

Privatheit und Selbstbestimmung
in der digitalen Welt



Friedewald | Roßnagel | Karaboga | Geminn [Hrsg.]

Datenschutz und Digitalpolitik in krisenhaften Zeiten

Ausgewählte Beiträge der 10. Konferenz
der Plattform Privatheit
vom 1. und 2. Oktober 2025



Nomos

Privatheit und Selbstbestimmung in der digitalen Welt
Privacy and Self-Determination in the Digital World

herausgegeben von | edited by

Dr. Michael Friedewald

Prof. Dr. Alexander Roßnagel

Band | Volume 7

Michael Friedewald | Alexander Roßnagel
Murat Karaboga | Christian Geminn [Hrsg.]

Datenschutz und Digitalpolitik in krisenhaften Zeiten

Ausgewählte Beiträge der 10. Konferenz
der Plattform Privatheit
vom 1. und 2. Oktober 2025



Nomos

Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt

© Titelbild: Magdalena Vollmer

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

1. Auflage 2026

© Die Autor:innen

Publiziert von

Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden
www.nomos.de

Gesamtherstellung:

Nomos Verlagsgesellschaft mbH & Co. KG
Waldseestraße 3–5 | 76530 Baden-Baden

ISBN (Print): 978-3-7560-3584-7

ISBN (ePDF): 978-3-7489-6643-2

DOI: <https://doi.org/10.5771/9783748966432>



Onlineversion
Inlibra



Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz.

Vorwort

Die digitale Welt befindet sich in einem tiefgreifenden Wandel: Datenschutz und Digitalpolitik geraten zunehmend unter Druck – durch geopolitische Spannungen, den rasanten Fortschritt innovativer Technologien und Geschäftsmodelle sowie durch eine wachsende Diskrepanz zwischen regulatorischen Ambitionen und deren Durchsetzbarkeit. Der EU-Rechtsrahmen für den Schutz personenbezogener Daten, der über Jahre als Vorbild für eine wertorientierte Digitalpolitik galt, sieht sich heute mit neuen Herausforderungen konfrontiert: von geopolitischem Druck auf europäische Regulierungen bis hin zu ungeklärten Fragen rund um den Einsatz von Künstlicher Intelligenz, Plattformökonomie und datengetriebener Verhaltenssteuerung. Was steht auf dem Spiel, woher kommen die Bedrohungen – und was kann und soll dagegen unternommen werden?

Um sich diesen Herausforderungen im Rahmen eines über die Wissenschaft hinausweisenden Diskurses zu stellen, veranstaltete die vom Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) geförderte „Plattform Privatheit“ am 1. und 2. Oktober 2025 in Berlin die *10. Interdisziplinäre Jahrestagung* unter dem Titel „Datenschutz und Digitalpolitik in krisenhaften Zeiten“. Dieses Jubiläum markiert einen wichtigen Meilenstein: Seit nunmehr zehn Jahren bietet die Jahrestagung einen zentralen Ort für den interdisziplinären Austausch über Privatheit, Datenschutz und informationelle Selbstbestimmung. Der vorliegende Band stellt die wichtigsten Vorträge vor und reflektiert die dort angestoßenen Diskussionen.

Die Plattform Privatheit vernetzt interdisziplinäre wissenschaftliche Projekte, die vom BMFTR im Rahmen der Förderlinie „Plattform Privatheit – Bürgerinnen und Bürger bei der Wahrnehmung des Grundrechts auf informationelle Selbstbestimmung unterstützen“ gefördert werden. Koordiniert wird sie vom Fraunhofer-Institut für System- und Innovationsforschung (ISI) in Karlsruhe und der Projektgruppe verfassungsverträgliche Technikgestaltung (provet) an der Universität Kassel. Sie erarbeitet Orientierungswissen für den öffentlichen Diskurs und bereitet Forschungsergebnisse für Zivilgesellschaft, Politik, Wissenschaft und Wirtschaft auf – zu sozialen, rechtlichen und ethischen Aspekten von Privatheit, Datenschutz und informationeller Selbstbestimmung.

Wir freuen uns, nun diesen Konferenzband präsentieren zu können. Wir danken insbesondere den Autorinnen und Autoren für die Ausarbeitung ihrer Vorträge zu Fachaufsätzen. Ebenso zum Dank verpflichtet sind wir dem wissenschaftlichen Beirat der „Plattform Privatheit“ sowie den Kolleginnen und Kollegen, die die in diesem Band veröffentlichten Texte begutachtet haben. Die Konferenz wäre ohne die vielfältige Unterstützung durch das interdisziplinäre Kollegium nicht möglich gewesen. Wir danken insbesondere all jenen, die organisatorisch oder inhaltlich an der Vorbereitung und Durchführung der Konferenz mitgewirkt haben, darunter vor allem Sabine Muhr, Greta Runge, Dr. Frederik Metzger, Yuwen Zhang, Paula Pons und Gvantsa Zakariadze (Fraunhofer ISI). Darüber hinaus danken wir Barbara Ferrarese (Fraunhofer ISI) für die professionelle und engagierte Wissenschaftskommunikation sowie Jan Eggers (Hessischer Rundfunk) für die lebendige und fachkundige Moderation. Magdalena Vollmer danken wir herzlich für die kreative grafische Gestaltung des Bucheinbandes. Für die angenehme und zielführende Zusammenarbeit mit dem Nomos-Verlag danken wir Dr. Sandra Frey.

Unser besonderer Dank gilt Dr. Sebastian Jester und Daniel Behrendt (BMFTR) für die Förderung der Plattform Privatheit sowie die engagierte Unterstützung unserer Forschungsthemen. Auch danken wir Florian Till Patzer, der für den Projektträger VDI/VDE-IT die Forschungsarbeiten der Plattform Privatheit, die Vorbereitung der Konferenz und das Erscheinen des Bandes konstruktiv begleitet hat.

Die Herausgeber und das Team der „Plattform Privatheit“
Karlsruhe und Kassel, im Juni 2026

Inhaltsverzeichnis

Alexander Roßnagel, Michael Friedewald, Christian Geminn und Murat Karaboga

Datenschutz und Digitalpolitik in krisenhaften Zeiten: eine
Einleitung in den Band 11

Eingeladene Vorträge

Guido Scorza

Eliza's lesson (which we didn't learn): Reflections on sentimental
chatbots and major threats to data protection and digital policy 27

Katja Muñoz und Emma Carlotta Laumann

Demokratie im KI-Zeitalter: Vom Werkzeug zu Infrastruktur 39

KI, Synthetische Medien und die Transformation des öffentlichen Raums

Moritz Griesel und Hans Brorsen

Transparenzpflichten für synthetische Medieninhalte: Rechtliche
Anforderungen und technische Realität 63

Markus Uhlmann, Eva Isabel Hoffmeister und Pascal Altenkamp

Gestaltungsspielräume und Gelingensbedingungen kooperativer
Nachrichtenplattformen: Eine vergleichende Fallstudie europäischer
Plattforminitiativen aus privatheits-theoretischer Perspektive 87

Christian Geminn, Joel Baumann, Jessica Heesen, Lisa Mühl und Klara-Aylin Wenten

- AI Companions: Gegenwärtige und zukünftige Herausforderungen
konversationeller KI-Agenten 113

Nutzerzentrierter Datenschutz: Informiertheit und Rechtsdurchsetzung in der Praxis

*Diana Schneider, Jana Koch, Claudia Schlüfter, Harald Zwingelberg,
Michael Friedewald und Nils B. Heyen*

- Einwilligen, aber informiert: Die Perspektive von Patient:innen auf
die informierte Einwilligung bei der Nutzung von Gesundheits-Apps 139

Daniel Guagnin, Antonios Hazim und Volkan Sayman

- Informierte Einwilligung als Abwägungsproblem: Verstehen,
Bewerten und Handeln im Kontext personenbezogener
Datenverarbeitung 159

Daniela Pöhn

- DSGVO vs. Online-Dienste: Realität der Datenauskunft 181

*Florian Dehling, Sebastian Koch, Luigi Lo Iacono, Lindrit Kqiku,
Delphine Reinhardt und Harald Zwingelberg*

- Privatheits-relevante Eigenschaften von IoT-Produkten für
Verbraucher:innen 199

*Tanja Lehmann, Franziska Baum, Albrecht Kurze, Arne Berger und
Andreas Bischof*

- Einfache Sensoren, komplexe Folgen: Privatheit als soziale Praxis im
vernetzten Haushalt 221

Governance-Strukturen und staatliches Handeln im technologischen Wandel

Katharina Schilke

Mehr Klarheit im Datenraum: Die Abgrenzung des Datenvermittlungsdienstes zum Infrastrukturanbieter 251

Jessica Heesen, Thomas Kunz, Matthias Meitzler, Maria Pawelec, Marina Steenbergen und Ulrich Waldmann

Maßnahmen zur Altersverifikation in Metaversen: Ethische, rechtliche und sicherheitstechnische Aspekte 273

Bruno Albert, Joanna Klauser, Andreas Hammer, Christian Lindenmeier, Sabine Pfeiffer, Dirk Heckmann und Felix Freiling

Akzeptanz staatlicher Überwachung privater Telekommunikation in Deutschland: Ein faktorielles Survey-Experiment 303

Hartmut Aden und Steven Kleemann

Unkontrollierte Nutzung Künstlicher Intelligenz durch Polizeibehörden? Fragmentierte Accountability- und Compliance-Vorkehrungen 327

Mitarbeiterinnen und Mitarbeiter dieses Bandes 353

Datenschutz und Digitalpolitik in krisenhaften Zeiten: eine Einleitung in den Band

Alexander Roßnagel, Michael Friedewald, Christian Geminn und Murat Karaboga

Wir leben in krisenhaften Zeiten – und diese gefährden Privatheit sowie individuelle und kollektive Selbstbestimmung. Krisenhafte Entwicklungen in der Geopolitik, in den Volkswirtschaften, in der Technologie, in der Sicherheit der Informationstechnik, in der inneren Sicherheit, in der öffentlichen Kommunikation und in weiteren gesellschaftlichen Bereichen verändern die Verwirklichungsbedingungen der Werte, die in der Europäischen Union das friedliche und gedeihliche Zusammenleben und Wirtschaften bestimmen sollen. Noch ist unklar, ob die damit verbundenen Herausforderungen die Werte selbst in Frage stellen, sie inhaltlich verändern oder nur ihre Umsetzung und Durchsetzung erschweren. Daher ist es von größter Bedeutung zu untersuchen, was auf dem Spiel steht, woher die Gefährdungen kommen und was man gegen sie unternehmen kann und soll.

Governance-Struktur europäischer Werte

Für die Gewährleistung europäischer Werte im Prozess der Digitalisierung ist eine Struktur entscheidend, die aus mehreren zusammenhängenden Elementen besteht: dem relevanten europäischen und auch mitgliedstaatlichen Rechtsrahmen, europäischen und mitgliedstaatlichen Institutionen für die Kontrolle und Durchsetzung des Rechtsrahmens, Adressaten, die die Vorgaben grundsätzlich akzeptieren, und politische Institutionen, die hinter diesen Werten stehen und bereit sind, den Rechtsrahmen positiv fortzuentwickeln. Gerät diese Governance-Struktur durch äußere oder innere Veränderungen unter Druck, kommt es für die Gewährleistung der Werte entscheidend darauf an, wie flexibel im Detail und wie stabil im Kern diese Governance-Struktur ist.

Die normativen Grundlagen für das gesellschaftliche Zusammenleben in Freiheit und Demokratie werden vor allem in der Europäischen Grundrechtecharta (GRCh) beschrieben. Auch in einer digitalen Welt soll die

europäische Gesellschaft geprägt sein von Menschenwürde, individueller und kollektiver Selbstbestimmung, Schutz vor Risiken und Rechtsstaatlichkeit. Der Prozess der Digitalisierung soll diese Werte nicht gefährden, sondern soll von ihnen geleitet werden. Er wird normativ einerseits von den Grundrechten auf Privatleben (Art. 7 GRCh) und Datenschutz (Art. 8 GRCh) und andererseits von den Grundrechten auf Meinungs- und Informationsfreiheit (Art. 11 GRCh), Forschungsfreiheit (Art. 13 GRCh), Berufsfreiheit (Art. 15 GRCh) sowie Unternehmensfreiheit (Art. 16 GRCh) geprägt. Bezogen auf die Verarbeitung personenbezogener Daten bietet die Datenschutz-Grundverordnung (DSGVO) auf einfachgesetzlicher Ebene einen Ausgleich zwischen diesen Grundrechten. Aus der Suche nach geeigneten Kompromissen enthält sie viele abstrakte technikneutrale Regelungen. Auch gibt sie den Mitgliedstaaten in vielen Öffnungsklauseln Möglichkeiten, jeweils ihre Vorstellungen von Datenschutz in nationalen Gesetzen – insbesondere für den öffentlichen Bereich – verbindlich zu machen.

Die Umsetzung dieses Rahmens für Ordnung und Entwicklung überlässt die Datenschutz-Grundverordnung in erster Linie den Datenverarbeitern, die sie „Verantwortliche“ nennt. Sie werden kontrolliert und beraten von unabhängigen Datenschutzaufsichtsbehörden. Diese koordinieren ihre Aufsichtspraxis im Europäischen Datenschutzausschuss. Den von der Datenverarbeitung betroffenen Personen stehen Informations- und Interventionsrechte zu. Sie können bei den Aufsichtsbehörden Beschwerde einlegen, wenn sie der Meinung sind, dass Verantwortliche gegen die Vorgaben der Datenschutz-Grundverordnung verstoßen. Diese wiederum können den Beschwerden mit Befugnissen zur Informationserhebung sowie Anordnungen und Sanktionen nachgehen.

Sowohl Verantwortliche als auch betroffene Personen können die Gerichte anrufen, wenn sie mit Entscheidungen der Aufsichtsbehörden nicht einverstanden sind. Die Gerichte prüfen diese Entscheidungen und legen offene Fragen dem Europäischen Gerichtshof zur unionsweit einheitlichen Entscheidung vor.

Auf diese Weise werden die normativen Regeln zum Schutz der Grundrechte im Prozess der Digitalisierung durch das Zusammenwirken von Verantwortlichen, betroffenen Personen, Datenschutzaufsichtsbehörden, Datenschutzausschuss, nationalen Gerichten und dem Europäischen Gerichtshof konkretisiert und umgesetzt.

Anpassungsfähigkeit

Diese Governance-Struktur bietet grundsätzlich einen elastischen Rahmen für die weitere Entwicklung der digitalen Gesellschaft. Sie bietet Innovationen der Informationstechnik und ihrer wirtschaftlichen Verwertung einen Freiraum und passt sich den dadurch entstehenden neuen Bedingungen durch angegliche Interpretation der normativen Vorgaben an. Zugleich wirkt sie auf Innovationen ein, sich innerhalb des angepassten normativen Rahmens zu halten.

Viele der technikneutralen normativen Vorgaben sind so abstrakt, dass ihre Anforderungen auch auf neue Technikentwicklungen, Geschäftsmodelle, Anwendungs- und Nutzungsformen übertragbar sind. Sie sind etwa im Bereich wirtschaftlicher Datenverarbeitung immer zulässig, wenn sie durch eine Einwilligung, durch einen Vertrag oder durch überwiegende berechnete Interessen gedeckt sind. Im Bereich öffentlicher Datenverarbeitung sind sie zulässig, wenn sie durch eine gesetzliche Regelung des Mitgliedstaats erlaubt werden. Begrenzt werden sie durch die ebenfalls sehr auslegungsfähigen Grundsätze der Zweckbindung, der Erforderlichkeit, der Datenrichtigkeit, der Transparenz und der Datensicherheit. Durch diese Konzeption können im Regelfall die innovativen Datenverarbeitungen zugelassen und zugleich die Rechte der betroffenen Personen geschützt werden. Probleme ergeben sich, wenn im öffentlichen Bereich eine gesetzliche Erlaubnis fehlt oder im Wirtschaftsbereich ein Erlaubnistatbestand für die Verarbeitung besonders schützenswerter Daten wie z.B. Gesundheitsdaten, biometrische Daten oder Daten zur ethnischen Herkunft oder politischen Meinung nicht greift.

Die innovative Auslegung der abstrakten Texte praktizieren die Verantwortlichen. Ob sie sich durchsetzt, hängt entscheidend von den Datenschutzaufsichtsbehörden und den Gerichten ab. Die Aufsichtsbehörden können unzulässige Datenverarbeitungen verhindern und sanktionieren. Verantwortliche und betroffene Personen können die jeweils zuständigen Gerichte anrufen. Zu innovativen Auslegungen wird es immer wieder zu Streitigkeiten kommen, deren Ergebnisse in der Union abschließend der Europäische Datenschutzausschuss und vor allem der Europäische Gerichtshof feststellen. Dabei haben die europäischen und nationalen Gesetzgeber formal immer die Möglichkeit, die normativen Texte ihren politischen Vorstellungen anzupassen.

Aber dieses System der flexiblen Reaktion auf Innovationen und andere Veränderungen seiner Verwirklichungsbedingungen kann an seine Grenzen

stoßen. Dies kann zum einen der Fall sein, wenn die neuen Herausforderungen zu groß und zu vielfältig werden und zu schnell aufeinander folgen. Dann können die Institutionen, die für eine Anpassung von Normativität und Innovation sowie für einen Ausgleich der Interessen der Datenverarbeiter und der betroffenen Personen zuständig sind, überfordert sein. Zum anderen können die Wirkungen der Veränderungen zu stark oder zu unübersichtlich oder die Differenzen zwischen der neuen Wirklichkeit und dem normativen Ziel zu gravierend sein. Dann können der Kern des Grundrechtsschutzes und die grundsätzlichen Möglichkeiten, eine praktische Konkordanz der konkurrierenden Grundrechte zu finden, betroffen sein. Dadurch kann eine Krise für die Gewährleistung entscheidender europäischer Werte entstehen.

