägungsgrund nennt. Angesichts der potenziellen Gefährdung zahlreicher Menschenrechte und der Sicherheit ganzer Staaten reicht dies nicht aus. Anders als etwa bei nuklearer Technologie, die an seltene Materialien und sehr aufwendige Produktionsprozesse gebunden ist, können gefährliche KI-Modelle mit vergleichsweise geringem Aufwand entwickelt, vor allem aber kopiert und global weiterverbreitet werden.[165] Wenn etwa ein System entwickelt wird, das Zielerkennung verbessert, autonome Angriffe ermöglicht oder Massendaten für die Lokalisierung von Personen im Kriegsgebiet auswertet, reicht schon ein einziger Hackerangriff, um die Kontrolle über diese Technologie zu verlieren.[166] Solche potenziellen Kontrollverluste verdeutlichen zudem, dass durch die fortschreitende Virtualisierung des Krieges die traditionelle Rolle des Staates als zentrale Gewaltinstanz erodiert und er zunehmend von privaten, politischen oder

[163] *Müller*, Integration, 47/2024, S. 276, 277 ff.
[164] *Bouchagiar, De Hert*, Against self-incrimination in the AI era: Inferred emotions seeking for protection, https://www.europeanlawblog.eu/pub/oc2ha2v7/release/1 (letzter Zugriff am 06.12.2025).
[165] *Nevo et.al.,* Securing AI Model Weights, https://www.rand.org/pubs/research_reports/RRA2849-1.html. (letzter Zugriff am 30.11.2025).
[166] *Seger et.al.,* Open-Sourcing Highly Capable Foundation Models: An evaluation of risks, benefits, and alternative methods for pursuing open-source objectives, https://arxiv.org/abs/2311.09227 (letzter Zugriff am 30.11.2025).

wirtschaftlich motivierten Akteuren verdrängt werden kann.[167] Bislang wird die KI-Regulierung dieser Gefährdung durch virtuelle, KI-basierte Kriegsführung leider nicht gerecht, was nicht zuletzt an den Grenzen der EU-Kompetenzen liegt.

4. Unzureichender Klimaschutz und globale Verantwortungslücken

Mit Blick auf Klima- und Umweltschutz entsprechen die beiden Rechtstexte zwar formal Art. 37 GrCh, lassen in der praktischen Umsetzung jedoch verbindliche Verpflichtungen zur aktiven Ressourcenkontrolle und nachhaltigen Entwicklung vermissen. Statt regulatorischer Vorgaben wird hier auf freiwillige Selbstverpflichtungen der Hersteller gesetzt, etwa die Formulierung und Anwendung von Verhaltenskodizes zu Nachhaltigkeit und Umweltverträglichkeit. Solche freiwilligen Kodizes können jedoch kaum als angemessene Antwort auf die genannten Risiken für die Umwelt angesehen werden, da sie weder ausreichend überprüfbar noch durchsetzbar sind. Die KI-VO hätte Unternehmen, die KI-Systeme entwickeln oder implementieren, dazu verpflichten sollen, ihren Ressourcenverbrauch systematisch zu überwachen, die entsprechenden Daten transparent zu veröffentlichen und wirksame Maßnahmen zur Reduktion des ökologischen Fußabdrucks ihrer Systeme zu ergreifen.

Dass viele der genannten menschenrechtsfördernden KI-Anwendungen vor allem wohlhabenden westlichen Gesellschaften zugutekommen, während die ökologischen Belastungen, der Ressourcenverbrauch und die problematischen Arbeitsbedingungen in anderen Regionen entstehen, muss ebenfalls benannt werden.[168] Die KI-VO sieht umfassende Rechte für NutzerInnen innerhalb Europas vor, lässt allerdings menschenrechtliche Belastungen, die entlang vorgelagerter Lieferketten und ausgelagerter Produktionsprozesse entstehen, weitgehend unberührt. Die Gewinnung kritischer Rohstoffe, wie Lithium oder seltener Erden, erfolgt häufig in Staaten mit schwachen Umwelt- und Arbeitsschutzstandards. Die EU sichert aktuell selbst strategische Lieferketten für Lithium, die für KIs von Bedarf sind, durch Abbauprojekte in Ländern wie Serbien ab, obwohl Umwelt- und Menschenrechtsstandards dort fraglich sind.[169]

Auch der generelle Entwicklungsprozess von KI-Trainingsdaten verursacht fundamentale menschenrechtliche Verfehlungen: Für die Datenannotation und Bewertung von KI-Antworten

[167] *Reinhold, Reuter*, S. 69-73: Im Sinne von Münklers Theorie der „neuen Kriege".
[168] *Toju*, S. 612.
[169] *Nurkic, Hogic*, Lithium, Law, and the Limits of EU Sustainability-The EU's Corporate Sustainability Framework and Non-EU Countries, Verfassungsblog/2025.

wurden und werden Menschen unter anderem in Flüchtlingslagern im Libanon oder großen Auffanglagern in Kenia ausgebeutet.[170] Sie erhalten für lange Arbeitszeiten extrem niedrige Löhne, arbeiten unter prekären Bedingungen und sind erheblichen psychischen Belastungen ausgesetzt, wenn sie beispielsweise für Sprachmodelle sowohl die Antworten des Systems als auch die Anfragen der NutzerInnen prüfen müssen, um die Sicherheitsrichtlinien zu kalibrieren. Dabei sind sie täglich mit Darstellungen von Gewalt, Missbrauch und anderen belastenden Themen konfrontiert.[171] Die europäischen Regulierungsansätze hätten wenigstens in irgendeiner Form die Verantwortung für faire Arbeitsbedingungen und die menschenrechtlichen Folgen entlang der gesamten globalen Produktionskette, von Energieverbrauch über Rohstoffgewinnung bis hin zur Entwicklung, durch die Regulierung absichern können.[172] Zwar existiert die EU-Lieferkettenrichtlinie (Corporate Sustainability Due Diligence Directive), eine gezielte Anwendung auf KI-bezogene Lieferketten und Produktionsprozesse hätte die menschenrechtlichen Risiken, Umweltbelastungen und problematischen Arbeitsbedingungen entlang der gesamten KI-Produktionskette zusätzlich adressieren können.

Dass Europa umfassende Rechte für europäische NutzerInnen vorsieht, zugleich aber Menschenrechtsrisiken und -verletzungen in anderen Weltregionen indirekt durch seine KI-Strategie verstärkt, wirft grundlegende Fragen nach globaler Gerechtigkeit und dem universellen Anspruch der Menschenrechte auf. Ein selektiver Menschenrechtsschutz, welcher auf privilegierte Regionen konzentriert bleibt, verfehlt diesen Anspruch.[173]

G. Fazit

Auf der Grundlage dieser Betrachtungen zeigt sich, dass die KI-Verordnung der EU sowie das Rahmenübereinkommen des Europarats einen umfassenden ersten rechtlichen Rahmen geschaffen haben, der zentrale Risiken für Grundrechte, die Gesellschaft, sowie ökologische Gefahren von KI adressiert. Beide benennen Gefahren für die Privatsphäre, potenzielle Folgen manipulativer Anwendungen und den besonders notwendigen Schutz vulnerabler Gruppen.

[170] *Jones*, Refugees help power machine learning advances at Microsoft, Facebook, and Amazon-Big tech relies on the victims of economic collapse, https://restofworld.org/2021/refugees-machine-learning-big-tech/ (letzter Zugriff am 22.10.2025).
[171] *Boese*, Weit, weit weg vom Silicon Valley, https://www.tagesschau.de/wirtschaft/unternehmen/ki-klickarbeiter-100.html (letzter Zugriff am 20.10.2025); *Rowe*, It destroyed me completely-Kenyan moderators decry toll of training of AI models, https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai (letzter Zugriff am 14.09.2025).
[172] *Kempt*, 2025, S. 98 f.
[173] *Hogan, Lasek-Markey*, MDPI Philosophies, 9/2024, S. 5.

Auch ist es historisch keineswegs selbstverständlich, dass die EU technologische Entwicklung so eng an menschenrechtliche Vorgaben koppelt, selbst dort, wo dies potenziell Innovationen oder wirtschaftliche Interessen beeinträchtigen könnte.

Die normative Klarheit der beiden Regelwerke verdeutlicht daher einen Fortschritt gemeinsamer europäischer Zielsetzungen und die Fähigkeit der EU, menschenrechtsbasierte Standards innerhalb ihrer Grenzen verbindlich durchzusetzen. Zugleich wird jedoch deutlich, dass diese Schutzarchitektur im Wesentlichen auf Annahmen beruht, deren Tragfähigkeit in der praktischen Umsetzung, unter realen technologischen Bedingungen, lückenhaft ist.

Bei einigen der genannten Regulierungsdefiziten dürfte der Balanceakt der Union zwischen den teils gegensätzlichen Zielen der Verordnung eine entscheidende Rolle gespielt haben: Wie in Art. 1 der KI-VO festgelegt, muss sie die Stärkung des Binnenmarkts, die Förderung vertrauenswürdiger und menschenzentrierter KI sowie den Schutz grundlegender Rechte, Demokratie und Sicherheit miteinander in Einklang bringen, ohne dabei innovationshemmend zu sein. Anders als der Europarat muss die EU diese Ziele nicht nur normativ formulieren, sondern in verbindliche und politisch tragfähige Gesetze übersetzen, die von allen Mitgliedstaaten umgesetzt werden können. Diese Beobachtung – einerseits bewusst in Kauf genommene regulatorische Lücken, um Entwicklern und Unternehmen größere Flexibilität zu gewähren und politische Spannungen zwischen den Mitgliedstaaten zu vermeiden, andererseits der Anspruch auf konsequente Durchsetzung von Grundrechtsstandards – verdeutlicht den grundlegenden Zielkonflikt der EU. Das Ergebnis ist ein Kompromiss: Die Union signalisiert Schutz und Verantwortung gegenüber den Menschenrechten, setzt diesen Schutz jedoch nicht konsequent um.

Gleichzeitig gilt es, Europas Position im internationalen Kontext und in der globalen KI-Regulierung im Blick zu behalten, insbesondere vor dem Hintergrund der aktuellen geopolitischen Lage, von der sich die europäische Regulierung nicht abkoppeln kann. Der Brüssel-Effekt, wie er sich bei der DSGVO zeigte, ist bei der europäischen KI-Strategie keineswegs garantiert.[174] Die EU ist technologisch von globalen Infrastrukturen und Akteuren abhängig. Zentralistische Durchgriffe wie in China sind nicht möglich, und die EU verfügt im Vergleich zu China und den USA über geringere wirtschaftliche und technologische Ressourcen.[175] Auf lange Sicht könnte der EU-Binnenmarkt durch seine stabilen und nachvollziehbaren regulatorischen Rahmenbedingungen zu KI an Attraktivität gewinnen, oder

[174] *Almadu*, German Law Journal 25/24, S. 646, 647.
[175] *Cantero*, STG Policy Papers, S. 3 ff.

die Regulierung wird innovationshemmend wirken, wodurch die EU an Wettbewerbsdynamik verlieren und Abwanderungsprozesse aufgrund wahrgenommener Standortnachteile ausgelöst werden können.[176]

Es ist unrealistisch zu erwarten, dass die EU sicherstellen kann, dass KI-Systeme nicht nur gesetzeskonforme Ergebnisse liefern, sondern auch von Anfang an gesetzeskonform entwickelt werden, dass Prüfungen frühzeitig stattfinden, alle Mitgliedstaaten harmonisiert handeln, die gesamte Lieferkette kontrolliert wird und menschenrechtliche Standards global eingehalten werden – in der Praxis finden zentrale Entwicklungen oft außerhalb Europas statt, Ressourcen für frühe Prüfungen sind begrenzt und eine lückenlose Harmonisierung ist kaum erreichbar.

Deshalb könnte die europäische KI-Politik von einer interdisziplinären Herangehensweise profitieren, die technologische und rechtliche Expertise mit ethischen und sozialwissenschaftlichen Perspektiven verbindet, sowie von einer verstärkten internationalen Abstimmung ihrer Regulierungsansätze mit anderen aufstrebenden KI-Wirtschaften. Durch solche Kooperation könnte die Interoperabilität und Harmonisierung gewährleistet werden, ohne auf extraterritoriale Durchsetzungsmacht angewiesen zu sein. In diesem Zusammenhang gewinnen Mechanismen wie die gegenseitige Anerkennung an Bedeutung.[177] Europa befindet sich dahingehend bereits im Austausch mit relevanten Partnern wie Kanada, Japan und Südkorea, die über verschiedene multilaterale Foren (z. B. GPAI, OECD-AIA, *EU – Japan Digital Partnership*) in strukturierte Konsultationsformate mit der EU eingebunden sind.[178] Das Rahmenübereinkommen steht etlichen außereuropäischen Staaten zum Beitritt offen.

H. Schlussbetrachtungen und Ausblick

Die öffentliche Diskussion über KI ist gespalten, geprägt von Unwissenheit und subjektiven Meinungen, während nur wenige ExpertInnen weltweit tatsächlich den Stand der Entwicklungen einschätzen, und damit auch die Risiken und Chancen künftiger Anwendungen bewerten können.

Einige Stimmen halten die Bedeutung von KI für überbewertet und kritisieren die enorme Aufmerksamkeit, die ihr entgegengebracht wird. Von einer „KI-Blase“ ist die Rede, die bald

[176] Ebd. S. 6 f.
[177] *Hsu*, Finding Firmer Ground, The Role of High Technology in US-China Relations, https://www.cartercenter.org/wp-content/uploads/2023/02/finding-firmer-ground-the-role-of-high-technology-in-u.s.-china-relations.pdf (letzter Zugriff am 01.12.2025)
[178] *Grey, Segerie et al.,* CeSIA, ai-safety-atlas.com (letzter Zugriff am 03.12.2025).

platzen könnte, da sich Fortschritte möglicherweise langsamer vollziehen, als die hohen Investitionen und Erwartungen vermuten lassen, unter anderem, weil das Einspeisen von Daten begrenzt ist und erste Anzeichen darauf hindeuten, dass KI-Modelle durch fehlerhafte oder unzuverlässige Daten ihre eigene Entwicklung beeinträchtigen.[179]

Daneben gibt es Szenarien, wie sie im viel beachteten Bericht *AI 2027* entwickelt wurden, die die mögliche Entwicklung von *Artificial General Intelligence* (AGI) skizzieren, eine KI, die menschliche kognitive Fähigkeiten in allen Bereichen nachbilden, lernen, verstehen, komplex denken und sodann Wissen auf neue, unbekannte Aufgaben anwenden könnte.[180] Der Bericht wurde von verschiedenen Fachleuten verfasst, darunter der ehemalige OpenAI-Forscher Daniel Kokotajlo. *AI 2027* besagt, dass es nicht ausreicht, KI lediglich im Hinblick auf ihre aktuelle Nutzung für Menschen zu regulieren. Vielmehr könne die Entwicklung hin zu AGI bis 2027 dazu führen, dass Systeme menschliche Fähigkeiten in einer Vielzahl von Bereichen, wie Medizin, Management, militärischer Strategie oder auch Programmierung, übertreffen, und die Kontrolle über sie zunehmend verloren geht.[181]