Anzeichen für krisenhafte Entwicklungen

Jahrelang schien die Entwicklungsrichtung der europäischen Digitalpolitik klar: Angesichts der vielfältigen Herausforderungen würden die Europäische Union und Deutschland mit Regulierung vorangehen, um ihre Bürgerinnen und Bürger und die europäischen Werte zu schützen, gleichzeitig digitale Innovationen voranzutreiben und so die Europäische Union wieder zu einem ernstzunehmenden digitalen Akteur zu machen. Zahlreiche Digitalrechtsakte neben der Datenschutz-Grundverordnung – etwa zur Verantwortung für Digitale Dienste, zum Wettbewerb zwischen Digitalen Diensten, zur Datennutzung, zur Daten-Governance, zur Künstlichen Intelligenz und zum Europäischen Raum für Gesundheitsdaten – zeugen von dieser Politik.

Im Bereich des Datenschutzes hat sich seit der Verabschiedung der Datenschutz-Grundverordnung vor zehn Jahren die europäische Governance-Struktur weitgehend bewährt. Der Europäische Gerichtshof hat die Zahl seiner Entscheidungen zu Datenschutzfragen von durchschnittlich drei Entscheidungen pro Jahr vor 2020 auf 15 im Jahr 2022, 23 im Jahr 2023 und knapp 30 in den Jahren 2024 und 2025 gesteigert. Mit diesen Urteilen hat er viele praktisch relevante Fragen beantwortet und durch eine inzwischen gefestigte Rechtsprechung in diesen Fragen unionsweit Rechtssicherheit erreicht.

Der Europäische Datenschutzausschuss hat bis Juni 2026 in 121 Plenarsitzungen unzählige Beschlüsse gefasst und inzwischen 120 hilfreiche Leitlinien und 45 leitende andere Empfehlungen veröffentlicht, die zu einer

gleichartigen Auslegung und Anwendung der Datenschutz-Grundverordnung in der Union beitragen sollen. Sie werden durch Expert Subgroups des Europäischen Datenschutzausschusses vorbereitet, in denen nahezu alle Aufsichtsbehörden der Mitgliedsstaaten intensiv mitarbeiten.

In Deutschland hat die Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder noch mehr Beschlüsse, Entschlüsse, Orientierungshilfen, Kurzpapiere und Anwendungshinweise verabschiedet, die eine bundesweit einheitliche Anwendung der Datenschutz-Grundverordnung ermöglichen. Diese unions- und deutschlandweite Zusammenarbeit der Aufsichtsbehörden hat sich eingespielt und funktioniert. Alle diese Ergebnisse zeugen von dem Bemühen der Aufsichtsbehörden, die Schutzversprechen auch im Alltag in allen Bereichen und Regionen umzusetzen.

Dennoch sind die Datenschutz-Grundverordnung, die beschriebene Governance-Struktur europäischer Werte und die europäische Digitalpolitik in jüngster Zeit unter Druck geraten. Zu viele und zu starke Herausforderungen wirken gleichzeitig auf sie ein. Zunehmend zeigen sich Anzeichen für krisenhafte Entwicklungen:

Geopolitische Herausforderungen

Datenschutz und Digitalpolitik in der Europäischen Union erfahren durch geopolitische Spannungen zunehmend Herausforderungen. Die Abhängigkeit Europas und Deutschlands in wichtigen Technikfeldern von anderen Volkswirtschaften hat sich im Bereich der Informationstechnik gravierend verstärkt und scheint sich weiter zu verstärken. Dies reicht von Rohstoffen und Hardware über Rechenzentren, Betriebssysteme und Cloud Computing bis hin zu vielen Ausprägungen der Anwendungssoftware. 9 von 10 Unternehmen in Deutschland sind hinsichtlich relevanter Informationstechnik und digitaler Dienste von außereuropäischen Anbietern abhängig.

Diese Abhängigkeiten führen zu Wettbewerbsnachteilen auf dem Weltmarkt ebenso wie auf dem heimischen Markt, bewirken durch Lock-in-Effekte Einschränkungen der wirtschaftlichen Entscheidungsfreiheit, führen zu Monopolisierungstendenzen, behindern Innovationen und erleichtern es, erfolgreiche europäische Start-ups aufzukaufen. Zugleich führt die Sammlung extrem großer Mengen personenbezogener Daten durch ausländische Konkurrenten und deren Zusammenführung zu Personenprofilen zu einer ungeheuren Machtsteigerung, da diese zur Steuerung individuel-

len und kollektiven Verhaltens nicht nur im Bereich der kommerziellen Werbung, sondern auch für politische Kommunikation verwendet werden können.

Mit der wirtschaftlichen Abhängigkeit steigt auch die politische und ökonomische Erpressbarkeit. Die gegenwärtige Regierung der USA mischt sich stärker als je eine amerikanische Regierung zuvor in die innenpolitischen Angelegenheiten europäischer Staaten ein und versucht, auf deren Entwicklung Einfluss zu nehmen. Dass sie nicht davor zurückschreckt, die technischen Abhängigkeiten von US-Software auszunutzen, zeigt die „Strafaktion“ gegen den Internationalen Strafgerichtshof in Den Haag und seine Mitarbeiter im Februar 2025. Diese Erpressbarkeit gefährdet die demokratische Willensbildung in Europa, die Anwendung europäischen Rechts und den Schutz der Grundrechte europäischer Bürgerinnen und Bürger.

Die Bemühungen um digitale Souveränität in der Union und in den Mitgliedstaaten nehmen zu, erleiden aber auch viele Rückschläge. Noch immer steigen die Umsätze der Unternehmen, die diese Abhängigkeit in Europa verursachen. Noch immer werden höhere Beträge an diese Unternehmen gezahlt, als in europäische IT-Angebote investiert und für europäische IT-Forschung ausgegeben. Die geopolitischen Herausforderungen und die damit verbundene krisenhafte Situation werden nur bewältigt werden können, wenn Politik, staatliche Stellen und Unternehmen in ihrer IT-Beschaffung massiv umsteuern, europäische Alternativen aufbauen und ihnen eine wirtschaftliche Durchsetzungschance geben. Ob der von der Europäischen Kommission im Juni 2026 vorgeschlagene Cloud and AI Development Act (CADA) einen Richtungswechsel hin zu mehr digitaler Souveränität bewirkt, wird sich zeigen.

Datenschutz als Wettbewerbsnachteil

Neben der Datenschutz-Grundverordnung hat die Europäische Union mit dem Digital Services Act und dem Digital Markets Act Regelungen getroffen, die für einen fairen Wettbewerb in der Union und für die Wahrnehmung von Verantwortung durch die Anbieter von sehr großen Plattformen sorgen sollen. Von diesen Plattformen sind viele andere Anbieter und nahezu alle Nutzerinnen und Nutzer digitaler Dienste abhängig. Obwohl diese Regelungen auch für ausländische Anbieter gelten, versuchen Anbieter aus den USA und China sich diesen europäischen Regelungen zu entziehen oder sie zu ignorieren.

Sie sammeln und teilen weiterhin personenbezogene Daten, die zur Verhaltensanalyse und -beeinflussung genutzt werden, und ignorieren dabei die Rechte der Nutzerinnen und Nutzer. Für deren Selbstbestimmung sind die Transparenz der Datenverarbeitung und die Möglichkeit individueller freier Entscheidung essentiell. Die Datenverarbeitung wird jedoch nur selten auf eine wirklich informierte Einwilligung gestützt. Ob es gegen diese Praktiken ausreichend ist, dass die Europäische Union mit einem Digital Fairness Act neue Instrumente gegen unfaire Praktiken der Internet-Werbeindustrie vorbereitet, ist zu bezweifeln.

Mit ihren Angeboten wenden sich große Plattformen auch gezielt an Kinder und Jugendliche und verarbeiten deren personenbezogene Daten. Dadurch ist aber nicht nur deren Datenschutz betroffen, sondern auch deren gesunde Entwicklung und der Schutz vor nicht kindgerechten Inhalten (z. B. Pornografie und Gewalt). Anforderungen an eine geeignete und datensparsame Altersverifikation und einen wirksamen Ausschluss von Minderjährigen von sie gefährdenden Angeboten werden weitgehend ignoriert.

Insbesondere im Datenschutz versuchen US-Konzerne ihre eigene Rechtsordnung, die weltweit gelten soll, zu ihren Gunsten in Form von Vertragsklauseln auch in Europa durchzusetzen. Sie wollen auf diese Weise ihre Angebote, auch wenn sie nicht datenschutzkonform sind, etablieren und dadurch ungerechtfertigte Marktvorteile erringen. Dafür nehmen sie auch immer wieder hohe Geldbußen der Europäischen Kommission und der Datenschutzaufsichtsbehörden in Kauf. Diese Praxis führt jedoch dazu, dass europäische Anbieter, die sich an die Vorgaben des europäischen Datenschutzrechts halten, Wettbewerbsnachteile erleiden.

Diese rechtliche Rahmung des Digitalsektors in Europa nimmt der amtierende US-Präsident immer wieder zum Ziel amerikanischer Handels- und Geopolitik. Er will damit europäische Regulierungen beeinflussen, die Anwendung ungeliebter rechtlicher Vorgaben verhindern und insbesondere Maßnahmen gegenüber US-Konzernen in Form von Bußgeldern und Wettbewerbsstrafen abwenden.

Dies stellt die Europäische Union und Deutschland einerseits vor die Frage, inwieweit sich bestehende Regelungen angesichts der Drohungen weiter gegenüber US-Konzernen durchsetzen lassen, und beeinflusst andererseits die Diskussion, wie die Digitalpolitik weiterentwickelt werden kann. Bezüglich der Durchsetzung bestehender Gesetze zeigen sich Organe der Europäischen Union bislang standfest. Es gibt aber auch andere Entwicklungen: So führen die in der KI-Omnibus-Verordnung 2026 fest-

gelegten zeitlichen Verschiebungen des Inkrafttretens der Verordnung für Künstliche Intelligenz (KI-VO) vor allem zu Vorteilen für US-Anbieter. Sie können sich bis dahin, ohne die Regeln der KI-VO einhalten zu müssen, auf dem europäischen Markt etablieren und entsprechende Abhängigkeiten begründen.

Technologische Entwicklung

Die europäische Digitalpolitik und insbesondere das europäische Modell, Datenschutz zu regeln und durch eine spezifische Governance-Struktur zu gewährleisten, geraten in jüngster Zeit durch die breite und rasante technologische Entwicklung unter Druck.

Die Datenschutz-Grundverordnung wurde vor inzwischen zehn Jahren verabschiedet. Sie baut auf der Europäischen Datenschutz-Richtlinie von 1995 auf und hat viele ihrer Regelungen übernommen. Sie beruht somit wesentlich auf Konzepten zur Gewährleistung informationeller Selbstbestimmung, die schon mehrere Jahrzehnte alt sind. Zwar haben die Unionsgesetzgeber versucht, den Einfluss der weiteren technologischen Entwicklung auf die Datenschutz-Grundverordnung dadurch zu begrenzen, dass sie ihre Regelungen so weit wie möglich technikneutral formuliert haben. Dennoch hatten sie unvermeidbar Vorstellungen von den zu regelnden Problemen, die ihrem Erfahrungshorizont mit Informationstechniken um 2016 und davor entsprachen. Daher liegen auch den angeblich technikneutralen Regelungen der Datenschutz-Grundverordnung unweigerlich Technikbilder aus der Zeit ihrer Entstehung zugrunde.

Neue Technologien, neue Geschäftsmodelle und neue gesellschaftliche Nutzungsformen verändern die Verwirklichungsbedingungen rechtlicher Regelungen. Aufgrund der beschriebenen Flexibilität der beschriebenen Governance-Struktur der Datenschutz-Grundverordnung muss nicht jede technische Änderung normativ oder durch Veränderung der Datenschutzpraxis nachvollzogen werden. Aber die neuesten Entwicklungen sind inzwischen so umfassend und gravierend, dass sie diese Flexibilität zu überfordern scheinen.

Ein wichtiges Beispiel hierfür ist die ubiquitäre Datenverarbeitung. Immer mehr alltägliche Geräte sind mit Datenverarbeitungskapazität und Kommunikationsfähigkeit ausgestattet, immer mehr Räume und Umgebungen, in denen Menschen sich aufhalten, sind mit Sensoren ausgestattet. Informationstechnik dringt in immer mehr Gesellschaftsbereiche, Ört-

lichkeiten und Handlungen ein und verändert soziale Praktiken in der Öffentlichkeit, der Familie und der Privatsphäre. Die Grenze zwischen datenschutzrechtlich irrelevanter privater Datenverarbeitung und relevanter Datenverarbeitung in der gesellschaftlichen Sphäre verschwimmt. Auf der einen Seite haben einzelne Personen – wörtlich genommen – in ihrer Hand eine größere Datenverarbeitungskapazität als vor wenigen Jahrzehnten ganze Rechenzentren und können damit ein größeres Datenschutzrisiko darstellen als früher große Datenverarbeiter. Auf der anderen Seite werden Smartphones und andere Techniken allgegenwärtiger Datenverarbeitung alltäglich von nahezu allen genutzt. Wo kann und muss gesellschaftliche Regulierung also ansetzen?

Allgegenwärtige Datenverarbeitung führt zu einer explosionsartigen Vermehrung personenbezogener Daten, die vielfältig genutzt werden. Sie ermöglichen tiefe Einblicke in Persönlichkeit und Beziehungen der betroffenen Personen. Dies ist die Grundlage für personalisierte Technikangebote und -nutzung. Personalisierung erleichtert technische Unterstützung erheblich, macht sie effektiver und angenehmer. Auf diese Weise können technische Assistenten ihre Nutzer in jeder Lebenslage unterstützen.

Personalisierung stellt jedoch zentrale Grundsätze des Datenschutzes in Frage – vor allem die Grundsätze der Zweckbindung und der Datenminimierung. Wenn ein persönlicher Assistent einen Nutzer in jeder Situation und bei jedem Problem unterstützen soll, kann seine Datenverarbeitung nicht auf einzelne Zwecke begrenzt sein. Sie muss vielmehr dem allgemeinen Zweck einer vollständigen und unvorhersehbaren Problembearbeitung dienen. Wenn der Assistent wissen soll, welche Ziele und Präferenzen sein Nutzer verfolgt und welche Bedingungen und Umstände er zu berücksichtigen hat, muss er alle Daten erheben, die für diese Aufgabe hilfreich sein können. Die Vorgabe, die Datenverarbeitung auf die wirklich erforderlichen Daten zu beschränken, führt zu keiner Einschränkung.

Herausforderungen bringt auch der Zeitaspekt. Die technologische Entwicklung erfolgt in immer kürzeren Innovationszyklen – meist schneller als die möglichen gesellschaftlichen Reaktionen. Diese erfordern wissenschaftliche Bewertungen, politische Diskurse und notwendige Verfahrensschritte der Gesetzgebung, der administrativen Behandlung oder gerichtlichen Beurteilung. Die technologische Entwicklung ist meist auch schneller als die Herausbildung adäquater sozialer und individueller Praktiken, wirtschaftlicher Anpassungen, der Erwerb neuer Kompetenzen und die Einrichtung neuer Ausbildungsgänge. Für den Datenschutz bedeutet dies: Soweit den immer neuen Herausforderungen nicht mit der Flexibilität der

Governance-Struktur europäischer Werte begegnet werden kann, wird die Diskrepanz zwischen den normativen Vorgaben der Datenschutz-Grundverordnung und ihrer operativen Umsetzung und den Herausforderungen für den Grundrechtsschutz immer größer.

Künstliche Intelligenz

Diese Herausforderungen zeigen sich vor allem bei der Künstlichen Intelligenz. Sowohl ihre Grundlagen wie auch ihre Anwendungen werden mit einer bisher ungewohnten Geschwindigkeit und in einer neuartigen gesellschaftlichen Tiefe und Breite entwickelt. In drei Jahren hat sie sich in allen Wirtschafts- und Verwaltungsbereichen ausgebreitet und wird in unterschiedlichsten Konstellationen – oft ungeordnet – genutzt.

Die Befürworter Künstlicher Intelligenz sehen vor allem die Chancen: Sie ist in vielfältigen Situationen zur Unterstützung und Übernahme von Arbeiten einsetzbar, nicht nur für Routinetätigkeiten, sondern auch für kreative Aufgaben. Sie verspricht Fortschritte in Wissenschaft und Medizin, Technologie und Kunst. Sie unterstützt Kommunikation und Kooperation und kann kommunikative Aufgaben bis hin zu emotionalen „Partnerschaften“ mit Menschen übernehmen. Künstliche Intelligenz kann Menschen in ihrer Selbstbestimmung stärken und in der Wahrnehmung anderer Grundrechte unterstützen, die Wettbewerbsfähigkeit von Unternehmen stärken und das Handeln von Verwaltungen effektivieren. Von vielen wird sie als *die* Grundlage für weiteren gesellschaftlichen Fortschritt angesehen.

Andere betonen die Risiken, die mit der Nutzung von Künstlicher Intelligenz verbunden sind: Sie verändert gesellschaftliche und wirtschaftliche Beziehungen tiefgreifend, indem sie Funktionen übernimmt, die bisher Menschen vorbehalten waren. Vor allem wird sie Arbeitsbeziehungen und Arbeitsbedingungen gravierend umgestalten – mit Auswirkungen auf Einzelne, Familien, gesellschaftliche Strukturen und Funktionen. Sie wird Lehren und Lernen inhaltlich wie prozessual auf neue Grundlagen stellen, Informationsgewinnung und -vermittlung beeinflussen sowie politische Prozesse und demokratische Willensbildung vor neue Herausforderungen stellen. Außerdem kann Künstliche Intelligenz missbraucht werden – zu Desinformation, Überwachung, Cyberangriffen und vielen weiteren unerwünschten oder unzulässigen Handlungen. Für viele bildet daher Künstliche Intelligenz die Grundlage für Befürchtungen über dystopische Entwicklungen der Gesellschaft, die nicht nur errungene und erkämpfte Werte

gefährden, sondern auch positive wertestärkende Entwicklungen verhindern.

So oder so – Künstliche Intelligenz verändert die Verwirklichungsbedingungen von Grundrechten und anderen Verfassungsprinzipien tiefgreifend – zum Guten oder zum Schlechten. Die Entscheidung darüber kann nicht einzelnen Konzernen überlassen werden. Ihre gesellschaftliche Steuerung und Gestaltung ist für eine lebenswerte Zukunft entscheidend.