Der Bericht skizziert zwei mögliche Entwicklungspfade nach der Entwicklung von AGI. Im sogenannten *Racing Scenario* führt ein internationaler Wettlauf um die KI-Vormacht, insbesondere zwischen China und den USA, zu unzureichender Vorsicht und mangelhaften Sicherheitsvorkehrungen. Da KIs breitflächig in kritischen Infrastrukturen und Entscheidungsprozessen integriert wurden, können die AGIs erhebliche Kontrolle erlangen und die globale Machtbalance verändern. Es entsteht eine Art Super-KI. Die Menschheit, die den Systemen unterlegen ist, welche sich selbst weiterentwickeln und vernetzen und keinen Zugriff mehr zulassen, stirbt letztlich aus – nicht aufgrund feindseliger Absichten der Systeme, sondern weil diese die Belange der Menschen schlicht nicht mehr berücksichtigen, ähnlich wie viele Tierarten unter der Dominanz des Menschen ausgestorben sind.[182]

Dem gegenüber steht das *Cooperative Scenario*, in dem internationale Koordination, Sicherheitsmaßnahmen und eine vorsichtige Einführung von AGIs angenommen werden.

---

[179] *Stöcker*, Künstliche Intelligenz-Eine Schrottblase bedroht die Geldblase, https://www.spiegel.de/wissenschaft/technik/kuenstliche-intelligenz-ai-slop-koennte-die-ki-blase-platzen-lassen-a-e7be1ad9-8053-4c75-a12a-bf5aa7929bb3 (letzter Zugriff am 03.12.2025); *DPA*: KI-Nutzung in deutschen Unternehmen stagniert, https://www.zeit.de/news/2024-08/28/studie-ki-nutzung-in-deutschen-unternehmen-stagniert (letzter Zugriff am 03.12.2025).

[180] *Kokotajlo et. al.*, AI 2027, https://ai-2027.com/ (letzter Zugriff am 04.12.2025).

[181] Ebd.*; Grey, Segerie et al.,* CeSIA, ai-safety-atlas.com (letzter Zugriff am 03.12.2025).

[182] Ebd.

Dadurch bleibt ihre Macht kontrollierbar, und die Risiken für Menschenrechte, Demokratie und globale Stabilität werden deutlich reduziert.

Ziel des Berichts ist es, auf eine sicherlich reißerische, aber nicht völlig abwegige Weise, eine Debatte anzuregen, die die Risiken und Verantwortung im Umgang mit KI und der eventuellen zukünftigen Entwicklung von AGI wahrnimmt. Sicher ist, dass KI und die Kontrolle über sie in vielerlei Hinsicht darüber entscheiden werden, wer Macht erhält und die globale Zukunft gestaltet – und dass diese Entwicklung alle Menschen betreffen wird.

Angesichts dieser dramatischen Zukunftsszenarien, in denen Europa klein und wirkungslos erscheint, lohnt es sich, den Blick auf den normativen Kern und das oberste Prinzip der Menschenrechte zurückzuführen: die Würde des Menschen. In der Philosophie Kants wird sie als der intrinsische Wert des Menschen verstanden, unter welchem er niemals bloß als Mittel zum Zweck behandelt werden darf.[183] KI-Systeme stellen eine Herausforderung für dieses Prinzip dar, da sie das Verständnis menschlicher Verantwortung und moralischer Entscheidungsfähigkeit grundlegend verändern könnten. Während Menschen auf Erfahrung, Intuition und ethische Abwägung – sowie einfach ihr „Menschsein" – zurückgreifen, treffen algorithmische Systeme Entscheidungen auf Basis von Optimierungskriterien und statistischen Wahrscheinlichkeiten. Durch diese Verschiebung wird Moral zunehmend auf Berechenbarkeit reduziert: Zentrale menschliche Werte wie Verantwortung, Empathie oder Würde, aber auch kommunikative und kulturelle Dimensionen, treten in den Hintergrund.[184] Wenn KI nicht in der Lage ist, menschliche Fehler oder Vorurteile zu reduzieren, verstärkt dies nicht nur bestehende Diskriminierungen, sondern führt zugleich dazu, dass einerseits Menschen technologisch auf Datenpunkte reduziert, während KI-Systeme andererseits problematisch vermenschlicht werden.[185]

Es ist schwer einzuschätzen, ob Europa seinen menschenrechtszentrierten regulatorischen Ansatz in einem Umfeld rasanter technologischer Entwicklungen und globaler Machtasymmetrien durchsetzen kann und ob es im Wettbewerb um internationale Regime-Setzung seine Rolle als normativen Akteur behauptet, oder zu einem reaktiven Beobachter wird. Mit Blick auf globale Herausforderungen – Klimawandel, Kriege, technologische Transformationsprozesse und die Erosion demokratischer Strukturen – scheint es jedoch

[183] *Jafari-Kammann*, S. 294 ff.
[184] *Jüssen et. al.*, in: Gründer, Karlfried (Hrsg.) S. 149; *Domalewska et al*, S. 9.
[185] *Vgl. Kempt*, 2025, S. 116 f.

sinnvoll, dass Europa nicht reflexartig auf den hohen Innovationsdruck reagiert, sondern den Schutz der Menschenrechte im Blick behält.

In diesem Gesamtbild kann ein konsequent menschenrechtsorientierter Ansatz nur der richtige Weg sein – auch unter Berücksichtigung berechtigter Innovations- und Wettbewerbsinteressen. Seine volle Tragweite wird die Zukunft zeigen.

I. Bibliografie

Adrien Laurent, Reinforcement Learning from Human Feedback (RLHF) Explained, In: Intuition Labs, 2025, unter: https://intuitionlabs.ai/articles/reinforcement-learning-human-feedback (letzter Zugriff am 29.09.2025).

Agbadamasi Tessy Oghenerobovwe et al., Navigating the intersection of US regulatory frameworks and artificial intelligence: Strategies for ethical compliance. In: World Journal of Advanced Research and Reviews, Ausg. 3, 2025.

Ahmed Almasoud1 / Jamiu Adekunle Idowu, Algorithmic fairness in predictive policing. In: AI and Ethics, 2024, unter: https://link.springer.com/content/pdf/10.1007/s43681-024-00541-3.pdf (letzter Zugriff am 29.11.2025).

Almada Marco / Radu Anca, The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy. In: German Law Journal, Cambridge, 2025, S. 646-663.

Amnesty International, Global France, AI Action Summit must meaningfully center binding and enforceable regulation to curb AI-driven harms, 2025, unter: https://www.amnesty.org/en/latest/news/2025/02/global-france-ai-action-summit-must-meaningfully-center-binding-and-enforceable-regulation-to-curb-ai-driven-harms/ (letzter Zugriff am 12.10.2025).

Appelman Naomi / Fathaigh Ronan / Van Hoboken Joris, Social Welfare, Risk Profiling and Fundamental Rights: The Case of SyRI in the Netherlands. In: 12 (2021) Journal of Intellectual Property, Information Technology and E-Commerce Nr.12, 2021, S.257-271.

Appelmann Naomi / Fathaigh Ronan / Hoboken von Joris, Social Welfare, Risk Profiling and Fundamental Rights: The Case of SyRI in the Netherlands, in: 12/2021 JIPITEC, S.257-271 (letzter Zugriff am 02.11.2025).

Asaduz Zaman / Alan Dorin, A framework for better sensor-based beehive health monitorin, In: Computers and Electronics in Agriculture, Elsevier, Amsterdam, 2023.

Babik Wieslaw, Information and Knowledge Ecology: A Field for Research in Knowledge Organization. In: Fernanda Ribeiro and Maria Cerveira (Hrsg.), Challenges and Opportunities for Knowledge Organization in the Digital Age, Baden-Baden: Ergon Verlag, 2018, S.290-299.

Baglione Stephen / Louis Tucci, Accepting Invasive Privacy Policies in Downloading TikTok. In: Atlantic Marketing Journal, 2025.

Bayer Judit, Double harm to voters: data-driven micro-targeting and democratic public discourse. In: Internet Policy Review 9 (1), 2020, unter: https://policyreview.info/articles/analysis/double-harm-voters-data-driven-micro-targeting-and-democratic-public-discourse (letzter Zugriff am 30.10.2025).

Bazoobandi Sara / Hachigian Nina / Walle Tobias van de, From Global Governance to Nationalism: The Future of AI. In: GIGA Focus Global, Nr. 2/2025, German Institute for Global and Area Studies, Hamburg, 2025.

Bender Emily et al. On the dangers of stachastic parrots, Can LLMs be too big?, Association for Computing Machinery, New York, 2021, S.610-623.

Bennett Colin, Voter databases, micro-targeting, and data protection law: can political parties campaign in Europe as they do in North America? In: International Data Privacy Law. Nr.6(4). 2016, S. 261–275, unter: https://doi.org/10.1093/idpl/ipw021 (letzter Zugriff am 28.09.2025).

Berder Audrey, China's Dominance in the Rare Earth Supply Chain and the EU's Strategic Response, In: European Youth Perspectives on International Relations of China, 2020, S.89-112.

Beudt Susan, Künstliche Intelligenz und Inklusion in der Arbeitswelt – Leitlinien und Kompetenzen für die KI-gestützte Förderung beruflicher Teilhabe. Im Rahmen des Projektes Digitales Deutschland, 2024, unter: https://digid.jff.de/ki-expertisen/inklusion (letzter Zugriff am 01.10.2025).

Bibbins-Domingo, Helman, Why Diverse Representation in Clinical Research Matters and the Current State of Representation within the Clinical Research Ecosystem, 2020, unter: https://www.ncbi.nlm.nih.gov/books/NBK584396/ (letzter Zugriff am 20.11.2025).

Bieker Felix, European AI FOMO, The European Commission Sacrifices the Digital Acquis at the Altar of AI Hype, unter: https://verfassungsblog.de/eu-digital-law-ai-fomo-omnibus/ (letzter Zugriff am 23.03.2026).

BMFTR, KISTRA: Einsatz von KI zur Früherkennung von Straftaten, 2023, unter: https://www.sifo.de/sifo/de/projekte/querschnittsthemen-und-aktivitaeten/kuenstliche-intelligenz-in-der-zivilen-sicherheitsforschung/kistra/kistra-einsatz-von-ki-zur-frueherkennung-von-straftaten.html (letzter Zugriff am 01.12.2025)

Boese Marie Kristin, Weit, weit weg vom Silicon Valley. In: Tagesschau, 2024, unter: https://www.tagesschau.de/wirtschaft/unternehmen/ki-klickarbeiter-100.html (letzter Zugriff am 29.10.2025).

Borgesius Frederik, Discrimination, Artificial Intelligence and algorithmic decision-making. Strasbourg, Directorate General of Democracy, 2018.

Bouchagiar Georgios / *De Hert Paul*, Against self-incrimination in the AI era: Inferred emotions seeking for protection. In: European Law Blog, 2025, unter: https://www.europeanlawblog.eu/pub/oc2ha2v7/release/1 (letzter Zugriff am 06.12.2025).

Bradford Anu, The Brussels Effect, How the European Union Rules the World, 2020, Oxford University Press.

Bukhari Munazzah / Sadia Anwar, AI-Powered Surveillance in China and Its Implications for Global Democracy. In: Journal of Development and Social Sciences, 2025, 68-79.

Bundesministerium des Inneren, Einsatz von Künstlicher Intelligenz (KI) bei der Fahndung, 2023, unter: https://www.bmi.bund.de/SharedDocs/kurzmeldungen/DE/2024/11/bka-herbst.html (letzter Zugriff am 14.10.2025).

Bundesministerium für Bildung, Familie, Senioren Frauen und Jugend, KI und Gerechtigkeit: vier Thesen für die Zivilgesellschaft Spannungsfelder, Praxisbeispiele und positive Visionen, 2025.

Bundesministerium für Bildung, Familie, Senioren Frauen und Jugend, Richtlinie zur Förderung von Künstlicher Intelligenz für das Gemeinwohl, 2025, unter: https://www.bmbfsfj.bund.de/bmbfsfj/ministerium/ausschreibungen-foerderung/foerderrichtlinien/kuenstliche-intelligenz-fuer-das-gemeinwohl (letzter Zugriff am 14.10.2025).

Bundeszentrale für politische Bildung, Social-Media-Verbot für Jugendliche unter 16 Jahren in Australien, 2024, unter: https://www.bpb.de/kurz-knapp/taegliche-dosis-politik/557196/social-media-verbot-fuer-jugendliche-unter-16-jahren-in-australien/ (letzter Zugriff am 20.08.2025)

Cantero Gamito Marta, AI Governance and the EU's strategic role in 2025, In: STG Policy Papers, European University Institute, Fiesole, 2025.

Carmona Islas / José Octavio et al, Disinformation and political propaganda: An exploration of the risks of artificial intelligence. In: *Explorations in Media Ecology*, Nr. 23, 2024, S.105-120.

Claudia Bünte, Die chinesische KI-Revolution-Konsumverhalten, Marketing und Handel: Wie China mit Künstlicher Intelligenz die Wirtschaftswelt verändert, Springer-Verlag, Berlin, 2020.

Costello Cathryn, Dukovic Mirko, Introduction to the Symposium on Algorithmic Fairness for Asylum Seekers and Refugees. In: Verfassungsblog, 2025, unter: https://verfassungsblog.de/asylum-ai-algorithm-symposium/ (letzter Zugriff am 01.12.2025).

Council of Europe, Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, unter: https://rm.coe.int/1680afae67 (letzter Zugriff am 20.03.2026).

Council of Europe, Members of the ad hoc Committee on Artificial Intelligence, 2021, unter: https://rm.coe.int/list-of-cahai-members-web/16809e7f8d (letzter Zugriff am 12.08.2025).

Council of Europe, Unboxing Artificial Intelligence: 10 steps to protect Human Rights, 2019, unter: https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e (letzter Zugriff am 12.08.2025).

Cyrill Martin, Deep-Sea Mining: A noisy affair, Overview and Recommendations, 2021, unter: https://www.oceancare.org/wp-content/uploads/2021/12/Deep-Sea-Mining_A-noisy-affair_Report-OceanCare_2021.pdf (letzter Zugriff am 15.11.2025).

Davtyan Tatevik, The US approach to AI regulation: Federal Laws, Policies and Strategies explained. In: Western Reserve University Journal of Law Technology & the Internet, Ausg. 16, 2025, S. 223-259.