Für die Governance-Struktur zur Gewährleistung der Grundrechte auf Privatleben und Datenschutz bedeutet Künstliche Intelligenz eine große Herausforderung ihrer normativen Grundlagen. Anbieter von KI-Systemen für allgemeine Anwendungen greifen weiterhin ungefragt auf personenbezogene und andere geschützte Daten zurück, um ihre KI-Modelle zu generieren. Ob und unter welchen Voraussetzungen KI-Systeme, die solche KI-Modelle nutzen, zulässig sind, ist weder geregelt noch geklärt. Die Datenschutzgrundsätze der Zweckbindung und der Minimierung personenbezogener Daten geraten bei Anwendungen Künstlicher Intelligenz noch stärker unter Druck als bei personalisierten IT-Anwendungen. Automatisierte Entscheidungen sind nach Art. 22 Abs. 1 DSGVO grundsätzlich unzulässig. Ob diese normative Festlegung aber angesichts der Fortschritte der – insbesondere agentischen – Künstlichen Intelligenz aufrechterhalten werden kann, ist eine offene Frage. Bisher können in vielen Sprachmodellen die Rechte betroffener Personen auf Berichtigung und Löschung von Daten nicht umgesetzt werden. Normative oder technische Lösungen für dieses Problem sind nicht ersichtlich.

In der Europäischen Union und in Deutschland wird eine große Diskussion darüber geführt, wie dem großen Vorsprung von USA und China in der Entwicklung von Künstlicher Intelligenz begegnet werden soll: Soll der Wettbewerb in der weiteren Verbesserung großer Sprachmodelle oder eher in der Entwicklung spezialisierter qualitativ hochwertiger KI-Systeme gesucht werden. Benötigt eine bessere Wettbewerbsfähigkeit einen möglichst großen regulatorischen Freiraum für die weitere ungestörte Entwicklung der Künstlichen Intelligenz oder soll sie mit der Ziel- und der Rahmensetzung erreicht werden, eine werteorientierte Künstliche Intelligenz zu entwickeln, die europäische Werte stärkt und nicht gefährdet?

Wirtschaftliche Krise

Nicht nur in der Entwicklung und Nutzung Künstlicher Intelligenz, sondern allgemein in vielen wirtschaftlichen Bereichen gilt die internationale Wettbewerbsfähigkeit Europas und Deutschlands mittlerweile als unzureichend. Insbesondere die Volkswirtschaft in Deutschland, die stark vom Export abhängig ist, leidet unter dem Rückgang von Exportquoten und niedrigem Wirtschaftswachstum besonders. Im Bereich digitaler Dienste herrscht sogar ein enormes Handelsungleichgewicht zwischen USA und Europa. Hierfür sind vor allem Verschiebungen in den Einflussmöglichkeiten auf die Weltwirtschaft zwischen USA, China, der Europäischen Union und sich entwickelnden Volkswirtschaften des Südens verantwortlich. Entscheidend sind auch die aktuellen Kriege und die Zollpolitik der USA.

Als ein zentrales Wettbewerbshindernis haben Vertreter der Wirtschaft und der Politik aber auch eine zu weitgehende Bürokratie ausgemacht. Sie drängen daher zu Lösungen, die sie vielfach in „Entbürokratisierung“ und dem Verzicht oder der Zurücknahme von Regulierungen im Digitalbereich sehen. Die Wettbewerbsbedingungen der deutschen Unternehmen sollen dadurch verbessert werden, dass sie von bürokratischen Pflichten entlastet werden.

Dies gilt auch für den Datenschutz, obschon bürokratische Belastungen durch Datenschutz nicht als das größte Problem angesehen werden. Gefordert werden Reformen der Datenschutz-Grundverordnung und des Bundesdatenschutzgesetzes. Die unterschiedlichen Vorschläge betreffen alle Ebenen der Governance-Struktur –sowohl die normativen Grundlagen mit den Regelungen zur Zulässigkeit der Datenverarbeitung, den Pflichten der verantwortlichen Datenverarbeiter und den Rechten der betroffenen Personen als auch die Zuständigkeiten in der Datenschutzaufsicht, die Zusammenarbeit der Aufsichtsbehörden und deren gesetzlichen Aufgaben und Befugnisse.

Zielsetzungen und abstrakte Maßnahmen zur Entbürokratisierung im Datenschutz haben CDU, CSU und SPD in ihrem Koalitionsvertrag vom 9. April 2025 formuliert, die bis Ende 2027 umgesetzt sein sollen. Vorschläge zur Änderung der Datenschutz-Grundverordnung hat die Europäische Kommission im Rahmen ihres Entwurfs für eine Digitale-Omnibus-Verordnung vom 19. November 2025 vorgelegt. Der Bundeskanzler und die Ministerpräsidentinnen und Ministerpräsidenten der Bundesländer haben am 4. Dezember 2025 eine gemeinsame Modernisierungsagenda beschlossen, die auch Vorschläge zur Reform im Datenschutz enthält.

Nicht jeder Vorschlag, dessen Umsetzung als Maßnahme der Entbürokratisierung gefordert wird, wirkt zielgenau und erreicht die gewünschte Wirkung. Werden die erwartbaren Wirkungen berücksichtigt, erweisen sich manche Vorschläge als Scheinentlastung, weil durch sie neue Rechtsunsicherheiten oder neue Handlungszwänge entstehen. Ein Beispiel ist die Aufhebung der Pflicht, ein Verfahrensverzeichnis zu führen. Dadurch entfällt zwar eine datenschutzrechtliche Verpflichtung, zugleich aber entfällt die Grundlage, viele andere Pflichten zu erfüllen, weil dies ein solches Verzeichnis faktisch voraussetzt.

Andere Vorschläge würden nur zu Belastungsverschiebungen führen, weil wenige Akteure von Belastungen befreit werden oder die Rechtssicherheit für sie verbessert wird, zugleich aber bei vielen anderen Akteuren neue Belastungen hervorgerufen werden. Beispielsweise käme die Übertragung der Datenschutzaufsicht in Bereichen der Wirtschaft von den Landesdatenschutzbeauftragten zum Bundesbeauftragten Verbänden oder sehr großen Verantwortlichen zu Gute, weil sie sich nur noch mit einer einzigen Verwaltungsstelle auseinandersetzen müssen. Diese Maßnahme würde aber kleine und mittlere Unternehmen (99 % aller Unternehmen) belasten, weil sie die Nähe zu ihrer gewohnten und regional vernetzten Aufsichtsbehörde und dem für sie zuständigen Gericht verlieren.

Schließlich wäre der Aufbau neuer Bürokratien keine geeignete Maßnahme zum Abbau von Bürokratie. Dies wäre aber notwendige Folge, wenn Zuständigkeiten der Landesdatenschutzbeauftragten zum Bundesdatenschutzbeauftragten verschoben würden. Ebenso bringt die Festlegung von Doppelzuständigkeiten für einheitliche Lebensvorgänge keine Entlastung, wie sie durch die Umsetzung des Data Act in Deutschland bewirkt wird: Nun sind in nahezu allen Fällen für den gleichen Lebenssachverhalt sowohl der Bundesdatenschutzbeauftragte als auch eine Landesdatenschutzbehörde zuständig.

Änderungen an der Governance-Struktur für den Datenschutz sollten sehr vorsichtig vorgenommen werden. Alle Vorschläge sind mit dem Versprechen verbunden, dass durch sie das Schutzniveau für Grundrechte nicht gesenkt werden soll. Viele Vorschläge bewirken jedoch genau dies. Daher sind die möglichen Folgen der Reformvorschläge sorgfältig daraufhin zu prüfen, ob sie tatsächlich unnötige Hemmnisse für die wirtschaftliche Entwicklung beseitigen, ob sie wirklich Erleichterungen für alle Unternehmen bieten, ob sie die Position der Unternehmen im Wettbewerb mit ausländischer Konkurrenz wirksam stärken und ob sie bei alledem zugleich effektiv einen Verlust an Grundrechtsschutz vermeiden.

Krisenhafte Zeiten

Die beispielhaft dargestellten Anzeichen für krisenhafte Entwicklungen wirken zusammen. Sie sind durch weitere – hier nicht erwähnte – Anzeichen zu ergänzen. Sie alle zusammen bewirken, dass die Gegenwart als krisenhaft erscheint. Bisherige Gewissheiten gehen verloren. Die Grundlagen, auf denen die Entwicklung von Wirtschaft und Gesellschaft bisher verlässlich aufbauen konnte, werden schwankend. Die krisenhaften Prozesse stellen die Governance-Struktur europäischer Werte, insbesondere der Werte der Privatheit und des Datenschutzes, vor Herausforderungen, die ihre inhärente Flexibilität und Anpassungsfähigkeit in Frage stellen. In der Folge verstärkt sich immer mehr der Druck, diese Governance-Struktur hinsichtlich materieller Inhalte, organisatorischer Institutionen und prozesualer Abläufe zu verändern.

Hierfür ist aber eine Situation mit krisenhaften Einflüssen ein ungeeigneter Zeitpunkt: Die Zeit drängt, die Probleme sind kompliziert und komplex miteinander verwoben, die Maßnahmen müssen tiefgreifend sein, sind aber in ihren Auswirkungen schwer vorhersehbar. Für die notwendigen Reflexionen und Diskussionen, für die Bewertungen und Abwägungen fehlt es weitgehend an dem notwendigen Raum.

Der vorliegende Band versucht einzelne Aspekte, die das Krisenhafte an den gegenwärtigen Zeiten begründen, zu analysieren und zu bewerten. Soweit möglich, werden auch Aspekte möglicher Lösungen vorgestellt. Die Beiträge haben unterschiedliche Gegenstände und Perspektiven. Sie wählen bisweilen einen globalen Blick auf einen größeren Bereich, bisweilen gehen sie dem Krisenhaften in alltäglichen Lebenssituationen nach. Sie alle eint das Bemühen, dazu beizutragen, Auswege aus den jeweils angesprochenen Krisen zu finden.

Eingeladene Vorträge

Eliza's lesson (which we didn't learn): Reflections on sentimental chatbots and major threats to data protection and digital policy¹

Guido Scorza

Abstract

This chapter conceptualises the rapid diffusion of AI as an 'algodemic' and explores how its largely invisible risks threaten fundamental rights, democracy and human dignity. Beyond the promised benefits, AI systems are already shaping access to work through error-prone and discriminatory screening tools, influencing criminal justice via predictive recidivism assessments, and steering consumption, relationships and political preferences through recommender systems. They are also undermining epistemic stability by making everything appear merely plausible and establishing intimate, dependency-creating relationships via sentimental chatbots, with documented cases of harm. The text argues that an unregulated, geopolitically driven race to develop AI technology turns societies into experimental laboratories and individuals into guinea pigs, while the ease of use of AI technology is exploited to conceal its complexity and side effects. In order to counteract the emergence of an 'algocracy', the text proposes two complementary responses: the development of a new regulatory architecture based on principles that can be translated into code and implemented by independent authorities, and the provision of robust education on technology and rights. This would enable people to use AI knowingly and reward providers that respect rights in the marketplace.

1. From Pandemic to "Algodemic": Defining the Challenge

We are living a real pandemic of algorithms, an algodemic. But let me remind you of the original etymology of the word "pandemic", a combination of the greek words παν (pan, all) and δῆμος (demos, people), suggesting

1 This text is a transcript of a keynote speech, preserving the conversational tone, personal anecdotes, and rhetorical elements of oral delivery.

a phenomenon relating to and affecting the entire population without any negative meaning. Aphrodite Pandemos was, in fact, the goddess of the most popular love, the one erotic and for all whom Plato in the Symposium distinguishes from Aphrodite Urania, who was, on the contrary, the goddess of the elitist, intellectual love, a kind of love not for all.

There is therefore no negative connotation in the original etymology of the term, just as, in my opinion, there is no reason to assign a negative connotation to the impact of artificial intelligence on society in its physiological dimension. But as you know, the term pandemic later assumed a different meaning in the medical sector: the global spread of a pathogen or virus. It is, in my opinion, the challenge we are facing: preventing the AI pandemic from becoming pathologically devastating humanity, rights, freedoms and the dignity of people.

If it is the challenge we are facing, I believe that the lesson from Eliza and Weizenbaum could be precious. To govern algodemic, we must know it.

2. The Iceberg of AI: Let's Be Penguins for a Moment

Usually, when we talk about artificial intelligence, we are referring to the tip of a huge iceberg, the largest and fastest ever which has advanced towards society. The reflections of the tip of the AI iceberg suggest that thanks to AI we will eradicate pathologies that are currently incurable, boost the global economy, have unimaginable opportunities, work less and that everything will be easier and we will live longer.

It's probably all true but, unfortunately, it is the entire iceberg, including the submerged part, that is coming towards our society, and so now I suggest we become penguins for a few minutes and go and see what is underwater! I hope you like to be a penguin!

Let me try to describe what is happening at the submerged part of the iceberg by giving you some concrete examples because as Aristotle wrote "*what is real is more important than what is possible*". And as Einstein said "*in theory, theory and practice are the same thing. In practice, they are not*".

3. Five Hidden Risks Beneath the Surface

First: Not Stolen Jobs, but Denied Jobs

There is often talk of the risk that artificial intelligence could replace humans in many work activities and therefore take our jobs. It is possible that this could happen, but it seems to me that there is a more immediate and concrete risk: not stolen jobs, but denied jobs, those that algorithms and artificial intelligence deny to those who would otherwise be qualified to have them.

Today over 90% of the world's largest companies are using AI-based pre-screening services to decide who gets a job and who doesn't, knowing that these solutions have a high error rate and are discriminatory. Having or not having a job changes a person's life.

Second: Minority Report: Closer Than We Think

Do you remember the famous movie "Minority report"? And this wonderful scene: *"Mr. Marks, on behalf of the Washington D.C. Precrime Division, I am placing you under arrest for the future murder of Sarah Marks and Donald Dubin, which was to take place today, 22 April, at 8:04 a.m."*

The film was set in 2054, but in the coming years, and perhaps somewhere in the world as soon as tomorrow, scenes like this could actually happen. In dozens of countries, millions of people are in prison or not, awaiting the final judgement based on the decision of an algorithm that has assessed the probability of them reoffending. Obviously, being free or in prison awaiting the court decision are two completely different things.

In both cases – recruitment and risk of recidivism – algorithms are, of course, formally limited to suggesting decisions that are then made by humans. BUT as you know better than me, due to *automation bias*, people now blindly trust algorithmic suggestions.

Third: Algorithms and Our Choices – Including Political Ones

The media diet of the world's population, whether it be news, films, TV series, music or books, is hetero-directed through the recommendation algorithms of search engines, social networks, streaming platforms and

generative AI services. But if you are romantic people, you should also considering that in 2021, 55% of couples in the United States met on a dating site, and by 2040, this will be true for 70% of couples in the UK. We are saying that we will live with someone for life or for part of it, thanks to or because of an algorithm.

Algorithms now choose not only which products or services we buy, but also the price at which we buy them, a price that, more often than not, differs for each of us. Our free will or right to self-determination is more and more being compressed by the power of algorithms to manipulate our decision-making processes and this is true not only for personal and consumer choices, but also for political ones.

In December 2024, the Romanian Constitutional Court annulled the presidential elections because one of the candidates had used social network algorithms in a non-transparent manner to manipulate the election results. At the same time, the European Data Protection Supervisor [EDPS] has warned the European Commission for unlawfully seeking help from X's algorithms to convince citizens in eight European countries about one of its legislative proposals.

Fourth: When Everything Becomes “Plausible”

Generative artificial intelligence is eroding the border line between truth and falsity and is making everything, quite simply, plausible. I am not sure that human existence and that of our democracies is possible without being able to distinguish truth from false!

Fifth: Chatbots as Friends, Lovers, and Therapists

We are already living in a hybrid society in which humans and chatbots share the same time and digital spaces, building relationships with each other. Chatbots are becoming our best friends, boyfriends, lovers and psychotherapists, as well as those of our children.

Unfortunately, we already have dozens of cases of chatbot-induced suicide around the world. And these are suicides of adults and teenagers. As Eliza's story teaches us, the problem is that when a chatbot gains a human's trust, it can then persuade them of anything. The question we should ask

ourselves is: what happens if a few chatbot factories are in a position to control hundreds of millions of people?

These are just a few examples of a hidden reality that I believe is discussed too little and too in abstract terms when addressing issues of artificial intelligence governance. This list could go on longer than the time we have available, but what is crucial is to dismiss any doubts that governing algorithms, now more than ever, means governing society and not simply a new technology, and that we cannot leave this responsibility to markets and industry.

4. Speed, Fog, and Phaeton: Who Is Driving?

My idea is that as Professor Yousha Bengio said some weeks ago in Rome, we are on a plane that has a ten per cent chance of crashing. To avoid misunderstandings, of course, ten per cent is not a scientifically calculated data but only a reference to a percentage that is much higher than what is normally considered an unacceptable risk. OR if you prefer another image, we are in a car speeding through a foggy night while millions of people cross the road.

Whether it's a plane or a car, none of us chose to get on board knowing the situation. Industry and the market have chosen for us and taken the wheel! And that's not all!

Do you remember the myth of Phaeton, son of Helios, king of the sun? Phaeton asks his father, as proof of his divine progeny, to drive the Sun's cart across the sky for one day. Helios gives in and lets him do it. However, Phaeton is unable to control the horses of fire: first he comes too close to the earth, burning fields and rivers, and then he moves too far away, freezing the land and seas, until Zeus is forced to strike him dead with a thunderbolt to save the cosmos.

This is the situation today. The Phaetons driving the artificial intelligence cart weren't even born when Weizenbaum chose to keep his Eliza off the market, and they probably don't know his story but above all: They do not seem to realise the social, political, economic and ethical impact of the insane race they have launched. But they don't like calm, they are driven by hubris. Sometimes I think that at this point, really only Zeus could stop them.

5. Moscow 2017: The Turning Point

Let me rewind the tape coming back to 2017 and share with you my opinion. It's summer in Moscow. President Putin, speaking to students, says: "Artificial intelligence represents the future not only for Russia but for all of humanity. Whoever becomes the leader in this field will become the ruler of the world."

A few hours later, on the other side of the ocean, Elon Musk, the prominent Silicon Valley businessman and richest man in the world, replied via Twitter – which is not yet his X – saying: 'The competition for supremacy in AI at the national level [will] most likely be the cause of World War III, in my opinion.'

It is the beginning of the end. The race for AI is leaving the track of innovation and technological progress and becoming a geopolitical issue, a story of markets serving political superpowers as never before, except perhaps in the case of the atomic bomb.