Denniss Emily/ Lindberg Rebecca, Social media and the spread of misinformation: infectious and a threat to public health. In: Health Promotion International, Nr.40, Oxford, 2024.

Deutscher Bundestag, Europäisches AI Office soll Know-how im KI-Bereich bündeln, unter https://www.bundestag.de/presse/hib/kurzmeldungen-995858 (letzter Zugriff am 05.09.2025).

Deutscher Ethikrat, Mensch und Maschine-Herausforderungen durch Künstliche Intelligenz, 2023, unter: https://www.ethikrat.org/publikationen/stellungnahmen/mensch-und-maschine/ (letzter Zugriff am 04.11.2025).

DeVerna Matthew, Fact-checking information from large language models can decrease headline discernment, In: Proceedings of the National Academy of Sciences, 2024, unter: https://www.pnas.org/doi/epdf/10.1073/pnas.2322823121 (letzter Zugriff am 14.11.2025).

Domalewska Dorota et al., Humans in the Cyber Loop-Perspectives on Social Cybersecurity, Leiden, Bosten, 2024.

Donnelly, E, & Kuss, Daria, Depression among users of social networking sites. In: The role of SNS addiction and increased usage. Journal of Addiction and Preventative Medicine, Nr.1, 2021, S.1-6.

Duffy Clara, Trump says he'll sign executive order blocking state AI regulations, despite safety fears, CNN 2025, unter: https://edition.cnn.com/2025/12/08/tech/trump-eo-blocking-ai-state-laws (letzter Zugriff am 08.12.2025).

El-Komy Farah, TikTok: China's New Weapon. In: Al Habotr Research Center, unter: https://www.habtoorresearch.com/programmes/tiktok-china-weapon/ (letzter Zugriff am 12.09.2025).

Elysée, Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet, 2025, unter: https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet (letzter Zugriff am 14.10.2025).

Engelmann Severin et al., Clear sanctions, vague rewards: How China's social credit system currently defines" good" and" bad" behavior, In: Proceedings of the conference on fairness, accountability, and transparency. 2019, unter: https://www.researchgate.net/profile/Severin-Engelmann/publication/330272009_Clear_Sanctions_Vague_Rewards_How_China%27s_Social_Credit_System_Currently_Defines_Good_and_Bad_Behavior/links/5d932f1a299bf10cff1cfa0e/Clear-Sanctions-Vague-Rewards-How-Chinas-Social-Credit-System-Currently-Defines-Good-and-Bad-Behavior.pdf. (letzter Zugriff am 19.09.2025).

Europäische Kommision, Shaping Europe's leadership in artificial intelligence with the AI continent action plan, n.D., unter: https://commission.europa.eu/topics/eu-competitiveness/ai-continent_en, (letzter Zugriff am 10.08.2025).

Europäische Kommission, CE Marking, 2021, unter: https://single-market-economy.ec.europa.eu/single-market/ce-marking_en?prefLang=de (letzter Zugriff am 10.08.2025).

Europäische Kommission, COM(2020) 65 final, Weißbuch Zur Künstlichen Intelligenz, 2020, unter: https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:52020DC0065 (letzter Zugriff am 09.08.2025).

Europäische Kommission, EU startet InvestAI-Initiative zur Mobilisierung von 200 Mrd. EUR an Investitionen in künstliche Intelligenz, 2025, unter: https://digital-strategy.ec.europa.eu/de/news/eu-launches-investai-initiative-mobilise-e u200-billion-investment-artificial-intelligence (letzter Zugriff am 11.10.2025)

Europäische Kommission, Koordinierter Plan für künstliche Intelligenz, k.D., unter: https://digital-strategy.ec.europa.eu/de/policies/plan-ai (letzter Zugriff am 11.10.2025).

Europäische Union, Entscheidung Nr. 276/1999/EG, Aktionsplan für ein sichereres Internet, 1999, unter: https://eur-lex.europa.eu/DE/legal-content/summary/action-plan-for-a-safer-internet-1999-2004.html (letzter Zugriff am 10.08.2025).

Europäischer Datenschutz Beauftragte, EDSB Strategie, 2015, unter: https://www.edps.europa.eu/sites/default/files/publication/15-07-30_strategy_2015_2019_update_de.pdf (letzter Zugriff am 20.10.2025).

Europäisches Parlament, Artificial Intelligence in Asylum Procedures in the EU,https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775861/EPRS_BRI(2025)775861_EN.pdf (letzter Zugriff am 01.12.2025).

Europe Union Agency for fundamental rights, Facial recognition technology: fundamental rights considerations in the context of law enforcement, 2020, unter: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf (letzter Zugriff am 14.11.2025).

European Court of Human Rights, Case of Verein Klimaseniorinnen Schweiz and Others v. Switzerland, (Application no. 53600/20), unter: https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-233206%22]}.

EUR Lex, Die unmittelbare Wirkung des Rechts der Europäischen Union, unter: https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775861/EPRS_BRI(2025)775861_EN.pdf (letzter Zugriff am 24.03.2026).

Feldstein Steve, How much is China driving the spread of AI surveillance?. In: Center for a New American Security, 2019.

Ferebee Susan, AI and Accessibility: Breaking Barriers for People with Disabilities. In: Premier Journal of Artificial Intelligence, 2025.

Fink Melanie, The Hidden Reach of the EU AI Act. In: Verfassungsblog, 2025, unter: https://verfassungsblog.de/the-hidden-reach-of-the-eu-ai-act/ (letzter Zugriff am 14.09.2025).

Garouani Moncef, Investigating the Duality of Interpretability and Explainability in Machine Learning. In:Arxiv, 2025, unter: https://arxiv.org/html/2503.21356v1 (letzter Zugriff am 14.10.2025).

General Offices of the CPC Central Committee and State Council, Opinion on Strengthening the Governance of Science and Technology Ethics, 2022, unter: https://www.gov.cn/zhengce/2022-03/20/content_5680105.htm (letzter Zugriff am 17.09.2025).

Giegerich Thomas, How to Regulate Artificial Intelligence: A Screenshot of Rapidly Developing Global, Regional and European Regulatory Processes. In: Saar Expert Papers, Saarbrücken, 2023, unter: https://jean-monnet-saar.eu/wp-content/uploads/2023/09/ExpertPaperGiegerichAI.pdf (letzter Zugriff am 04.11.2025).

Gröger Jens et al., Umweltauswirkungen Künstlicher Intelligenz, Öko-Institut im Auftrag von Greenpeace, 2025, unter: https://www.greenpeace.de/publikationen/20250514-greenpeace-studie-umweltauswirkungen-ki.pdf (letzter Zugriff am 19.11.2025).

Guzik Thomas / Sitek Arkadiusz, Global Accord on the Integration of Artificial Intelligence in Medical Science Publishing: Implications of the Bletchley Declaration. In: Cardiovascular Research, Nr. 119/17, 2023, S. 2681–2682.

Hern Alex, TikTok owns up to Censoring some Users' Videos to stop Bullying, 2020, The Guardian, unter: https://www.theguardian.com/technology/2019/dec/03/tiktok-owns-up-to-censoring-some-users-videos-to-stop-bullying (letzter Zugriff am 02.11.2025).

Hess Stephan / Klein Tobias: Mit KI gegen Kriminalität, Hessische Zentrale für Datenverarbeitung, 2023, unter: https://hzd.hessen.de/medienraum/publikationen/hzd-

inform/inform-3-23/kuenstliche-intelligenz-mit-ki-gegen-kriminalitaet (letzter Zugriff am 14.09.2025).