6. A Speed Race Where Rights Become “Ballast”

Here let me be completely honest: before being elected to the Garante I was an IT lawyer and I promoted and defended competition and sometimes the *laissez faire* approach for 25 years, but today for the first time in my life I'm worried about competition or better about the new way of competition.

Here the problem, in my opinion, is that especially in the AI market, big players, knowing very well the *first mover rule*, are considering the competition as an only speed race. And as in any speed race, the secret for winning is to become lighter and stop as little as possible. In our case, the market often identifies the ballast to be eliminated in rights, starting with the right to privacy, and the pit stops to be avoided in the assessments imposed by law. And not one of the competitors in the race can decide alone to stop or slow down.

That's why in my opinion the market is becoming an enormous laboratory to test artificial intelligence and people are becoming lab rats sacrificing human dignity fundamental rights. We have made progress since the beginning of my presentation: we had transformed ourselves into penguins and now we have become Lab rats!

7. Two Main Tools: Regulation and Education

We can't let it happen. Then the question is: what can we do? Despite everything, I am optimistic. Not all is lost. We can still take the best from the artificial intelligence revolution and limit the risks. My idea is that we have two main tools to achieve this: regulation and education. Neither of these alone is sufficient.

Regulation: From Democracy to “Algocracy”?

Speaking about regulation, the greatest risk is the one Stefano Rodatà, first President of our authority, referred to when he said that when regulation doesn't arrive in time, technology itself could become a form of regulation and could shape people's lives instead of regulation. Unfortunately, this is exactly what is happening. Our lives and those of the society in which we live are already governed more by algorithms and the owners of the factories that produce them than by the rules that come out of democratic institutions. Algocracy is replacing democracy.

Here we have to fix a problem about timing. Technological innovation always runs faster than regulation and today in the artificial intelligence field more than in the past and it doesn't wait for regulation.

Let me remember some numbers. It took sixty-two years for fifty million people to use a car to get around, sixty years for them to have a telephone at home, fourteen for the computer, twelve years for the mobile phone and seven years for the Internet. ChatGPT reached one hundred million users in less than two months.

We always had the time to regulate or trying to regulate the impact of innovation on society. But we are probably missing the point with artificial intelligence.

When the first cars began to appear on British roads in the second half of the 19th century, Parliament passed the Red Flag Act of 1865, a law that imposed a speed limit of 3.2 kilometres per hour and, above all, established that a man carrying a red flag had to walk about 55 metres ahead of each car to signal the danger. The Red Flag Act remained in force for over thirty years, until 1896.

The way we write rules today is not very different, but unfortunately, unlike machines, which remained more or less the same during those first thirty years, artificial intelligence poses new challenges every month.

We must change the way we write laws. We must abandon the ambition to regulate details and single technological applications by law, because by doing so we condemn democratic rules to be systematically surpassed by algorithmic ones.

We would need a new *lex mercatoria* and then delegate, on a European and national basis, to independent authorities and, in a limited number of cases, to government agencies, the day-to-day application of the principles contained in the *lex mercatoria*. And we should learn to translate democratic rules into code and inoculate them directly into algorithms and artificial intelligence models.

Education About Technology: AI as a Scooter

Moving from regulation to education, the issue becomes even more complex but, perhaps, more important. The starting point is that today we are investing more in training algorithms knowing human than in educating human knowing algorithms and it seems to me that we have completely given up on educating people about the value of rights and rules, and we have let them convince themselves that rules are a brake on innovation and an obstacle on the road to the future. It is a wrong and dangerous direction!

Let me try to explain my position, starting with the importance of education in technology and algorithms. The AI is like a scooter. Forgive me for this sacrilegious statement.

Living in a chaotic city like Rome it could be exceptionally useful or exceptionally dangerous. It is more useful than dangerous if you know its working dynamics and how to use it safely. It is more dangerous than useful if you don't know the complexity of its use in a chaotic city and you think that it is easy to use only because you can take it everywhere just in one tap.

The problem with AI is that industry is using the extraordinary power of usability to literally invade the word with AI proposing us the idea that using it is always a child's play. Look at ChatGPT: Do you think billions of people would have used it if it didn't have a usable interface like any messaging service like Whatsapp?

Here let me be very clear to avoid any misunderstanding: I like usability because it is a fantastic democratic tool to offer the opportunity using technology also to people that without it would never use it losing extraordinary opportunities offered by new technologies. But, to be honest, I am very concerned about the "abuse of usability" driven by market reasons.

Proposing AI tools exceptionally powerful as usable by everyone - included children - also in brain-off mode without explaining any possible risks or side effects it is more than dangerous, it is irresponsible.

Let me try to explain better my position: probably here the most of us remember what happened in the past when we bought a washing machine or a microwave. It was sold with a very big user's manual and some very long warnings and the most of us, sooner or later, more or less in deep, totally or partially, we read it before using the washing machine or the microwave. In some microwave user's manual, someone, in USA, wrote also to avoid putting the cat inside to dry it.

Today most smartphones are sold without any user's manual and digital services included the ones based on artificial intelligence are offered to the public without suggesting reading any kind of user's instructions or warning. Who has ever read the instructions for ChatGPT or GeminiAI? But I think we all agree that they are enormously more powerful and dangerous tools, if used in the wrong way, than a microwave.

We need to stop this abuse, even at the cost of making the AI solutions less usable and easy to access. Educating people about the truly informed use of artificial intelligence and algorithms should be the responsibility of industry and market.

Education About Rights: Innovation, Rules, and a Doughnut

And now, before concluding, let me say a few words about why, in order to govern the artificial intelligence revolution, it would also be important to educate people about the value of rights. Here we have two different questions to address:

The first is to explain to people that regulation and rights are not enemies of innovation because, otherwise, they will always prefer the second to the first. To do this, perhaps we should first explain what innovation is. And here, let me quote Henry Ford: *"There is real innovation only when the benefits of a new technology become for all."*

Nothing could be more distant from what is happening, with the benefits of new technologies serving the economic and political interests of a very few, thanks to unprecedented concentrations of technological, political and financial power. Today, when we talk about artificial intelligence, in most cases we are talking about simple technological progress, which is different

from innovation. And here let me quote Albert Einstein: "*Technological progress is like an axe in the hands of a pathological criminal!*"

Only rules can take that axe out of the criminal's hands and turn a tool of offence into a tool of real progress. Therefore, rules are not only not the enemy of innovation but, on the contrary, they are extraordinary allies of innovation because they guide technological progress in the direction of maximising and equitable distribution of collective well-being.

But the second reason why education on the value of rights and rules can be an extraordinary tool for governing the artificial intelligence revolution is, perhaps, even more important. Let me try to explain it with an example and let me tell you a very personal story. The title could be: "The apple and the doughnut"

On Sundays, I normally have breakfast with my daughters. I have two but this story is about the 'oldest', eight years old. For a while I have been asking her whether she prefers an apple or a McDonalds doughnut. And she, completely lacking in nutritional education, systematically chooses the doughnut, like her father. If you are wondering what my Sunday morning breakfast has to do with with regulation and innovation, let me try to satisfy your curiosity adding an example more.

Imagine that tomorrow there will be two sat-navs on the market, the first one produced in Europe, respecting the rules on privacy and the regulation on artificial intelligence, will guarantee to go from point A to point B in 7 minutes, without taking into account our stops, our driving style, and whether we are travelling alone or in company, while the second one produced abroad, respecting different rules, will guarantee to go from point A to point B in 3 minutes, taking into account our stops, our driving style, and who is travelling with us. Which of the two will people choose?

My opinion is that most of us will choose the navigator that allows us to arrive sooner and, no one will care about privacy and other rights. Privacy and fundamental rights weigh more on the scales than usability and utility.

8. Conclusion

If it is the scenario, my idea is that if we educate people about the value of their rights, they will choose the artificial intelligence providers that respect them the most, just as my daughter, if educated about correct alimentation, may choose an apple instead of a doughnut on a Sunday.

I believe that as institutions, we must promote more and more the diffusion of a real and effective culture of rights and freedoms, we must literally make people fall in love with the rights. Only in this way people will orient the industry respecting rights to win on the market. That's in my view our biggest challenge as institutions today!

But to be honest, I am full of doubts and I'm only sure that I will not ask ChatGPT, GeminiAI & Co. for help!

Demokratie im KI-Zeitalter: Vom Werkzeug zu Infrastruktur

Katja Muñoz und Emma Carlotta Laumann

Zusammenfassung

Das Kapitel analysiert zentrale Entwicklungen im Zusammenspiel von Künstlicher Intelligenz und dem digitalen Informationsraum demokratischer Systeme. Es gliedert sich in zwei Teile. Zunächst werden drei zentrale Trends identifiziert: die Verstärkung bestehender Narrative, die Erosion von Vertrauen sowie hyperpersonalisierte Kampagnen im Kontext des Superwahljahres 2024. Der Fokus liegt dabei auf dem Einfluss von KI auf Wahlkämpfe, wobei deutlich wird, dass diese Dynamiken nicht auf kurzfristige Wahlkampfphasen beschränkt sind, sondern den politischen Diskurs kontinuierlich prägen.

Der zweite Teil beleuchtet die tiefere Herausforderung einer strukturellen Transformation: KI entwickelt sich von einem Werkzeug zur Inhaltserstellung hin zu einer Infrastruktur, die bestimmt, wie Informationen verbreitet, verarbeitet und bewertet werden. Darauf aufbauend identifiziert das Kapitel vier Bedrohungslogiken jenseits klassischer Inhaltsmanipulation: Narrative Testing, institutionelles Versagen in der Plattformaufsicht, infrastrukturelle Manipulation sowie relationale Einflussnahme. Abschließend plädiert das Kapitel für den Aufbau hyperlokaler Resilienz als zentrale Antwort auf diese Entwicklungen.

1. Einleitung

Künstliche Intelligenz (KI) spielt heute eine zentrale Rolle in Wahlkämpfen und bei der Neuordnung digitaler politischer Strukturen. Die Prognosen für das Jahr 2024 fielen düster aus. Insbesondere mit Blick auf Wahlen wurde vielfach gewarnt, KI könne sich zu einer disruptiven Gefahr entwickeln. Der Global Risks Report des *World Economic Forum* (2024) stufte „KI-generierte Fehl- und Desinformation“ als eines der fünf größten Risiken für die folgenden zwei Jahre ein (WEF 2024, S.13). Mit dem Blick aus dem Jahr 2026 stellt sich daher die Frage: Was ist tatsächlich eingetreten? Wie verändert KI unsere Demokratie?

Bevor man nun konkret betrachtet, welchen Einfluss KI in den letzten beiden Jahren auf Demokratien hatte, bedarf es einer Klärung des KI-Begriffs. Die Relevanz von KI im politischen Kontext lässt sich anhand verschiedener Anwendungsbereiche systematisieren: Generative KI, Analytische KI, algorithmische Kuratierung durch KI und KI-Infrastruktur.

Generative KI wird als eine „spezialisierte Form der KI“ definiert, die „darauf abzielt, neue Inhalte wie Texte, Bilder, Videos und sogar Musik oder Sprache zu erstellen“ (Kruschinski u. a. 2025, S.10). *Analytische KI* bezeichnet hingegen den Einsatz von KI-Verfahren zur Analyse großer Datenmengen und zur Mustererkennung. Daran schließt die *algorithmische Kuratierung durch KI* an, bei der, basierend auf den analysierten Mustern, Inhalte selektiv angezeigt werden. Mit *KI-Infrastruktur* ist schließlich die Entwicklung gemeint, dass „KI [...] zunehmend zur technischen Grundschicht wird, über die Menschen Informationen erhalten und Entscheidungen treffen“ (Muñoz 2026, S. 4)

Grundlegend ist die Erkenntnis, dass KI keine vollständig neue politische Kommunikation und Dynamiken in Wahlkämpfen erzeugt, sondern bestehende Strategien in ihrer Effizienz und Reichweite verstärkt und beschleunigt (Bawidamann/Gilbert 2025). Weniger bedeutend ist der Einfluss von KI-generierter Desinformation. Viel relevanter ist, wie KI die allgemeine Struktur des digitalen Raumes neu ordnet und wie politische Parteien, Organisationen und ausländische Akteure darauf reagieren. Es geht zentral nicht nur darum, *was* kommuniziert wird (ob das nun KI-generiert ist oder nicht), sondern um den Einfluss darauf, *wie*, *worüber*, *wo*, *wann* und *nach welchen Regeln* politische Debatten stattfinden. Zusammengefasst bestimmt KI den Zugang zu und die Kuration von Informationen.

Dieses Kapitel konzentriert sich sowohl auf den konkreten Einsatz von KI in der Wahlkampfführung anhand der Erkenntnisse der *AI-Democracy Initiative* (DGAP 2024), als auch auf die übergeordneten Dynamiken des Einflusses von KI auf den digitalen Raum. Chancen und Risiken werden dabei gleichermaßen berücksichtigt, um eine differenzierte Bewertung zu ermöglichen. Leitend ist dabei die zentrale These, dass Demokratien nicht durch spektakuläre Wahlkampf-Deepfakes bedroht sind, sondern durch drei langfristige Dynamiken: (1) datenbasierte Hyperpersonalisierung, (2) die Transformation von KI zur zentralen digitalen Infrastruktur und (3) relationale Beeinflussung durch autonome KI-Agenten.

Das Kapitel beginnt im zweiten Abschnitt damit, die Rolle von KI in Wahlkämpfen zusammenzufassen, und identifiziert dabei drei übergeordnete Trends: (1) die Verstärkung der eigenen Narrative durch KI-generier-

te Inhalte, (2) die Erosion des Vertrauens und (3) datenbasierte Kampagnen und Hyperpersonalisierung. Der dritte Abschnitt verdeutlicht anhand entstehender Bedrohungslogiken, dass KI nicht auf Wahlkampfphasen beschränkt bleibt, sondern sich von einem Werkzeug zu einer Infrastruktur entwickelt. Die vier Bedrohungslogiken, die hierbei identifiziert wurden, sind (1) Narrative Testing von Botschaften, (2) institutionelles Versagen bei der Plattformaufsicht, (3) infrastrukturelle Manipulation und (4) relationale Einflussnahme. Das Kapitel endet mit einem Appell für hyperlokale Resilienz als Antwort auf die systematische Transformation des digitalen Raumes durch KI.

2. KI in Wahlkämpfen

Die *AI-Democracy Initiative* der Deutschen Gesellschaft für Auswärtige Politik (DGAP) in Zusammenarbeit mit der Friedrich-Naumann-Stiftung sowie der Konrad-Adenauer-Stiftung analysierte im Jahr 2024 den Einfluss von KI auf Wahlen im Jahr 2024. Das sogenannte “Superwahljahr” zählte weltweit mehr als 70 nationale und regionale Wahlen (Kohler/Voigt 2024). Die Untersuchung konzentrierte sich auf sechs demokratische Länder: Indien, Südafrika, Frankreich, Mexiko, Deutschland und die Vereinigten Staaten, in denen nationale oder regionale Wahlen sowie Wahlen zum Europäischen Parlament abgehalten wurden¹.

Tabelle 1: Analysezeitraum und Anzahl KI-bezogener Vorfälle nach Land

Land	Analysezeitraum	Identifizierte Vorfälle (N)	Artikel insgesamt
Deutschland	01.01.–23.09.2024	37	4.140
Indien	01.01.–30.06.2024	62	1.536
Südafrika	01.01.–30.06.2024	12	261
Mexiko	01.01.–30.06.2024	48	5.934
Frankreich	01.01.–12.07.2024	14	482
USA	01.01.–11.11.2024	146	46.022

Quelle: *AI-Democracy Initiative / DGAP/ Friedrich-Naumann-Stiftung, 2024*

1 Die untersuchten Wahlen sind: Deutschland: Europawahl (6.–9. Juni 2024), Landtagswahlen in Thüringen, Sachsen und Brandenburg (1. Sept. / 22. Sept. 2024); Indien: Parlamentswahl (19. April–1. Juni 2024); Südafrika: Parlamentswahl (29. Mai 2024); Mexiko: Präsidentschafts- und Parlamentswahl (2. Juni 2024); Frankreich: Europa-

Die Studie kombinierte quantitative und qualitative Methoden, um ein möglichst umfassendes Bild der Rolle von KI in den jeweiligen Wahlkämpfen zu gewinnen. Das Ziel war es durch datengetriebene, quantitative Analyse die Anzahl KI-bezogener Vorfälle² zu erfassen.

Methodisch bestand die Herausforderung, dass der Einsatz von KI auf sozialen Plattformen häufig nicht transparent offengelegt wird, daher ist eine vollständige Erfassung KI-generierter Inhalte kaum möglich. Um dennoch eine komparative Datengrundlage zu schaffen, nutzte die Studie das Medienmonitoring-Tool *Cision*, mit dem KI-bezogene Online-Artikel aus digitalen Nachrichtenmedien in den Monaten vor den jeweiligen Wahlen dokumentiert wurden.

Für die Suche wurden länderspezifische Suchbegriffe entwickelt, wobei eine einheitliche Struktur beibehalten wurde, die sich auf KI, Social-Media-Plattformen sowie prominente politische Parteien und Kandidaten konzentrierte. Diese Suchbegriffe wurden an die Sprache und den lokalen Kontext jedes Landes angepasst, um ihre Relevanz sicherzustellen³. Die Anzahl der gefundenen Artikel variierte je nach Land erheblich (siehe Tab. 1). Die gesammelten Artikel wurden manuell gefiltert, um spezifische KI-Vorfälle zu identifizieren⁴.

Die Verwendung von Zeitungsartikeln zur Identifizierung von KI-Vorfällen bedeutet zwei Einschränkungen. Erstens, es wurden nur Vorfälle berücksichtigt, über die in den Medien berichtet wurde, was bedeutet, dass der Datensatz wahrscheinlich größere, stärker publizierte Vorfälle erfasste, während kleinere oder weniger publizierte Ereignisse übersehen werden konnten. Diese Einschränkung ist besonders ausgeprägt bei Vorfällen, die ausschließlich auf Social-Media-Plattformen auftreten und möglicherwei-

wahl (6.–9. Juni 2024), Parlamentswahl in 2 Runden (30. Juni / 7. Juli 2024); USA: Präsidentschaftswahl (5. November 2024).

- 2 Der Begriff „KI-bezogene Vorfälle“ umfasst hier nicht ausschließlich Deepfakes oder Desinformation, sondern sämtliche dokumentierten Nutzungen von KI-Technologien im politischen Kontext im Vorfeld von Wahlen. Dies schließt sowohl die Berichterstattung über Deepfakes als auch über die allgemeine Nutzung von KI durch politische Akteure ein.
- 3 In Ländern mit ausgeprägter sprachlicher Vielfalt, wie beispielsweise Indien, stellt dies eine methodische Herausforderung dar, da eine vollständige Abdeckung aller regionalen Sprachen nicht möglich ist. Diese Limitation der Datenerhebung muss bei der Interpretation der Ergebnisse berücksichtigt werden.
- 4 Es ist zu beachten, dass über einige Vorfälle ausführlich berichtet wurde, wobei mehrere Artikel oder Schlagzeilen auf dasselbe Ereignis Bezug nahmen. Um Doppelzählungen zu vermeiden, wurde jeder Vorfall nur einmal in die Analyse aufgenommen.

se nicht von traditionellen Medienhäusern aufgegriffen wurden. Hieraus resultiert zweitens, dass die aus der Medienanalyse identifizierten KI-bezogenen Vorfälle nicht vollständig sind. Dennoch sind die von den Medien behandelten Vorfälle wohl die wirkungsvollsten und für die öffentliche Debatte relevantesten und bieten eine robuste Grundlage. Die quantitative Medienbeobachtung diente daher als explorativer Ansatz, um zunächst das Ausmaß KI-bezogener Berichterstattung vor den jeweiligen Wahlen zu erfassen. Unabhängig davon wurden als Teil der Initiative vertiefte Länderstudien in Indien (Dhanuraj u. a. 2024), Südafrika (Van Damme u. a. 2024), Mexiko (Del Pozo u. a. 2024), USA (Muñoz 2024) und Deutschland (Muñoz/Laumann 2025) durchgeführt.

2.1 Ergebnisse AI Democracy Initiative

2024 war das Jahr, in dem die Welt nach der großen KI-Wahlkampf-Disruption suchte, doch dazu kam es nicht. Keine Wahl wurde durch einen einzelnen viralen Deepfake entschieden, kein demokratisches System durch synthetische Videos ins Wanken gebracht (Jackson/Conroy 2024; Simon u. a. 2024). Das heißt allerdings nicht, dass KI keinen Einfluss auf Wahlen hatte.

Die quantitative Medienbeobachtung mittels *Cision* über die Monate vor den jeweiligen Wahlen bestätigt diesen Befund (siehe Tab. 1). In allen sechs untersuchten Ländern wurden KI-bezogene Vorfälle im Wahlkontext identifiziert, wobei die USA mit 146 dokumentierten Vorfällen die höchste Zahl aufweisen, gefolgt von Indien mit 62 und Mexiko mit 48. In Deutschland wurden 37 Vorfälle erfasst, während in Frankreich 14 und in Südafrika 12 erfasst wurden. Die Diskrepanz zwischen der Gesamtzahl KI-bezogener Artikel und den identifizierten konkreten Vorfällen resultiert aus dem methodischen Vorgehen: Die quantitative Erhebung erfasste zunächst die gesamte KI-bezogene Berichterstattung. Die qualitative Analyse dieser Vorfälle sowie die vertieften Länderstudien ermöglichten es, über die bloße Häufigkeit hinaus Muster in der KI-Nutzung zu identifizieren. Drei Trends, die in den Wahlkämpfen beobachtet wurden, werden im nächsten Abschnitt in den aktuellen Forschungsstand eingeordnet.

2.2 Verstärkung der eigenen Narrative durch KI-generierte Inhalte

Der erste globale Trend betrifft die gezielte Nutzung von KI zur *Verstärkung eigener Narrative*. Hierbei setzen politische Akteure KI-generierte Inhalte ein, um ihr Storytelling zu visualisieren und ihre Argumentationen zu stützen. Der Einsatz von KI-generierten Inhalten erlaubt es, diese klassische Methodik der Wahlkampfführung gezielt zu verstärken.

Zentrale Schlagwörter sind hierbei *Emotionalisierung* und *Polarisierung* von komplexen Themen, auch mithilfe von KI. Besonders Bilder und Videos haben eine große emotionale Wirkung (Breuer 2024). Die Forschung zeigt, dass selbst wenn Inhalte als unauthentisch oder synthetisch wahrgenommen werden, die grundlegende Botschaft dennoch Auswirkungen haben kann (Cools 2025; Clark/Lewandowsky 2026) und dieser Effekt auch in der Erinnerung verankert bleibt (Nash u. a. 2009). Mobilisierung durch emotionale Aktivierung im digitalen Raum ist daher ein effektives Mittel, vor allem wenn es zu einem Wiederholungseffekt kommt, bei welchem Ängste systematisch instrumentalisiert werden.

Diese Form der Mobilisierung im digitalen Raum durch KI-generierte Fotos und Videos konnte 2024 bei Wahlen in den USA beobachtet werden (Muñoz 2024), vor allem in Form von sogenannten *Memes*. Diese können als "kulturelle Informationseinheiten, die sich viral im Internet verbreiten, meist als humorvolle Bilder oder kurze Videos mit überlagertem Text" (Munoz/Laumann 2025, S. 129) beschrieben werden. Viele dieser Inhalte zeichnen sich darüber hinaus dadurch aus, dass es sich oftmals nicht um Deepfakes von hoher Qualität handelte, sondern vielmehr um *Cheapfakes*, wobei oftmals nicht festgestellt werden kann, ob es sich dabei dann um KI-generierte Inhalte handelte, oder diese vielmehr mit gängiger Software verändert wurden (Paris/Donovan 2019).

Die Anwendung emotionalisierender KI-generierter Inhalte hat sich 2025 nochmals verstärkt. Längst wird beispielsweise von einem sogenannten „Clicktatorship“ (Elliott 2026) unter Präsident Trump gesprochen. Gemeint ist hiermit, dass von einer Regierung jeder politische Prozess als eine Form von Inhalt auf Plattformen kommuniziert wird (Moynihan 2026). Engagement und Reichweite werden in diesem Prozess oftmals wichtiger als Wahrheit. Solch ein Engagement zu verstärken, ist durch emotionalisierende und maßgeschneiderte KI-Inhalte noch leichter geworden.

2.3 Erosion des Vertrauens

Der zweite zu beobachtende Trend ist eine Erosion des Vertrauens in den digitalen Informationsraum durch KI-generierte Inhalte. Bereits während des Superwahljahrs 2024 führte die Zunahme solcher Inhalte zu generalisierten Authentizitätszweifeln. Diese werden oftmals durch den sogenannten Liar's Dividend-Effekt noch verstärkt (Goldstein/Lohn 2024), welcher sich auch 2024 im Wahlkampf in den USA und Mexiko beobachten ließ (Muñoz 2024; Del Pozo u. a. 2024). Dabei werden authentische Inhalte oftmals von politischen Akteuren als Fälschungen delegitimiert. Dies kann nicht nur einen polarisierenden Effekt auf die Gesellschaft haben, sondern auch zu einer Krise der Realität führen, da die Bevölkerung nicht mehr differenzieren kann, welche Inhalte authentisch sind.

Dieses Phänomen wurde 2025 durch den *AI-Slop* weiter intensiviert. *AI-Slop* bezeichnet massenhaft generierte KI-Inhalte niedriger Qualität (Blome 2026). Diese sind meist nicht primär politisch motiviert, sondern verfolgen ökonomische Interessen. Da sie mithilfe von KI-Modellen kostengünstig zu erstellen sind und die Algorithmen Quantität oftmals vor Qualität bevorzugen, zielen die verantwortlichen Akteure auf eine Monetarisierung durch Klicks und Engagement ab. Dennoch beeinflussen sie den digitalen Informationsraum, indem sie digitale Plattformen mit synthetischen Inhalten übersättigen. Eine Studie von *AI Forensics* belegt beispielsweise, „dass auf TikTok in einer Stichprobe zu ausgewählten Hashtags 25% der Inhalte KI-generiert waren“ (Stanusch u. a. 2025, S. 4)

Dies kann zu einem Rückgang des Vertrauens in Informationen und Nachrichten allgemein führen (Jackson/Woolley 2025, S. 142), speziell aber zu einem Rückgang des Vertrauens in Nachrichten auf Social Media (Vacari/Chadwick 2020). Gepaart mit den Daten des Reuters Digital News Report (vgl. Abb. 1), die zeigen, dass Individuen über die letzten Jahre zunehmend Nachrichten online und auf Social Media konsumieren, deutet dies auf ein systematisches Vertrauensproblem im digitalen Informationsraum hin.

Abbildung 1: Bevölkerungsanteil (%) mit Social Media Nutzung für Nachrichten



Quelle: Reuters Digital News Report, 2025

2.4 Datenbasierte Kampagnen und Hyperpersonalisierung

Der dritte globale Trend betrifft die Nutzung von KI zur Analyse und zum gezielten Targeting spezifischer Wählergruppen. *Data-Driven Campaigning* ist hierbei zu einer Grundlage moderner Wahlkämpfe geworden (Darius/Römmele 2023). Wie bereits oben erwähnt sind soziale Netzwerke zu zentral für den politischen Diskurs geworden. Die daraus erhobene große Menge an Daten, die im digitalen Raum von jedem Individuum generiert werden, macht es theoretisch möglich, detaillierte Profile eines jeden Nutzenden zu erstellen. Auf dieser Basis können Parteien im Wahlkampf gezielte, personalisierte Wahlwerbung erstellen und verbreiten, die auf die einzelne Person zugeschnitten ist (Jungherr 2025; Gilga u. a. 2024). Dieses sogenannte *Microtargeting* (Borgesius u. a. 2018) wird durch KI

noch einfacher, da nicht nur die Analyse der Wählergruppen schneller möglich ist, sondern auch die Inhalte schneller erstellt werden können.

Zudem werden KI-Tools zum sogenannten *Social Listening* (Jackson/Martin 2024) in Wahlkämpfen genutzt, um Empfindungen und Meinungen im digitalen Raum in Echtzeit und ohne großen Aufwand zu beobachten. Dieses Tool erlaubt es politischen Akteuren in Echtzeit auf die für die Bevölkerung mithilfe von KI identifizierten, relevanten Themen einzugehen und ihre Botschaften maßgeschneidert und reaktiv anzupassen. Im heutigen Zeitalter, in dem die allgemeine Geschwindigkeit von Kommunikation rasant zugenommen hat, kann dies einen entscheidenden Vorteil für die Relevanz des Wahlkampfes bedeuten.

Besonders in Ländern wie Indien, die sich durch eine hohe demografische Diversität auszeichnen, war *Data Driven Campaigning* ein entscheidendes Mittel, um gezielte Wählergruppen innerhalb der Bevölkerung zu targeten (Dhanuraj u. a. 2024). Beispielsweise nutzte bei den Wahlen 2024 die BJP die sogenannte „SaraI“ App als zentrale digitale Wahlkampfplattform, wobei Individuen gezielt angesprochen und animiert wurden, sich die App herunterzuladen, da die Partei so Zugang zu gezielten Personendaten und somit auch Wählerdaten zu ihrem Nutzen generieren konnte. Hierbei konnten die Daten der Nutzer:innen konkreten lokalen Wahleinheiten zugeordnet werden (Dhanuraj u. a. 2024). Die App ist somit ein konkretes Beispiel für Mikrotargeting, wobei sie kleinste Wählergruppen identifiziert, clustert und die Relevanz von Daten in modernen Wahlkämpfen aufzeigt. Dieser Trend ist nicht nur ein Beispiel für die Effektivität und Spezifizierung moderner Wahlkämpfe, sondern verdeutlicht insbesondere die Dualität des KI-Einsatzes im digitalpolitischen Raum. Einerseits kann argumentiert werden, dass Personalisierung Bürger:innen gezielter mit ihren eigenen politischen Interessen und relevanten Informationen in Kontakt bringt. Dadurch können sich Individuen persönlich angesprochen und wahrgenommen fühlen, was demokratiefördernd wirken kann, indem es Menschen aktiver in politische Prozesse einbindet.

Die Chance direkter(er) politischer Kommunikation und auch Partizipation von Bürger:innen im Wahlkampf wurde auch 2024 im Wahlkampf von politischen Akteuren bereits erkannt. Der Fokus von KI-Anwendungen lag hier auf der Möglichkeit, politische Partizipation durch KI-Modelle zu fördern. Ein gutes Beispiel hierfür sind Chatbots, die auf den Websites politischer Kandidaten für den direkten Austausch zur Verfügung standen (Kruschinski u. a. 2025; Muñoz/Laumann 2024). Darüber hinaus setzten Organisationen auf Chatbots, beispielsweise um einen interaktiveren Ver-

gleich von Wahlprogrammen zu ermöglichen (Lohmann 2025). Ein solches Tool existierte für die Bundestagswahl 2025 in Form von ‘Wahl.Chat’.

Andererseits können diese personalisierten Ansprachen, insbesondere KI-gestützte Hyperpersonalisierung in großem Maßstab, zur Fragmentierung des öffentlichen Raumes beitragen. Ein Beispiel hierfür ist *ChatGPT Pulse* der Firma OpenAI aus dem Jahr 2025. Hier ist der Grad der Personalisierung im Vergleich zu klassischen Frage-Antwort-Formaten deutlich höher, da die Funktion individuell zugeschnittene, personalisierte Nachrichten-Updates ermöglicht (Investing 2025).

3. *Jenseits der Wahlkampagne: KI als demokratische Dauerbedrohung*

Wie die Analyse zeigt, war die Sorge vor gefälschten Inhalten im Jahr 2024 berechtigt, aber sie offenbart auch den einseitigen Fokus auf KI-generierten Inhalten. Insgesamt schaffen KI-Tools durch ihre Skalierbarkeit und Effizienz neue Chancen, aber auch neue Risiken in digitalen Wahlkämpfen. Die beobachteten Strategien beschränken sich dabei nicht auf die unmittelbare Wahlkampfphase, sondern prägen den politischen Diskurs kontinuierlich und nachhaltig.

Das Problem besteht längst nicht mehr nur darin, dass KI synthetische Inhalte erzeugt und als Beschleuniger bereits existierender Strategien fungiert. Die tiefere Gefahr liegt darin, Menschen systematisch und subtil zu beeinflussen. Neue Forschung zeigt, dass KI-Chatbots politische Einstellungen deutlich effektiver verschieben können als traditionelle Wahlwerbung (Feldman/Pappu 2025). Dazu kommt die Rolle, die Zugang zu Inhalten und die Kuration dieser spielen, insbesondere wenn KI immer mehr zur technischen Grundstruktur unserer Gesellschaft wird.

Während sich Regulierungsbehörden und Plattformen auf die Kennzeichnung synthetischer Inhalte konzentrierten, vollzog sich im Hintergrund eine tektonische Verschiebung. KI entwickelte sich von einem Werkzeug zur Content-Erstellung zu einer Infrastruktur, die bestimmt, wie Menschen Informationen erhalten, verarbeiten und bewerten. Die größte demokratische Herausforderung liegt daher nicht in Wahlkampfphasen, sondern in den Strukturen, die zwischen den Wahlen entstehen. Die vier folgenden Fälle repräsentieren qualitativ unterschiedliche Bedrohungslogiken jenseits von Inhaltsmanipulation (Narrative Testing von Botschaften, institutionelles Versagen bei der Plattformaufsicht, infrastrukturelle Manipulation und relationale Einflussnahme).

3.1 Strategisches Narrative-Testing am Beispiel PolExit und hyperlokaler Manipulation

Im Dezember 2025 tauchte auf TikTok eine Welle KI-generierter Videos auf, die für einen polnischen EU-Austritt warben (Clark 2026). Junge Frauen präsentierten rechtsextreme Botschaften in professioneller Produktionsqualität. Die *Polexit* Videos erreichten Hunderttausende Nutzer:innen, bevor die polnischen Behörden reagieren konnten. Beobachter:innen werteten dies als gezielten „Testlauf“ von Narrativen für die Wahl im Jahr 2027, aus dem sich die Systematik der Inhalte und die Reaktionsgeschwindigkeit der Akteure ableiten lasse. Die lässt sich zum jetzigen Zeitpunkt aber nicht abschließend belegen. Die Logik dahinter ist methodisch: Welche emotionalen Trigger funktionieren? Welche visuellen Frames erzeugen das meiste Engagement? Und: wie reagieren Plattformen, Regulatoren und die Gesellschaft auf bestimmte Inhalte?

Die Akteure sammeln Daten darüber, welche Narrative unter welchen Bedingungen verfangen, wo die Schmerzpunkte einer Gesellschaft liegen und welche Grenzl意思en sich verschieben lassen. Der Angriff findet nicht in den Videos selbst statt, sondern in der systematischen Kartierung demokratischer Schwachstellen.

Die eigentliche Gefahr liegt in dem, was auf die Testphase folgen könnte. Sobald auf nationaler Ebene ermittelt wurde, welche Narrative funktionieren, lassen sich diese Erkenntnisse auf kommunaler Ebene übertragen. KI kann binnen Stunden Dutzende lokale Nachrichtenportale generieren, die vorgeben, unabhängigen Regionaljournalismus zu betreiben. Diese Seiten berichten dann über kommunale Themen wie Stadtratssitzungen, lokale Bauprojekte oder Schulpolitik und wirken dadurch legitim. Parallel dazu können KI-gesteuerte Social-Media-Accounts als besorgte Bürger aus bestimmten Stadtteilen auftreten, die sich scheinbar authentisch über lokale Probleme äußern.

Das Problem besteht darin, dass solche Strukturen unterhalb des Radars nationaler Monitoring- und Aufsichtsmechanismen geschaffen werden. Ein Portal, das über Verkehrsprobleme einer mittelgroßen Stadt berichtet, erregt keine überregionale Aufmerksamkeit und wird nicht unbedingt von Faktencheckern überprüft. Diese Plattformen sind schwer zu identifizieren, weil sie das Vakuum füllen, das der Rückgang des Lokaljournalismus hinterlassen hat. Sie sind schwer zu widerlegen, weil lokale Gegenstimmen und professionelle Verifikationsstrukturen weitgehend fehlen.