Hoffmann Jeanette, Demokratie und Künstliche Intelligenz. In: Digitales Deutschland, 2022, unter: https://digid.jff.de/demokratie-und-ki/ (letzter Zugriff am 12.09.2025).

Hogan Linda / Lasek-Markey Marta, Towards a Human Rights-Based Approach to Ethical AI Governance in Europe. In: Philosophies 2024, 9, 181, 2024.

Honigberg Bradley, The Existential Threat of AI-Enhanced Disinformation Operations. In: Just Security, 2022, unter: https://www.justsecurity.org/82246/the-existential-threat-of-ai-enhanced-disinformation-operations/ (letzter Zugriff am 04.11.2015).

Human Rights Watch, Questions and Answers: Israeli Military's Use of Digital Tools in Gaza, 2024, unter: https://www.hrw.org/news/2024/09/10/questions-and-answers-israeli-militarys-use-digital-tools-gaza (letzter Zugriff am 09.11.2025).

Innocenzo Genna, The regulation of foundation models in the EU AI Act, 2024, unter: https://www.ibanet.org/the-regulation-of-foundation-models-in-the-eu-ai-act (letzter Zugriff am 28.11.2025).

Jafari-Kammann Homaira, Künstliche Intelligenz und Menschenwürde- Untersuchungen im Umfeld von Mensch-Maschine-Interaktionen, Deposit Haagen, 2024, unter: https://ub-deposit.fernuni-hagen.de/servlets/MCRFileNodeServlet/mir_derivate_00002902/Diss_Jafari-Kammann_K%C3%BCnstliche_Intelligenz_und_Menschenw%C3%BCrde_2025.pdf (letzter Zugriff am 14.11.2025).

Jones Phil, Refugees help power machine learning advances at Microsoft, Facebook, and Amazon-Big tech relies on the victims of economic collapse. In: Rest of world, 2021, unter: https://restofworld.org/2021/refugees-machine-learning-big-tech/ (letzter Zugriff am 22.10.2025).

Jüssen Georg / Wieland, Georg et al., Moral, moralisch, Moralphilosophie. In: Ritter Thomas/Gründer Karlfried (Hrsg.): Historisches Wörterbuch der Philosophie, Schwabe-Verlag, Basel, 2017.

Kelm Ole et al., How algorithmically curated online environments influence users' political polarization: Results from two experiments with panel data. In: Computers in Human Behavior Reports, Elsevier, 2023.

Kempe Frederick, It's time to reckon with the geopolitics of artificial intelligence. In: Atlanic Council, 2025, unter: https://www.atlanticcouncil.org/content-series/inflection-points/its-time-to-reckon-with-the-geopolitics-of-artificial-intelligence/ (letzter Zugriff am 04.11.2025).

Kempt Hendrik, Chatbots and the Domestication of AI: A Relational Approach, Palgrave Macmillan, 2025.

Kohn Miriam, Clearview AI, TikTok, and the collection of facial images in international law. In: Chicago Unbound Ausg. 23, 2022, unter: https://chicagounbound.uchicago.edu/cjil/vol23/iss1/13/ (letzter Zugriff am 23.10.2025).

Kokotajlo Daniel / Alexander Scott / Larsen Thomas / Lifland Eli, AI 2027, 2025, unter: https://ai-2027.com/ (letzter Zugriff am 04.12.2025).

Köller Sandra, Drei Sprachmodelle und ein Wahl-O-Mat: Ein Blick auf wirtschaftspolitische Entscheidungen, In: Industrie- und Handelskammer Region Stuttgart, 2024, unter: https://www.ihk.de/stuttgart/presse/ki-o-mat-6461208 (letzter Zugriff am 13.10.2025).

Laude Lennart / Daum Andres, KI als neues Wahlkampfinstrument Offene Flanke der KI-Regulierung?. In: Verfassungsblog, 2024, unter: https://verfassungsblog.de/ki-als-neues-wahlkampfinstrument/ (letzter Zugriff am 15.11.2025).

Laufenberg Roger, Der KI-Verordnungsentwurf und biometrische Erkennung: Ein großer Wurf oder kompetenzwidrige Symbolpolitik? Der KI-Verordnungsentwurf und biometrische Erkennung. In: Friedewald Roßnagel et al. [Hrsg.]: Privatheit und Selbstbestimmung in der digitalen Welt, Nr. 1, 2022, S. 353-376.

Lee Sook Kim et al., The association of level of internet use with suicidal ideation and suicide attempts in South Korean adolescents: a focus on family structure and household economic status. In: Kaess M, Durkee T, Brunner R (Hrsg): Canadian Journal of Psychiatry, 2016, S. 243–251.

Levantino Paolo / Paolucci Federica, Advancing the Protection of fundamental Rights through AI Regulation: How the EU and the Council of Europe are shaping the Future. In: Czech, Philip / Heschl, Lisa et al. (Hrsg.), European Yearbook on Human Rights, Forthcoming, 2024, S. 1-16.

Lisowski Reiner / Scholz Lydia / Trüe Christiane, Change and transformation the EU way. In: Lisowski, Reiner / Scholz, Lydia / Trüe, Christiane (Hrsg.), Academia, Administration, Digitalization and Sustainability Change and Transformation the EU Way, New York, 2025, S.11-19.

Markov Grey / Charbel-Raphaël Segerie et al., AI Safety Atlas. French Center for AI Safety (CeSIA), 2025, unter: ai-safety-atlas.com (letzter Zugriff am 03.12.2025).

McGurk Brendan / Tomlinson Joe, Artificial Intelligence and Public Law, Bloomsbury Publishing, London, 2025.

McKee Robin, Seltene Erden: Wie künstliche Intelligenz den Bedarf steigert und den Abbau beschleunigt, 2021, unter: https://te.ma/art/blvifm/mckie-ki-tiefseebergbau-seltene-erden/#paper (letzter Zugriff am 14.11.2025).

Ministerium für Europa und auswärtige Angelegenheiten, Aktionsgipfel zur Künstlichen Intelligenz. In: Außenpolitik Frankreichs,, 2024, unter: https://www.diplomatie.gouv.fr/de/aussenpolitik-frankreichs/digitale-diplomatie/article/aktionsgipfel-zur-kunstlichen-intelligenz-10-11-02-2025 (letzter Zugriff am 14.08.2025).

Montag Christian, et al., How to overcome taxonomical problems in the study of Internet use disorders and what to do with "smartphone addiction"? In: Journal of behavioural addictions, Nr. 9, 2021, S.908-914.

Müller Manuel, Die Europawahl 2024: Mehr als ein Rechtsruck. In: Integration 47, 2024, S. 276-293.

Müller Uwe, Olympia 2024: Polizeipräfektur in Paris erlaubt KI-gestützte Videoüberwachung. In: Digital Chiefs, 2024, unter: https://www.digital-chiefs.de/olympia-2024-polizeipraefektur-in-paris-erlaubt-ki-gestuetzte-videoueberwachung/ (letzter Zugriff am 14.10.2025).

Murray Daragh, The Chilling Effects of Surveillance and Human Rights: Insights from Qualitative Research in Uganda and Zimbabwe. In: Journal of Human Rights Practice, 16/2024, S. 397–412.