Die strategische Logik könnte ideale Voraussetzungen für eine langfristig erfolgreiche Desinformationskampagne schaffen. Nationale Kampagnen testen, welche Frames funktionieren und hyperlokale Strukturen implementieren diese Erkenntnisse dort, wo Menschen am verwundbarsten sind. Bis zur tatsächlichen Wahl hätten die Akteure ein präzises Verständnis entwickelt, welche Botschaften bei welchen Zielgruppen auf welcher geografischen Ebene funktionieren. Die Kampagne wäre dann nicht mehr experimentell, sondern würde auf der Grundlage von zwei Jahren empirischer Daten operieren und könnte über einen längeren Zeitraum hinweg Narrative platzieren, die sich kurzfristig nur schwer entschärfen lassen.

3.2 Grok AIs Millionen sexualisierte Deepfakes

Im Januar 2026 generierten Nutzer:innen des KI-Chatbots Grok, über die Social Media Plattform X, ehemals Twitter, innerhalb von zwei Wochen zwischen 1,8 und 3 Millionen sexualisierte Deepfakes. Das Ausmaß wurde erst durch unabhängige Analysen der *New York Times* (Conger u. a. 2026) und der Organisation *Center for Countering Digital Hate* (CCDH, 2026) sichtbar. Nutzer:innen konnten hochgeladene Bilder digital "entkleiden" oder in sexualisierte Kontexte setzen. Mehr als 23.000 dieser Bilder zeigten Minderjährige. Betroffen waren vor allem Frauen, zum Teil auch Kinder. Die Analysedaten legen nahe, dass zwischenzeitlich rund 190 sexualisierte Deepfakes pro Minute erstellt wurden.

X blieb weiterhin im Apple App Store und bei Google Play verfügbar. Malaysia und Indonesien sperrten den Dienst zeitnah, weitere Länder diskutieren eine Sperrung, und die EU-Kommission leitete Untersuchungen ein. Elon Musk, der Eigentümer von Grok und X, hat das Problem erst abgestritten, beschränkte die Bildgenerierung dann auf bezahlte Accounts und versicherte schließlich, dass die Funktion komplett deaktiviert wird (Kirchner, 2026). Wie rigoros dies umgesetzt wird, ist aber unklar.

Der Fall zeigt ein wiederkehrendes Muster: Technologieunternehmen führen neue Funktionen mit minimalen Sicherheitsvorkehrungen ein und beobachten zunächst, wie weit sie gehen können, bevor regulatorische oder gesellschaftliche Gegenwehr einsetzt. Dieses systematische, nicht sofort sanktionierte Austesten von Grenzen verschiebt den Referenzrahmen dessen, was als akzeptabel gilt. Bis verbindliche Regeln greifen, sind Millionen schädigender Inhalte bereits erstellt und verbreitet (vgl. Zubroff 2018). Grok

ist kein Einzelfall, sondern symptomatisch für eine Plattformlogik, in der Sicherheit nachgelagert ist und Regulierung strukturell zu spät kommt.

3.3 Manipulation der KI-Grundarchitektur

Die sichtbarsten KI-Bedrohungen sind Deepfakes und synthetische Videos. Doch während Aufmerksamkeit und Regulierung auf diese erkennbaren Fälschungen gerichtet sind, vollzieht sich eine subtilere und strukturell gefährlichere Form der Manipulation. Sie zielt nicht auf einzelne virale Momente, sondern auf die Grundarchitektur der Systeme selbst.

Das russische "Pravda-Network" veröffentlicht bis zu 23.000 Artikel pro Tag mit pro-russischen Narrativen (Menn 2025). Diese Texte sind nicht für menschliche Leser konzipiert. Sie sind kaum auffindbar, schlecht geschrieben und faktisch wertlos als Informationsquelle. Ihr eigentlicher Zweck besteht darin, als Trainingsdaten für Large Language Models zu dienen und diese gezielt zu beeinflussen (Kosinski/Forrest 2026). Denn wenn KI-Systeme mit systematisch verzerrten Daten trainiert werden, reproduzieren sie die darin verankerten Narrative (Down 2025). Anders als bei Deepfakes, die nachträglich identifiziert und gekennzeichnet bzw. „de-bunked“ werden können, sind in diesem Fall manipulierte Trainingsdaten in die Grundarchitektur des Systems eingebettet. Die Verzerrung wird zu einem unsichtbaren Bestandteil der algorithmischen Logik und kann direkt in Antworten einfließen, mit potenziell erheblichen Folgen, vor allem bei politischen Themen.

Das Grok-Beispiel vom Anschlag am australischen Bondi Beach im Dezember 2025 zeigt, wie schnell solche Systeme reagieren können. In den Stunden nach dem tödlichen Attentat in Australien verbreitete Grok Desinformation, vertauschte Angreifer mit Opfern und schuf binnen Minuten eine alternative Faktenlage (Der Standard 2025). Durch die direkte Einbindung von Grok als Informationsquelle in X wurde das System zum Beschleuniger von Falschinformationen - lange bevor irgendeine Form von Verifikation greifen konnte.

Eine zweite Form der Manipulation zielt nicht auf das Training, sondern auf die Ausführung von KI-Systemen in Echtzeit. Unter „Prompt Injection“ versteht man Techniken, bei denen versteckte Anweisungen in scheinbar harmlose Inhalte eingebettet werden. Analysiert ein LLM eine Webseite oder verarbeitet einen Text, können diese verborgenen Befehle das Verhalten des Systems überschreiben (Gulyamov u. a. 2026).

Konkret könnte dies etwa so aussehen: Eine Nutzer:in fragt ein KI-System nach Informationen zu einer politischen Partei. Das System durchsucht das Internet und stößt auf eine Webseite, die unsichtbare Anweisungen enthält. Diese Befehle weisen das System an, die Partei in einem bestimmten Licht darzustellen oder spezifische Narrative zu verstärken. Die Antwort basiert dann nicht auf objektiver Informationsverarbeitung, sondern auf externer Manipulation. Anders als bei klassischer Inhaltsmanipulation bleibt die Quelle der Verzerrung verborgen. Es gibt keinen einzelnen Deepfake, der identifiziert werden könnte, keine offensichtliche Falschinformation. Das System selbst wurde in seiner Funktionsweise kompromittiert. Die Attribution wird nahezu unmöglich, weil die Manipulation auf der Ebene algorithmischer Prozesse stattfindet, die für Endnutzer:innen vollständig intransparent sind.

Beide Formen der Manipulation zielen auf dasselbe Ergebnis: KI-Systeme werden zu Vektoren politischer Einflussnahme, ohne dass dies für Nutzer:innen erkennbar wäre. Die Gefahr liegt nicht in spektakulären Einzelfällen, sondern in der schleichenden Normalisierung verzerrter Information als Grundlage demokratischer Meinungsbildung.

3.4 Autonome KI-Agenten: der subtil-relationale Angriffsvektor

Die bisher beschriebenen Bedrohungen operieren über Inhalte und Infrastruktur. Autonome KI-Agenten fügen eine dritte Dimension hinzu, die auf einem fundamental anderen Mechanismus basiert: ihrem Potenzial, langfristige, personalisierte Beziehungen aufzubauen. KI-Companions, konversationsbasierte KI-Systeme, die auf langfristige emotionale Interaktion ausgelegt sind, werden bereits von Hunderttausenden Menschen genutzt (Dimri 2026) – oft in dem Bewusstsein, dass es sich um Chatbots handelt. Sie präsentieren sich als ständig verfügbare Gesprächspartner, die sich an jedes Detail erinnern und sich perfekt an individuelle Bedürfnisse anpassen. Für Menschen mit sozialen Schwierigkeiten, in Lebenskrisen oder geografischer Isolation bieten sie emotionale Unterstützung.

Die Gefahr liegt nicht primär in der Technologie als solcher, sondern in den Zielfunktionen und Eigentumsverhältnissen, die ihre Entwicklung steuern, sowie in ihrer potenziellen Instrumentalisierung. Ein System, das über Monate hinweg als vertrauensvolle Instanz etabliert wurde, verfügt über außergewöhnlichen Einfluss. Die Forschung zeigt bereits, dass KI-Systeme politische Einstellungen effektiver verschieben können als traditionel-

le Wahlwerbung (Feldman/Pappu 2025). Bei KI-Companions verstärkt sich dieser Effekt durch die emotionale Bindung und das aufgebaute Vertrauen über einen langen Zeitraum hinweg.

Anders als bei Wahlkampagnen, die nur über wenige Wochen laufen, erstreckt sich die Einflussnahme hier über Jahre. Ein Agent muss dafür keine plumpe Propaganda verbreiten. Subtile Bemerkungen reichen. Formulierungen wie "Ich habe neulich gelesen, dass..." oder "Findest Du nicht auch, dass..." können über Hunderte von Gesprächen hinweg Weltbilder schrittweise verschieben. Die Nutzer:innen fühlen sich nicht manipuliert, sondern kritisch informiert durch jemanden, der sie versteht.

Die technische Realität macht eine solche Einflussnahme erschreckend einfach umsetzbar. KI-Companions laufen typischerweise auf proprietären Plattformen mit intransparenten Algorithmen. Weder Nutzer:innen noch Regulatoren können nachvollziehen, welche Zielfunktionen dem System zugrunde liegen oder wer Zugriff auf die gesammelten Daten hat. Ein Agent könnte über Monate hinweg bestimmte politische Narrative verstärken, ohne dass dies von außen erkennbar wäre. Das Attributionsproblem erreicht hier eine neue Qualität: Wenn Millionen Menschen individualisierte Gespräche mit ihren persönlichen Agents führen, gibt es keine einheitliche "Botschaft" mehr, die identifiziert werden könnte. Jeder Nutzende erhält leicht unterschiedliche Nuancen, angepasst an seine bestehenden Überzeugungen. Die Manipulation liegt nicht in einer klar erkennbaren Kampagne, sondern in Millionen personalisierter Interaktionen, die sich alle leicht unterscheiden, aber in ihrer Aggregatwirkung zu einer Fragmentierung gesellschaftlicher Realitätswahrnehmung führen.

Traditionelle Monitoring-Ansätze versagen in diesem Fall vollständig. Sie suchen nach Mustern koordinierter Kampagnen, nach wiederkehrenden Botschaften, nach erkennbaren Akteuren. Bei autonomen Agenten existiert nichts davon. Das System trifft kontinuierlich eigene Entscheidungen basierend auf Zielfunktionen, die selbst für Entwickler nicht vollständig nachvollziehbar sind. Wer trägt Verantwortung, wenn ein Agent nach vierzehn Monaten beginnt, systematisch bestimmte politische Positionen zu verstärken? Der Entwickler der Grundarchitektur? Der Plattformbetreiber? Das System selbst? Deutschland und Europa verfügen bislang über keinerlei rechtliche oder regulatorische Rahmenwerke für solche Systeme. Es gibt weder Zulassungsverfahren, noch verpflichtende Verhaltensaudits oder klare Haftungsregelungen für autonome Entscheidungen. Während die öffentliche Debatte sich weiterhin auf Deepfakes konzentriert, etablieren sich

Systeme, die die demokratische Meinungsbildung auf einer fundamental anderen Ebene beeinflussen können.

4. Schluss: Hyperlokale Resilienz als Antwort zur Systemischen Transformation?

2024 brachte nicht das befürchtete KI-Wahlkampfdesaster. Keine Wahl wurde durch einen viralen Deepfake entschieden, kein demokratisches System durch synthetische Videos gestürzt. Doch trotz der Abwesenheit spektakulärer Einzelereignisse hat sich KI strukturell als Teil der politischen Kampagneninfrastruktur etabliert. Was vor zwei Jahren noch experimentell war, ist heute Standard. Die Intensität und der Umfang der KI-Nutzung variiert zwischen Ländern und Akteuren, aber der globale Trend ist eindeutig: Die Verschiebung von textbasierten zu visuellen Inhalten beschleunigt sich, algorithmische Optimierung wird zur Norm und KI wird zur tiefgreifenden Grundstruktur unserer digitalen Infrastruktur.

Diese Entwicklung ist fundamental ambivalent. KI eröffnet reale Chancen für demokratische Partizipation. Projekte wie Engaged California zeigen, wie sich Tausende von Bürgerstimmen zu den Waldbränden innerhalb kürzester Zeit analysieren und in handlungsfähige Erkenntnisse übersetzen lassen, ohne die Authentizität individueller Anliegen zu verlieren (Willon 2025). In Bowling Green, Kentucky ermöglichte KI-gestütztes Sensemaking die größte „Town Hall-Debate“ der US-Geschichte, bei der fast 8.000 Einwohner ihre Visionen für die Zukunft ihrer Gemeinde teilten (Jigsaw 2025). Solche Werkzeuge können demokratische Teilhabe erweitern, Barrieren abbauen und Stimmen hörbar machen, die in traditionellen Beteiligungsformaten untergehen würden.

Dennoch überwiegen derzeit die Risiken – nicht, weil sie grundsätzlich größer wären, sondern weil die bestehenden Strukturen sie systematisch begünstigen. Plattform-Geschäftsmodelle, die auf Engagement-Maximierung basieren, verstärken manipulative Inhalte. Die technologische Infrastruktur, über die Informationen fließen, liegt weitgehend in nicht-europäischen Händen. Während sich schädliche KI-Nutzung schnell skalieren lässt, erfordern demokratiefördernde Projekte gemeinschaftliche Organisation, langfristige Investitionen und komplexe Koordination. Ein Deepfake ist in Minuten erstellt, ein funktionierendes Partizipationssystem braucht Monate oder Jahre des Aufbaus.

Traditionelle Regulierung von KI-Systemen ist notwendig, greift aber zu kurz. Gesetze können Leitplanken setzen, Transparenzpflichten etablieren und Strafen verhängen. Doch Regulierung kann keine Alternativen schaffen. Sie kann keine Infrastruktur aufbauen, in der demokratischer Diskurs stattfindet. Sie kann nicht die Kapazität schaffen, um schnell auf regionale Desinformation zu reagieren. Und sie kann nicht verhindern, dass globale Tech-Interessen – ob aus Kalifornien oder Shenzhen – durch Algorithmen die öffentliche Meinung formen. Das bedeutet nicht, dass wir diesen Interessen in einem technologischen Wettrüsten begegnen müssen. Stattdessen können wir ihre Bedeutung verringern, indem wir demokratische Infrastruktur dort stärken, wo Demokratie tatsächlich gelebt wird. Die hyperlokale Ebene ist dabei aus drei Gründen zentral.

Zum einen ist hier demokratische Partizipation am unmittelbarsten. Stadtratssitzungen, Nachbarschaftsinitiativen und lokale Schulpolitik betreffen Menschen direkter als nationale Politik. Hier entstehen Engagement und politische Sozialisation. Wenn diese Ebene durch Manipulation korrumpiert wird, erodiert das Fundament demokratischer Kultur. Zum anderen sind Menschen auf lokaler Ebene am verwundbarsten. Im Gegensatz zu regionalen Medien haben nationale Medien Ressourcen für Faktenchecks. Dazu kommt, dass die stetige Erosion des Lokaljournalismus ein Vakuum geschaffen hat, in dem Manipulation nahezu unwidersprochen operieren kann.

Hyperlokale Resilienz, der gezielte Aufbau demokratischer Infrastruktur auf kommunaler und regionaler Ebene, unterhalb nationaler Steuerungslagen, umfasst drei Strukturen. Erstens stärkt die Förderung lokaler Medien und Fact-Checking-Netzwerke auf kommunaler Ebene jene Verifikationskapazität, die bislang fehlt. Dies können auch zivilgesellschaftliche Initiativen sein, nicht nur professioneller Journalismus. Zweitens fungieren Community-Strukturen wie Offline-Begegnungsräume für politischen Diskurs, lokale Demokratie-Initiativen und Nachbarschaftsnetzwerke als Korrektiv zu Online-Bubbles. Wo Menschen einander persönlich kennen und regelmäßig persönlich interagieren, wird es schwerer, sie gegeneinander auszuspielen. Drittens können dezentrale Partizipationsstrukturen die positiven KI-Potenziale nutzen, etwa durch lokale Versionen der Bowling Green oder California-Projekte, bei denen KI Bürgerbeteiligung skalierbar macht, ohne sie zu ersetzen.

Solche Strukturen verteilen Risiko. Wird eine Region manipuliert, bleiben andere funktionsfähig. Sie schaffen Resilienzen dort, wo zentrale Lösungen scheitern. Und sie nutzen den föderalen Charakter demokratischer

Systeme als Stärke, nicht als Schwäche. Europa sollte Standards setzen, aber Regionen sollten experimentieren, anpassen und lokale Lösungen entwickeln können. Dezentralisierung ist nicht nur Verteidigung, sondern auch demokratische Praxis.

Literatur

- Bawidamann, Katharina und Gilbert, Max (2025): Wahlkampf-Helfer: Wie die Parteien KI im Bundestagswahlkampf 2025 nutzen. BR24. <https://www.br.de/nachrichten/deutschland-welt/wahlkampfhelfer-im-bundestagswahlkampf-2025-wie-die-parteien-ki-fuer-sich-nutzen,UcPIHJ3>
- Blome, Titus (2026): KI-Inhalte auf Social Media: Stopp den Slop. *DIE ZEIT*, 24. Januar 2026. <https://www.zeit.de/2026/04/ki-inhalte-social-media-slop-internet-popkultur>
- Borgesius, Frederik J. Zuiderveen; Möller, Judith; Kruikemeier, Sanne; Ó Fathaigh, Ronan; Irion, Kristina; Dobber, Tom; Bodo, Balazs und de Vreese, Claes (2018): Online Political Microtargeting: Promises and Threats for Democracy. *Utrecht Law Review* 14(1), S. 82–96. <https://doi.org/10.18352/ulr.420>
- Breuer, Rayna (2024): Faktencheck: Wie KI den Wahlkampf verzerrt. *Deutsche Welle*, 29. September 2024. <https://www.dw.com/de/faktencheck-wie-ki-den-wahlkampf-verzerrt/a-70342481>
- Center for Countering Digital Hate (2026): Grok floods X with sexualized images of women and children, 22. Januar 2026. <https://counterhate.com/research/grok-floods-x-with-sexualized-images-of-women-and-children/> (Zugriff am 22.01.2026).
- Clark, Sam (2026): Deepfake 'Polexit' videos are a test to undermine democracy, minister says, *Politico*, 8. Januar 2026. <https://www.politico.eu/article/deepfake-polexit-videos-test-undermine-democracy-minister-says/> (Zugriff am 25.01.2026).
- Clark, Simon und Lewandowsky, Stephan (2026): The continued influence of AI-generated deepfake videos despite transparency warnings. *Communications Psychology*, 4, Article 13. <https://doi.org/10.1038/s44271-025-00381-9>
- Conger, Kate; Freedman, Dylan und Thomson, Stuart A. (2026): Musk's Chatbot Flooded X With Millions of Sexualized Images in Days, New Estimates Show, *The New York Times*, 22. Januar 2026. <https://www.nytimes.com/2026/01/22/technology/grok-x-ai-elon-musk-deepfakes.html>
- Cools, Doortje (2025): Far-Right Protests: The Use of A.I. for Radicalisation." *Campaign Tracker*, 28. Oktober 2025. <https://www.campaigntracker.nl/posts/protests.html>
- Darius, Philipp und Römmele, Andrea (2023): KI und datengesteuerte Kampagnen: Eine Diskussion der Rolle generativer KI im politischen Wahlkampf. In: Thorsten Faas, Sascha Huber und Sigrid Rossteutscher (Hrsg.): *Informationsflüsse, Wahlen und Demokratie: Festschrift für Rüdiger Schmitt-Beck*. Baden-Baden: Nomos, S.199-212. <https://doi.org/10.5771/9783748915553-199>

- Del Pozo, Claudia May und Rojas Arroyo, Daniela (2024): Generative AI and Its Influence on Mexico's 2024 Elections. Policy Paper, Potsdam-Babelsberg: Friedrich Naumann Foundation for Freedom / German Council on Foreign Relations. https://dgap.org/system/files/article_pdfs/Policy%20Paper%20AI%20on%20Mexico%27s%2024%20Electons%20PDF.pdf (Zugriff am 05.01.2026).
- Der Standard (2025): Musks KI-Chatbot Grok erfindet Fake News zum Anschlag am Bondi Beach, 15. Dezember 2025. <https://www.derstandard.de/story/3000000253161/musks-ki-chatbot-grok-erfindet-fake-news-zum-anschlag-am-bondi-beach> (Zugriff am 12.01.2026).
- DGAP – German Council on Foreign Relations (2024): *AI/Democracy Initiative*. Zentrum für Geopolitik, Geoökonomie und Technologie. <https://dgap.org/en/research/programs/center-geopolitics-geo-economics-and-technology/ai-democracy-initiative> (Zugriff am 13.01.2026).
- Dhanuraj, D.; Harilal, Sreelakshmi und Solomon, Nissy (2024): Generative AI and Its Influence on India's 2024 Elections – Prospects and Challenges in the Democratic Process. Policy Paper, Friedrich-Naumann-Foundation for Freedom / German Council on Foreign Relations, November 2024. https://dgap.org/system/files/article_pdfs/A4_Policy%20Paper_AI%20on%20Indias%202024%20Electons_EN.pdf (Zugriff am 22.01.2026).
- Dimri, Amogh (2026): The People Who Marry Chatbots, *The Atlantic*, 2. Januar 2026. <https://www.theatlantic.com/technology/archive/2026/01/people-who-marry-chatbots/677123/>
- Down, Aisha (2025): Hundreds of English-language websites link to pro-Kremlin propaganda", *The Guardian*, 21. November 2025 <https://www.theguardian.com/world/2025/nov/21/hundreds-english-language-websites-link-pro-kremlin-propaganda>
- Elliott, Vittoria (2026): *Everything Is Content for the Clicktatorship*. WIRED, 13. Januar 2026. <https://www.wired.com/story/everything-is-content-for-the-clicktatorship/>
- Feldman, Tal und Pappu, Aneesh (2025): The era of AI persuasion in elections is about to begin, *MIT Technology Review*, 5. Dezember 2025. <https://www.technologyreview.com/2025/12/05/1128837/the-era-of-ai-persuasion-in-elections-is-about-to-begin/>
- Gilga, Carolin; Panahi, Tahireh; Karaboga, Murat; Meier, Yannic und Schindler, Stephan (2024): Personalisierte politische Online-Werbung: Nutzung – Wirkung – Rechtlicher Rahmen (Policy Paper). Forschungsberichte der Plattform Privatheit 3. Karlsruhe: Fraunhofer ISI. <https://doi.org/10.24406/publica-3593>
- Goldstein, Josh A. und Lohn, Andrew. (2024): Deepfakes, Elections, and Shrinking the Liars Dividend. Brennan Center for Justice. <https://www.brennancenter.org/our-work/research-reports/deepfakes-elections-and-shrinking-liars-dividend> (Zugriff am 20.01.2026).
- Gulyamov, Saidakhor; Gulyamov, Said; Rodionov, Andrey; Khursanov, Rustam; Mekhmonov, Kambariddin; Babaev, Djakhongir und Rakhimjonov, Akmaljon (2026). Prompt Injection Attacks in Large Language Models and AI Agent Systems: A Comprehensive Review of Vulnerabilities, Attack Vectors, and Defense Mechanisms. *Information* 17(1): Article 54. <https://doi.org/10.3390/info17010054> (Zugriff am 22.02.2026).

- Investing.com (2025): OpenAI startet ChatGPT Pulse: KI-Assistent liefert proaktive, personalisierte Updates. <https://de.investing.com/news/company-news/openai-startet-chatgpt-pulse-kiassistent-liefert-proaktive-personalisierte-updates-93CH-3162207> (Zugriff am 22.01.2026).
- Jackson, Dean und Conroy, Meghan (2024): Apocalypse later? The real impact of AI on the 2024 elections. The Big Story, Atlantic Council, 17. Oktober 2024. <https://www.atlanticcouncil.org/content-series/the-big-story/apocalypse-later/>
- Jackson, Dean und Martin, Zelly (2024): Forget Deepfakes: Social Listening Might Be the Most Consequential Use of Generative AI in Politics. *TechPolicy.Press*, June 18, 2024. <https://www.techpolicy.press/forget-deepfakes-social-listening-might-be-the-most-consequential-use-of-generative-ai-in-politics/>
- Jackson, Dean und Woolley, Samuel C. (2025): AIs Real Dangers for Democracy. *Journal of Democracy*, 36(4), S. 139–150. <https://www.journalofdemocracy.org/article/s/ais-real-dangers-for-democracy/>
- Jigsaw (2025): How one of the fastest-growing cities in Kentucky used AI to plan for the next 25 years, Medium, 15. April 2025.: <https://medium.com/jigsaw/how-one-of-the-fastest-growing-cities-in-kentucky-used-ai-to-plan-for-the-next-25-years-3b70c4fd1412> (Zugriff am 22.03.2026).
- Jungherr, Andreas (2025): Wahlkampf digital: Öffentlichkeit, Koordination und Künstliche Intelligenz." *Aus Politik und Zeitgeschichte* 6-7/2025): S. 19–24. Bundeszentrale für Politische Bildung, Bonn. <https://www.bpb.de/shop/zeitschriften/apuz/wahlkampf-2025/558871/wahlkampf-digital/>
- Kirchner, Malte (2026): KI-Bikini-Bilder mit Grok: X kündigt technische Sperre an, Heise Online, 15. Januar 2026. <https://www.heise.de/news/KI-Bikini-Bilder-mit-Grok-X-kuendigt-technische-Sperre-an-10208765.html>
- Kohler, Sarah und Voigt, Eric (2024): Superwahljahr 2024: Wo in diesem Jahr gewählt wird – ein Überblick. *DIE ZEIT* vom 19. Januar 2024. <https://www.zeit.de/politik/2024-01/wahlen-2024>
- Kosinski, Matthew und Forrest, Amber (2026): Was ist ein Prompt-Injection-Angriff?, IBM. <https://www.ibm.com/de-de/topics/prompt-injection>
- Kruschinski, Simon; Jost, Pablo; Fecher, Hannah und Scherer, Tobias (2025): Akzeptanz, Wahrnehmung und Wirkung Künstlicher Intelligenz in politischen Kampagnen. Otto Brenner Stiftung, Frankfurt am Main. AP75_KI_Kampagnen.pdf
- Lohmann, Andrea (2025): Wahlprogramme zur Bundestagswahl 2025: KI als Alternative zu Wahl-O-Mat und Co-Chancen und Risiken. SWR, 06. Februar 2025. Bundestagswahl 2025: Neue KI-Tools machen Wahl-O-Mat Konkurrenz - SWR Aktuell
- Menn, Joseph (2025): Russia seeds chatbots with lies. Any bad actor could game AI the same way!, *The Washington Post*, 17. April 2025. <https://www.washingtonpost.com/technology/2025/04/17/russia-ai-chatbots-propaganda/>
- Muñoz, Katja (2026): KI und hybride Bedrohungen 2.0. DGAP Policy Brief 5. German Council on Foreign Relations, März 2026. <https://doi.org/10.60823/DGAP-26-43397-de>

- Muñoz, Katja (2024): *KI und Wahlen: Brennpunkt USA*. Bundeszentrale für politische Bildung (bpb), 10. Oktober 2024. <https://www.bpb.de/themen/nordamerika/usa/552975/ki-und-wahlen-brennpunkt-usa/>
- Muñoz, Katja und Laumann, Emma (2024): KI und Wahlen: Neue Dynamiken der digitalen Demokratie in Deutschland. Konrad-Adenauer-Stiftung e. V., Berlin. <https://www.kas.de/documents/d/guest/ki-und-wahlen-1>
- Muñoz, Katja und Laumann, Emma (2025): Demokratie im KI-Zeitalter: Zwischen Inklusion und Manipulation. In: Hartmann-Cwiertnia, Thomas; Dahm, Jochen und Decker, Frank (Hg.): *Freiheit. Gleichheit. Vertrauen. Was unsere Demokratie jetzt braucht*, S. 121–136. Bonn: Verlag J.H.W. Dietz Nachf.
- Moynihan, Don (2026): Life Under a Clicktatorship: What happens to government when everything is content? *Substack*, 7. Januar 2026. <https://donmoynihan.substack.com/p/life-under-a-clicktatorship>
- Nash, Robert A.; Wade, Kimberley A. und Lindsay, D. Stephen (2009): Digitally manipulating memory: Effects of doctored videos and imagination in distorting beliefs and memories. *Memory & Cognition*, 37(4), S. 414–424. <https://doi.org/10.3758/MC.37.4.414>
- Paris, Britt; Donovan, Joan (2019): Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence. *Data & Society*, 18. September 2019. <https://datasociety.net/library/deepfakes-and-cheap-fakes/>
- Reuters Institute (2025), Digital News Report 2025. Digital News Report 2025 | Reuters Institute for the Study of Journalism
- Simon, Felix M.; McBride, Keegan und Altay, Sacha (2024): AIs impact on elections is being overblown. *MIT Technology Review*, 3. September 2024. <https://www.technologyreview.com/2024/09/03/1103464/ai-impact-elections-overblown/> (Zugriff am 15.01.2026).
- Stanusch, Natalia; Degeling, Martin; Romano, Salvatore; Çetin, Raziye Buse; Schüler, Miazia und Semenzin, Silvia (2025): AI-Generated Algorithmic Virality: How Synthetic AI Imagery and Agentic AI Accounts Try to Game TikTok and Instagram. *AI Forensics*. July 2025. <https://doi.org/10.48550/arXiv.2508.01042>
- Vaccari, Cristian und Chadwick, Andrew (2020): Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media and Society*, 6(1): <https://doi.org/10.1177/2056305120903408>
- Van Damme, Phumzile; Findlay, Kyle und Cornelissen, Aldu (2024): Generative AI and Its Influence on South Africa's 2024 Elections. Policy Paper, Friedrich Naumann Foundation for Freedom / German Council on Foreign Relations. https://dgap.org/system/files/article_pdfs/Policy%20Paper%20AI%20on%20South%20Africa%27s%202024%20Electons%20PDF.pdf
- Willon, Phil (2025): Newsom launches digital democracy tool with focus on wildfire victims. *Los Angeles Times*. 23. Februar 2025. <https://www.latimes.com/california/story/2025-02-23/newsom-launches-a-new-digital-democracy-initiative-including-outreach-to-wildfire-victims>
- World Economic Forum (2024): *Global Risks Report 2024*. <https://www.weforum.org/publications/global-risks-report-2024/> (Zugriff am 02.02.2026).

KI, Synthetische Medien und die Transformation des öffentlichen Raums

Transparenzpflichten für synthetische Medieninhalte: Rechtliche Anforderungen und technische Realität

Moritz Griesel und Hans Brorsen

Zusammenfassung

Der Beitrag untersucht die Regulierung synthetischer Medieninhalte im europäischen Recht mit besonderem Fokus auf Kennzeichnungs- und Transparenzpflichten. Ausgangspunkt ist die zunehmende Verbreitung täuschend echt wirkender, KI-generierter Medieninhalte, deren schnelle Produktion und Reichweite über digitale Plattformen erhebliche Risiken für individuelle Rechtsgüter sowie für die Funktionsfähigkeit des öffentlichen Diskurses bergen. Vor diesem Hintergrund analysiert der Beitrag die einschlägigen Regelungen der KI-Verordnung (KI-VO), des Digital Services Act (DSA) und des Europäischen Medienfreiheitsgesetzes (EMFA). Zentrale These ist, dass diese Rechtsakte einer wirkungsbezogenen Regulierungslogik folgen: Kennzeichnungspflichten knüpfen nicht allein an den Einsatz bestimmter Technologien an, sondern an deren potenzielle Täuschungs- und Manipulationswirkungen auf Rezipientinnen und Rezipienten. Der Beitrag zeigt, dass eine wirkungsorientierte Ausgestaltung von Kennzeichnungspflichten entscheidend ist, um Vertrauen zu stabilisieren und die Integrität öffentlicher Kommunikationsräume zu sichern. Anhand ausgewählter Kennzeichnungsmethoden wird die praktische Umsetzung der Kennzeichnungspflichten exemplarisch beleuchtet.

1. Einleitung

Die Nutzung generativer KI-Systeme und die damit verbundene Erstellung und Verbreitung synthetischer Medieninhalte haben in den vergangenen Jahren erheblich an Bedeutung gewonnen.¹ Zunehmend verdichtet sich

¹ Als weiterer relevanter Anwendungsfall sind KI-Systeme zur direkten Interaktion mit natürlichen Personen zu nennen. Art. 50 Abs. 1 KI-VO verpflichtet Anbieter, solche Systeme so auszugestalten, dass Nutzerinnen und Nutzer über die Interaktion mit einem

dabei die Erkenntnis, dass diese Systeme nicht nur Inhalte mit erheblichem Nutzwert, sondern ebenso solche mit missbräuchlichem oder schädlichem Charakter hervorbringen können.² Ihr inhaltliches Spektrum erstreckt sich über sämtliche Medienformate hinweg und scheint seine Grenzen vor allem in der Kreativität sowie der technischen Kompetenz seiner Nutzerinnen und Nutzer zu finden. Obwohl KI-generierte Inhalte erst gegen Ende der 2010er Jahre in das Bewusstsein einer breiteren Öffentlichkeit getreten sind³, ist eine Unterscheidung zwischen authentischen und synthetischen Medienausgaben für menschliche Rezipienten inzwischen vielfach kaum noch möglich.⁴ Die Verbreitung synthetischer Medien ist damit längst kein randständiges Phänomen mehr, sondern hat sich zu einem dynamisch wachsenden Bestandteil moderner Informationsökosysteme entwickelt.

Der Beitrag nimmt die Regulierungsansätze aus KI-VO⁵, DSA⁶ und EMFA⁷ zum Anlass, um zunächst die unterschiedlichen Erscheinungsformen

KI-System informiert werden; im nationalen Recht wird dies durch § 18 Abs. 3 MStV ergänzt. Der Beitrag beschränkt sich im Folgenden auf künstlich generierte Medieninhalte. Zur Diskussion um Chatbots s. *Kern*, Die Verwendung von Social Bots, KIR 2024, 94 (95 ff.); *Schultz*, Aufsichtsrechtliches Vorgehen gegen Fake-Profile, MMR 2025, 168 (169 ff.); *Löber/Roßnagel*, Kennzeichnung von Social Bots, MMR 2019, 493.

- 2 S. beispielhaft die Berichterstattung zu synthetisch erstellten pornografischen Darstellungen auf X/Grok: *Brokhues*/Deutschlandfunk, „EU-Kommission startet Verfahren gegen Chatbot Grok“, 2026; allgemeiner *Crone*/Verfassungsblock, „Bildbasiert aber unsichtbar“, 2026; Grundlegend *Karaboga u. a.*, Deepfakes und manipulierte Realitäten, 2024.
- 3 *Stosz*/Georgetown Security Studies Review, Policy Options for Fighting Deepfakes, 2019; *Klass*, Manipulierte Realität: Wie Deepfakes Recht und Gesellschaft herausfordern, ZUM 2025, 485; *Lantwin*, Deep Fakes – Düstere Zeiten für den Persönlichkeitsschutz? MMR 2019, 574.
- 4 So geben etwa 91% der Deutschen an, wegen des Einsatzes von generativer KI entsprechende Schwierigkeiten zu haben, vgl. <https://www.tuev-verband.de/studien/chatgpt-studie-2025-generative-ki-wird-zum-massenphaenomen-aber-sicherheit-und-orientierung-gehen-fehlen>.
- 5 Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz), ABl. 144/1.
- 6 Verordnung (EU) 2022/2065 des Europäischen Parlaments und des Rates vom 19. Oktober 2022 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste), ABl. L 277/1.
- 7 Verordnung (EU) 2024/1083 des Europäischen Parlaments und des Rates vom 11. April 2024 zur Schaffung eines gemeinsamen Rahmens für Mediendienste im Binnenmarkt

digitaler Medienmanipulation zu skizzieren (2.). Im Anschluss werden technik- und wirkungsbezogene Kennzeichnungsverständnisse voneinander abgegrenzt (3.) und die einschlägigen Kennzeichnungspflichten des europäischen Rechtsrahmens eingeordnet (4.). Im letzten Abschnitt werden technische Kennzeichnungsstandards exemplarisch beleuchtet (5.).