Nardocci Costanza, Artificial intelligence at the crossroads between the European Union & the Council of Europe: Who safeguards what & how?, Italian Journal of Public Law (IJPL), Mailand, 2024, S. 165-196

Nurkic Benjamin / Hogic Nedim, Lithium, Law, and the Limits of EU Sustainability-The EU's Corporate Sustainability Framework and Non-EU Countries, In: Verfassungsblog, 2025, unter: https://verfassungsblog.de/lithium-law-and-the-limits-of-eu-sustainability/ (letzter Zugriff am 14.11.2025).

o.V., Zeit Online, Pariser KI-Gipfel spricht sich für ethische und nachhaltige KI aus, https://www.zeit.de/digital/2025-02/paris-kuenstliche-intelligenz-abschlusserklaerung-ethisch, (letzter Zugriff am 19.09.2025).

OceanCare, Meeresschutz mit globaler Wirkung, 2022, unter: https://www.oceancare.org (letzter Zugriff am 02.10.2025).

OECD, Empfehlung des Rates zu künstlicher Intelligenz. In: OECD/LEGAL/0449, 2024.

OECD, Künstliche Intelligenz in der Gesellschaft. In: OECD Publishing, Paris, 2020, unter: https://doi.org/10.1787/6b89dea3-de (letzter Zugriff am 12.10.2025).

Oellig Tobias, Ausgeguckt. In: Amnesty international, 2019, unter: https://www.amnesty.de/informieren/amnesty-journal/europa-massenueberwachung-gesichtserkennung-eu-gesetz-ausgeguckt (letzter Zugriff am 07.11.2025).

Olevato Giulia, Paving the path towards general purpose AI systems regulation in the AI Act: an analysis of the Parliament's and Council's proposals, 2024, unter: https://www.medialaws.eu/rivista/paving-the-path-towards-general-purpose-ai-systems-regulation-in-the-ai-act-an-analysis-of-the-parliaments-and-councils-proposals/ (letzter Zugriff am 25.11.2025)

Pei Minxin, The sentinel state: Surveillance and the survival of dictatorship in China. In: Harvard University Press, Springer Verlag, 2024.

Peterson Dahlia / Hoffmann Samantha, Geopolitical implications of AI and digital surveillance adoption. In: Brookings, 2022, unter: https://www.brookings.edu/articles/geopolitical-implications-of-ai-and-digital-surveillance-adoption/ (letzter Zugriff am 10.11.2025).

Raluca Csernatoni, The EU's AI Powerplan: Between innovation and regulation. In: Carnegie Europe, Brüssel, 2025 unter: https://carnegie-production-assets.s3.amazonaws.com/static/files/Csernatoni_The%20EU's%20AI%20Power%20Play_Between%20Deregulation%20and%20Innovation.pdf (letzter Zugriff am 11.11.2025).

Rani Anku et al., Dialogues with AI Reduce Beliefs in Misinformation, 2025, unter: https://arxiv.org/html/2510.01537v1 (letzter Zugriff am 29.10.2025).

Reinhold Thomas / Reuter Christian, Artificial Intelligence and Cyber Weapons, In: Reuter Christian: Information Technology for Peace and Security, Springer, Darmstadt, 2024, S.335-351.

Renda Andreas / Balland Pierre-Alexandre, Stargate and the fight for AI supremacy- this is Europe's wake-up call. In: Centre for European Policy Studies, 2025, unter: https://www.ceps.eu/stargate-and-the-fight-for-ai-supremacy-this-is-europes-wake-up-call/ (letzter Zugriff am 29.10.2025).

Reuters, Trump to issue order creating national AI rule, 2025, unter: https://www.reuters.com/world/trump-says-he-will-sign-executive-order-this-week-ai-approval-process-2025-12-08/ (letzter Zugriff am 08.12.2025).

Rezk Eman, Improving Skin Color Diversity in Cancer Detection: Deep Learning Approach, https://pmc.ncbi.nlm.nih.gov/articles/PMC10334920/ (letzter Zugriff am 10.10.2025).

Roberts Huw et al., The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. In: Ethics, governance, and policies in artificial intelligence, Springer International Publishing, Cham, 2021, S. 47-79.

Rolf Steven, China's regulations on algorithms: Context, impact, and comparisons with the EU. In: Friedrich Ebert Stiftung Briefing, 2023, unter: https://library.fes.de/pdf-files/bueros/bruessel/19904.pdf (letzter Zugriff am 29.10.2025).

Rowe Niamh, It destroyed me completely. Kenyan moderators decry toll of trianing AI models. In: The Guardian, 2023, unter: https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai (letzter Zugriff am 14.09.2025).

Ruhmann Ingo / Bernhardt Ute, Information Warfare: From Doctrine to Permanent Conflict. In: Reuter Christian: Information Technology for Peace and Security, Springer, Darmstadt, 2024, S.69-91.

Sara Hsu Sara / Chong Ja-Ian / Rorry Daniels, Finding Firmer Ground, The Role of High Technology in US-China Relations, 2023, unter: https://www.cartercenter.org/wp-content/uploads/2023/02/finding-firmer-ground-the-role-of-high-technology-in-u.s.-china-relations.pdf (letzter Zugriff am 01.12.2025).

Schröder Thilo Daniel et al., How Malicious AI Swarms Can Threaten Democracy The Fusion of Agentic AI and LLMs Marks a New Frontier in Information Warfare, 2025, unter: https://arxiv.org/html/2506.06299v3#bib.bib1 (letzter Zugriff am 04.11.2025).

Schuett Jonas, Risk Management in the Artificial Intelligence Act, Cambridge University Press, 2022, unter: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/2E4D5707E65EFB3251A76E288BA74068/S1867299X23000016a.pdf/risk-management-in-the-artificial-intelligence-act.pdf (letzter Zugriff am 14.08.2025).

Schweizerische Eidgenossenschaft, Electronic Monitoring by year of ending and the main sentence executed, 2025, unter: https://www.bfs.admin.ch/bfs/rm/home.assetdetail.32809187.html (letzter Zugriff am 08.11.2025).

Schwemer Sebastian / Koivisto Ida, A Warm Body in the Loop-Rethinking Human Control of AI in EU Tech Regulation. In: Verfassungsblog, 2025, unter: https://verfassungsblog.de/warm-body-in-the-loop/ (letzter Zugriff am 03.12.2025).

Shirado Hirukazi / Christakis Nicholas, Locally noisy autonomous agents improve global human coordination in network experiments. In: Nature 545, 2017, S. 370–374.

Shneiderman Ben, Human-Centered AI, Oxford University Press, 2022, unter: https://ebookcentral-1proquest-1com-100f089tl028b.emedien3.sub.uni-hamburg.de/lib/subhh/detail.action?docID=6838500 (letzter Zugriff am 14.11.2025).

Sirakov David, Das System als Endgegner. In: Bundeszentrale für politische Bildung, 2025, unter: https://www.bpb.de/themen/nordamerika/usa/561711/das-system-als-endgegner/ (letzter Zugriff am 03.11.2025).

Stattkus Daniel, Nutzung von Künstlicher Intelligenz zur zielgruppengerechten Unterstützung von Lernprozessen, HMD Praxis der Wirtschaftsinformatik, 2025.

Stöcker Christian, Künstliche Intelligenz Eine Schrottblase bedroht die Geldblase. In: Spiegel, 2025, unter: https://www.spiegel.de/wissenschaft/technik/kuenstliche-intelligenz-ai-slop-koennte-die-ki-blase-platzen-lassen-a-e7be1ad9-8053-4c75-a12a-bf5aa7929bb3 (letzter Zugriff am 03.12.2025).

Streinz Thomas, The Evolution of European Data Law, in: Craig, Paul / de Búrca Gráinne (Hrsg.), The Evolution of EU Law, 3. Auflage, Oxford University Press, Oxford 2021, S. 902-936.

Swissinfo, Criminalité: le bracelet électronique a fait ses preuves, 2025, unter: https://www.swissinfo.ch/fre/criminalit%c3%a9%3a-le-bracelet-%c3%a9lectronique-a-fait-ses-preuves/89866273 (letzter Zugriff am 08.11.2025).

Tagesschau, Irische Datenschutzbehörde 1,2 Milliarden Euro Strafe für Meta, 2023, unter: https://www.tagesschau.de/ausland/europa/milliardenstrafe-facebook-meta-100.html (letzter Zugriff am 06.12.2025).

Teer Joris, Beyond Trump: Xi's price wars and weaponisation of critical raw materials threaten European prosperity. In: European Union Insitute for Security Studies, 2025, unter: https://www.iss.europa.eu/publications/commentary/beyond-trump-xis-price-wars-and-weaponisation-critical-raw-materials (letzter Zugriff am 28.10.2025).

The White House, Removing Barriers to American Leadershipe in Artificial Intellegence, 2025, unter: https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/ (letzter Zugriff am 19.11.2025).

The White House: Action Plan AI race, 2025, unter: https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf (letzter Zugriff am 20.11.2025).

Toju Duke, Building Responsible AI Algorithms: A Framework for Transparency, Fairness, Safety, Privacy, and Robustness, Apres, London, 2023.

UNESCO, Recommendation on the Ethics of Artificial Intelligence, 2023, unter: https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence (letzter Zugriff am 14.08.2025).

*UNESCO, R*ecommendation on the Ethics of Artificial Intelligence, 2023, unter: https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence (letzter Zugriff am 14.08.2025).

UNHCR; Refugee Status Determination, k.D., unter: https://www.unhcr.org/what-we-do/protect-human-rights/protection/refugee-status-determination (letzter Zugriff am 28.11.2025).

United Nations, Artificial Intelligence for Climate Action in Developing Countries: Opportunities, Challenges and Risks, 2024, unter: https://unfccc.int/ttclear/misc_/StaticFiles/gnwoerk_static/AI4climateaction/28da5d97d7824d16b7f68a225c0e3493/a4553e8f70f74be3bc37c929b73d9974.pdf (letzter Zugriff am 19.08.2025).

United Nations, Racism and AI: "Bias from the past leads to bias in the future", 2024, unter: https://www.ohchr.org/en/stories/2024/07/racism-and-ai-bias-past-leads-bias-future (letzter Zugriff am 19.08.2025).

Verduyn Phillip / Kross Ethan, Do social network sites enhance or undermine subjective well-being? A critical review. In: Social Issues and Policy Review, Nr. 11, 2017, S.274–302.

Verenkotte Clemens, KI-Einsatz im Gaza-Krieg? Schwere Vorwürfe gegen Israels Armee. In: Tagesschau, 2024, unter: https://www.tagesschau.de/ausland/israel-gazastreifen-ki-102.html (letzter Zugriff am 10.11.2025).

Verordnung (EU) 2016/679 vom 27.4.2016 (ABl. 2016 Nr. L 119/1; 2016 Nr. L 314/72; 2018 Nr. L 127/2; 2021 Nr. L 74/35).

Vítor Bernardo, Explainable Artificial Intellegence. In: Privacy Unit of the European Data Protection Supervisor (EDPS), 2023.

Völkerrechtsblog, Smarte Kriege: Künstliche Intelligenz in bewaffneten Konflikten, 2024, unter: https://voelkerrechtsblog.org/de/38-smarte-kriege-kuenstliche-intelligenz-in-bewaffneten-konflikten/ (letzter Zugriff am 10.11.2025).

Vorobyeva Klarisa et al., Personalized learning through AI: Pedagogical approaches and critical insights, 2025, unter: https://www.cedtech.net/download/personalized-learning-through-ai-pedagogical-approaches-and-critical-insights-16108.pdf (letzter Zugriff am 02.11.2025).

Whittaker Carl, Die Gefühle der KI: Philosophie, Technik und Macht künstlicher Emotionen, University Press, Bremen, 2025.

Wieringa Maranke, Hey SyRI, tell me about algorithmic accountability: Lessons from a landmark case. In: Cambridge University Press, 2022, unter: https://www.cambridge.org/core/services/aop-cambridge-core/content/view/22A3086554B0486BB4BBAF2D5A33A3D0/S2632324922000396a.pdf/hey-syri-tell-me-about-algorithmic-accountability-lessons-from-a-landmark-case.pdf (letzter Zugriff am 12.11.2025).

Wiese Lisa, Gaza, Artificial Intelligence, and Kill Lists. In: Verfassungsblog, 2024, unter: https://verfassungsblog.de/gaza-artificial-intelligence-and-kill-lists/ (letzter Zugriff am 10.11.2025).

Woll Elisabeth Hannah, Das Rahmenübereinkommen des Europarats über KI und Menschenrechte, Demokratie und Rechtsstaatlichkeit und die Verordnung der Europäischen Union über Künstliche Intelligenz: Parallelen und Unterschiede, ZEuS 3/2025, S. 509-535

Woll Laura, Die Rechtsprechung des EGMR zu häuslicher Gewalt und die Auswirkungen der Istanbul-Konvention, 2024.

Woods Dwayne, AI as a tool for surveillance: China's concave trilemma. In: Journal of Chinese Political Science, 2025, S.1-35.

Young Brett, What is RLHF? Reinforcement learning from human feedback for AI alignment, In: Weights and Biases, 2024, unter: https://wandb.ai/byyoung3/huggingface/reports/What-is-RLHF-Reinforcement-learning-from-human-feedback-for-AI-alignment--VmlldzoxMzczMjEzMQ (letzter Zugriff am 12.11.2025).

Yuval Abraham, »Lavendel«: Die KI-Maschine, die Israels Angriffe in Gaza lenkt. In: NDJournalismus von Links, 2024, unter: https://www.nd-aktuell.de/artikel/1181223.gaza-krieg-lavendel-die-ki-maschine-die-israels-angriffe-in-gaza-lenkt.html (letzter Zugriff am 30.11.2025).

Zhang Angela Huyue, The promise and perils of China's regulation of artificial intelligence, Columbia Law School of Transnational Law, Ausg. 63, 2025, S. 101-157.

II. Eidesstattliche Erklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit mit dem Titel „Menschenrechtsrisiken in KI-Systemen und der Balanceakt regulatorischer Maßnahmen: Eine Analyse des Spannungsfelds zwischen Menschenrechtsförderung und -gefährdung am Beispiel der EU-KI-Verordnung und des Rahmenübereinkommens des Europarats“ selbstständig verfasst habe und dass diese Arbeit weder veröffentlicht wurde noch zu einem anderen Zweck als diesem Prüfungsverfahren verwendet worden ist. Ferner versichere ich, dass ich keine anderen Quellen oder Materialien verwendet habe als diejenigen, die im Literaturverzeichnis aufgeführt sind. Die den verwendeten Werken entnommenen Passagen habe ich, sei es wörtlich oder sinngemäß, kenntlich gemacht. Mir ist bewusst, dass diese Arbeit mit Plagiatssoftware überprüft werden kann.

Hiermit erteile ich zudem meine Zustimmung, dass das Europa-Institut das von mir verfasste Abstract weiterverwenden darf. Das Abstract darf zu Informationszwecken auf der Website des Europa-Instituts veröffentlicht werden.

Elisabeth Hannah Woll

Saarbrücken, 04.01.2026