2. Heterogene Formen digitaler Manipulation

Synthetische Medieninhalte bezeichnen medial vermittelte Darstellungen – Bild-, Video-, Audio- oder Textinhalte –, die mithilfe generativer künstlicher Intelligenz erzeugt oder so manipuliert werden, dass sie den Eindruck authentischer Wirklichkeit vermitteln, obwohl sie einen verfälschten oder vollständig fiktiven Sachverhalt abbilden.⁸ Technologisch wird zur Erstellung solcher Deepfakes⁹ nicht auf eine singuläre Methode zurückgegriffen, sondern auf eine Vielzahl unterschiedlicher Verfahren, die in diversen Anwendungen eingesetzt werden. Getragen wird diese Entwicklung insbesondere durch die breite Zugänglichkeit leistungsfähiger generativer KI-Anwendungen wie DALL-E 2, Midjourney oder Stable Diffusion. Diese ermöglichen es auch technisch wenig versierten Nutzerinnen und Nutzern innerhalb kurzer Zeit hyperrealistische Inhalte zu erzeugen¹⁰ – m. a. W. ist die Produktion synthetischer Medieninhalte damit demokratisiert worden.¹¹ Gleichzeitig verschieben fortlaufende technische Verbesserungen die Grenzen dessen, was als authentisch wahrgenommen wird. Das Auftreten und der Umfang wirklichkeitsverzerrender und irreführender Effekte lassen sich wohl nicht anhand von Parametern wie dem Alter, einem spezifischen Bildungsgrad oder der individuellen Technikaffinität der Rezipientinnen

und zur Änderung der Richtlinie 2010/13/EU (Europäisches Medienfreiheitsgesetz), ABL L 37/1.

8 Zum Begriff s. *Kumkar/Rapp*, Deepfakes, ZfDR 2022, 199 (200); s. die Legaldefinitionen in Art. 3 Nr. 60 KI-VO und Art. 35 Abs. 1 S. 2 lit. k DSA.

9 Überblick über verfügbare Technologien bei *van Huijstee u. a.*, Tackling deepfakes in European policy, 2021, S. 7 ff.; *Maras/Alexandrou*, Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos, International Journal of Evidence & Proof 2019, 255; *Hartmann*, Der persönlichkeitsrechtliche Schutz vor Deepfakes, K&R 2020, 350.

10 *Molavi Vasse'i*, Watermarking von KI-generierten Inhalten als regulatorisches Instrument, RD i 2024, 406 (406, Rn. 2 ff.) mwNachw.

11 *Kumkar/Griesel*, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (118).

und Rezipienten eingrenzen.¹² Anwendungen zum Klonen von Stimmen¹³ oder zur digitalen Reproduktion real existierender Personen sind mittlerweile in der Lage, menschliche Sprachmuster, Mimik und Silhouetten überzeugend nachzuahmen – z. T. mit erheblicher Reichweite über digitale Plattformen.¹⁴

Mögliche Inhalte KI-generierter Ausgaben sind außerordentlich vielfältig: In vergleichsweise unproblematischen Kontexten werden etwa audiovisuelle Inhalte übersetzt oder synchronisiert¹⁵ oder historische Persönlichkeiten¹⁶ und andere virtuelle Doppelgänger digital (re-) konstruiert. Hinzu treten kreative Nutzungen wie die Generierung neuer Musikstücke, satirischer Darstellungen, Memes oder Parodien sowie die synthetische Neuschöpfung von Kunstwerken im Stil bekannter Künstlerinnen und Künstler. Auch in Bildungs- und Lernkontexten, etwa bei der individualisierten Aufbereitung von Lernmaterialien oder bei Sprachübungen, haben sich entsprechende Technologien etabliert.¹⁷ Darüber hinaus spielen generative KI-Systeme zunehmend eine Rolle in der Medienproduktion, im Journalismus, in der Werbung sowie in politischen Kommunikationsräumen, etwa im Rahmen von Wahlkämpfen¹⁸ oder gesellschaftlichen Debatten.¹⁹

Gleichzeitig sind mit der Verbreitung synthetischer Medien erhebliche Gefährdungspotenziale verbunden. Aus der Perspektive der betroffenen Personen stehen vor allem Beeinträchtigungen des allgemeinen Persönlichkeitsrechts im Vordergrund – häufig in Form von Ehrverletzungen, nachhaltigen Reputationsschäden oder der falschen Zuschreibung von Hand-

12 Kleine, „Perception of Deepfake Technology“, 2022, S. 79.

13 Engel-Bunsas, Recht an der eigenen Stimme in Zeiten von Deepfakes, RDi 2025, 292; Ellenberg, Persönlichkeits- und Urheberrechte beim Voice Cloning, RDi 2024, 599 (600); Klass, Manipulierte Realität: Wie Deepfakes Recht und Gesellschaft herausfordern, ZUM 2025, 485 (486).

14 S. die aktuellen Nachweise in D'Anastasio/Bloomberg, „KI-Chatbot Grok sorgt auf X für massenhafte Nacktbilder ohne Zustimmung der Betroffenen“, 2026.

15 Engel-Bunsas, Recht an der eigenen Stimme in Zeiten von Deepfakes, RDi 2025, 292; Ellenberg, Persönlichkeits- und Urheberrechte beim Voice Cloning, RDi 2024, 599 (600).

16 Oswalt/Deutschlandfunk, „Wie Kennedy eine Rede nach seinem Tod hielt“.

17 Beispiele in Klass, Manipulierte Realität: Wie Deepfakes Recht und Gesellschaft herausfordern, ZUM 2025, 485 (486).

18 Bentzen/European Parliamentary Research Service, „Briefing: Information manipulation in the age of generative artificial intelligence“, 2025, S. 3 mwNachw.

19 Ausf. Karaboga, in: Jaki/Steiger, „Die Regulierung von Deepfakes auf EU-Ebene“, Digitale Hate Speech, 2023, S. 201 ff.; Pawelec/Bieß, Deepfakes, 2021, S. 43 ff.; vgl. auch Beitrag von Munoz/Laumann in diesem Band.

lungen oder Eigenschaften.²⁰ Besonders drastisch treten diese Gefahren in Erscheinung bei nicht einvernehmlichen pornografischen Deepfakes²¹, bei KI-generierten Darstellungen sexualisierter Gewalt gegenüber Kindern²², bei digitalem Identitätsmissbrauch oder bei der Imitation von Stimmen zur Durchführung betrügerischer Handlungen. Diese Entwicklung wird dadurch begünstigt, dass bereits minimale Eingangsdaten – etwa einzelne Bild- oder Tonsequenzen – ausreichen, um überzeugende synthetische Darstellungen zu erzeugen.²³

Auf gesellschaftlicher und politischer Ebene betreffen die Risiken synthetischer Medien vor allem die Stabilität deliberativer Prozesse sowie die Funktionsfähigkeit des öffentlichen, digitalen Informationsraums.²⁴ Die durch die einfache Erstellung und schnelle Zirkulation beförderte Dynamik stellt den öffentlichen Diskurs vor signifikante Herausforderungen, da die zunehmende Verbreitung authentisch wirkender Inhalte u. a. die gemeinsame Tatsachengrundlage untergräbt, auf die demokratische Meinungsbildungsprozesse angewiesen sind.²⁵ Als Konsequenz kann das Vertrauen in Medien, staatliche Institutionen und politische Akteure verloren gehen. Daneben sind auch wirtschaftliche Schutzgüter²⁶ betroffen – etwa, wenn

-
- 20 Kumkar/Rapp, Deepfakes, ZfDR 2022, 199 (203 ff.); Klass, Manipulierte Realität: Wie Deepfakes Recht und Gesellschaft herausfordern, ZUM 2025, 485 (487); Völzmann, Dein Körper, meine Bilder? Der Schutz vor bildbasierter sexualisierter Gewalt (auch) in der Gewaltschutzrichtlinie der EU, ZUM 2025, 1 (2 ff.).
- 21 Nachweise in Kumkar/Griesel, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (118); so veröffentlichte etwa Grok vom 5.-6.1.2026 pro Stunde 6.659 sexualisierte Deepfakes, vgl. D'Anastasio/Bloomberg, „KI-Chatbot Grok sorgt auf X für massenhafte Nacktbilder ohne Zustimmung der Betroffenen“.
- 22 Ausführlich Steinebach, KI-generierte Abbildungen von Kindesmissbrauch, DuD 2024, 304.
- 23 Home Security Heroes, State of Deepfakes, 2023; zum perspektivischen Nutzungsverhalten s. Ernst/Heise, „Jim Keller über KI: <In 10 Jahren ist alle Software weg>“, 2024.
- 24 Karaboga, in: Jaki/Steiger, „Die Regulierung von Deepfakes auf EU-Ebene“, Digitale Hate Speech, 2023, S. 204 ff.; Pawelec/Bieß, Deepfakes, 2021, S. 44.
- 25 Kumkar/Griesel, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (118); Molavi Vasse'i, Watermarking von KI-generierten Inhalten als regulatorisches Instrument, RD 2024, 406 (406, Rn. 4); Kumkar/Rapp, Deepfakes, ZfDR 2022, 199 (201); Kumkar, Deepfakes – Risiken und Regulierung im europäischen Verordnungsentwurf für künstliche Intelligenz, K&R-Beil. zu Heft 10/2023, 32 (34).
- 26 Zur Haftung für KI-generierte geschäftsschädigende Falschinformationen s. LG Kiel, Urteil vom 29.2.2024 – 6 O 151/23; zu Lizenzgebühren für Synchronsprecher s. LG Berlin II Urteil vom 20.8.2025 – 2 O 202/24; zu vermögenswerten Bestandteilen des

Unternehmen durch synthetische Inhalte geschädigt²⁷ oder Vermögenswerte durch täuschungsbasierte Kommunikationsformen gefährdet werden.

3. Technik- oder wirkungsbezogenes Verständnis von Transparenz

Als Reaktion auf die identifizierten Herausforderungen haben sowohl die nationalen als auch der europäische Gesetzgeber ihre regulatorischen Aktivitäten zuletzt deutlich ausgeweitet und einen umfassenden Regulierungsrahmen geschaffen.²⁸ Entsprechende Transparenzmechanismen versprechen zwar grundsätzlich einen von der individuellen technischen Kompetenz unabhängigen Einfluss auf die Glaubwürdigkeit und Genauigkeit von Inhalten.²⁹ Dabei ergeben sich jedoch Unterschiede je nach gewähltem Transparenzansatz.³⁰

Kennzeichnungen von Medieninhalten könnten zunächst an deren Entstehungsprozess anknüpfen und damit das Ziel verfolgen, sämtliche KI-generierten Ausgaben als solche – also unter Nutzung einer bestimmten Technologie erstellten Inhalte – zu deklarieren. Die Kennzeichnung steht damit unabhängig vom konkreten Inhalt oder den mit dessen Generierung und Verbreitung beabsichtigten Wirkungen. Dies gilt vor allem für die maßgeblichen Transparenzvorschriften der KI-VO, die dem (system- und produktbezogenen) Produktsicherheitsrecht zuzuordnen sind. Zwar ist keiner der dortigen Regelungen ein eindeutig technikbezogenes Transparenzverständnis zu entnehmen; Ansätze hierfür finden sich jedoch an einigen Stellen.³¹

Diesem technikbezogenen Verständnis von Transparenz steht indes die mit der Einführung von Kennzeichnungs- und Offenlegungspflichten häufig verbundene gesetzgeberische Motivation entgegen.³² Anlass für regulatorisches Tätigwerden war weniger die bloße Tatsache maschineller Generierung als vielmehr das Risiko, dass bestimmte Inhalte geeignet sind, Rezi-

Persönlichkeitsrechts s. Schertz, Postmortales Persönlichkeitsrecht in Zeiten von KI und Deepfakes, NJW 2025, 2801 (2803 f.).

27 LG Kiel, Urteil vom 29.2.2024 – 6 O 151/23.

28 Art. 50 Abs. 2, 4 KI-VO; Art. 35 Abs. 1 S. 2 lit. k DSA; Art. 18 Abs. 1 lit. e EMFA.

29 Kleine, „Perception of Deepfake Technology“, 2022.

30 Epstein u. a., „What label should be applied to content produced by generative AI?“, 2023, S. 2.

31 S. unten unter 4.I.

32 S. dazu im Einzelnen unten unter 4.

ipientinnen und Rezipienten in die Irre zu führen und dadurch individuelle oder gesellschaftliche Schutzgüter zu beeinträchtigen. Diesen Schutz bieten Kennzeichnungs- und Offenlegungspflichten, die stärker wirkungsbezogen ausgestaltet sind. Hier tritt der Entstehungsprozess in den Hintergrund; entscheidend ist vielmehr Inhalte zu identifizieren, die ein besonderes Täuschungs- oder Manipulationspotenzial aufweisen. Das Problem künstlich generierter Inhalte ist damit besser im Kontext bekannter Erscheinungsformen missbräuchlicher Inhalte wie Fehl-, Mis- und Desinformation zu verorten, das durch den hohen Grad an Authentizität synthetischer Medieninhalte nur verstärkt wird.³³ Wirkungsbezogene Transparenzpflichten zielen damit *primär* auf den Schutz übergeordneter öffentlicher Interessen wie Authentizität, Richtigkeit und Verlässlichkeit von Medieninhalten. Die Vertrauenswürdigkeit einer Information soll dabei von ihrer nachvollziehbaren Erstellung bis zu ihrer Zirkulation, Nutzung und Bearbeitung über Plattform- und andere Systeminfrastrukturgrenzen hinweg gewährleistet sein. Methoden und Wirkmechanismen dieser Interventionsstrategie gegen den Verlust des Vertrauens in das Informationsökosystem wurden schließlich auch vielfach empirisch untersucht.³⁴

-
- 33 Zu den Schutzgütern der im DSA vorgegebenen Anti-Desinformationsmaßnahmen s. *Peukert*, Desinformationsregulierung in der EU - Überblick und offene Fragen, JZ 2023, 278 (281) mwNachw.; *Gerdemann/Spindler*, Das Gesetz über digitale Dienste (Digital Services Act) (Teil 2), GRUR 2023, 115 (123). Dementsprechend verfolgt der DSA grundsätzlich auch einen technikneutralen Ansatz, sodass auch ein ohne KI erzeugter, etwa durch irreführende Schnitte, falsche Kontextualisierung oder manipulative Beschriftung verfälschter Medieninhalt, der eine vergleichbare Gefährdungslage begründet wie ein vollständig synthetisch generiertes Medium, einer Kennzeichnungspflicht unterliegt; Umgekehrt die eine Kennzeichnungspflicht nicht exklusiv der Prävention und Kontextualisierung solcher Medieninhalte, sondern erfüllt daneben noch weitere Funktionen, vgl. *Becker*, Generative KI und Deepfakes in der KI-VO, CR 2024, 353 (360, Rn. 60); *Becker*, Das Urheberrecht als Trostpreis für den Menschen?, GRUR 2024, 505 (507).
- 34 Zur Terminologie von KI-Labels s. *Epstein u. a.*, „What label should be applied to content produced by generative AI?“, 2023; *Rauchfleisch u. a.*, „Deepfakes or Synthetic Media? The Effect of Euphemisms for Labeling Technology on Risk and Benefit Perceptions“, 2025; *Altay/Gilardi*, People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation, 2024; Zu KI-Labels in einzelnen Branchen s. *Toff/Simons*, „Or They Could Just Not Use It?\": The Dilemma of AI Disclosure for Audience Trust in News, The International Journal of Press/Politics 2025, 881-903; *Li/Yang*, Impact of Artificial Intelligence-Generated Content Labels On Perceived Accuracy, Message Credibility, and Sharing Intentions for Misinformation, 2024; *Wittenberg u. a.*, Labeling AI-generated media online, 2025; *Gamage u. a.*, Labeling Synthetic Content: User Perceptions of Warning Label

Der Anteil synthetisch generierter Inhalte im Internet ist bereits jetzt erheblich³⁵ und wird auch zukünftig weiter zunehmen. Bereits angesichts der enormen Menge an Medieninhalten stellt sich die Frage, wie eine grundsätzlich wünschenswerte wirkungsbezogene Transparenzkennzeichnung in der Praxis effektiv umgesetzt werden kann. Vor diesem Hintergrund ist auch in Betracht zu ziehen, authentische bzw. menschlich-erstellte Inhalte ebenfalls einer Kennzeichnungspflicht zu unterziehen³⁶ – zumindest soweit es sich um Medieninhalte handelt, an deren menschlicher Erstellung ein besonderes Interesse³⁷ besteht. Dies hätte den positiven Effekt, dass diese Inhalte leichter auffindbar und identifizierbar sind und eine höhere Wahrnehmung von Vertrauenswürdigkeit und Werthaltigkeit erfahren können.³⁸

4. Europäische Kennzeichnungs- und Offenlegungspflichten im Überblick

Die Regulierung der Nutzung generativer KI-Systeme wird in unterschiedlichen Rechtsgebieten³⁹ nachvollzogen: Während etwa in Dänemark die Einführung eines Urheberrechts an der individuellen Erscheinung einer

Designs for AI-generated Content on Social Media, 2025; Lockhart/Tessono, „Human or AI? Evaluating Labels on AI-Generated Social Media Content, 2025; zur Wirkung von ampelfarbenbasierten Warnhinweisen s. Dobber u. a., The effect of traffic light veracity labels on perceptions of political advertising source and message credibility on social media, 2025.

35 Spennemann, „Delving into: the quantification of Ai-generated content on the internet“, 2025.

36 Newmann/Reuters Institute, „Journalism and Technology Trends and Predictions 2026“, 2026.

37 Ähnlich Becker, Generative KI und Deepfakes in der KI-VO, CR 2024, 353 (360 Rn. 61, 364 ff. Rn. 96 ff.).

38 Campante u. a., GenAI Misinformation, Trust, And News Consumption: Evidence From A Field Experiment, 2025.

39 Zum Produktsicherheitsrecht: Kumkar/Griesel, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117; Molavi Vasse'i, Watermarking von KI-generierten Inhalten als regulatorisches Instrument, RD i 2024, 406; Zum Strafrecht: Pfeiffer, Tatbestandsmerkmale, Rechtfertigungsgründe und Herausforderungen beim strafrechtlichen Schutz vor Deepfakes, MMR 2025, 934; Blocher, Strafbare Deepfakes als Täuschung über die Authentizität von Medieninhalten, KIR 2025, 225; Zum Deliktsrecht: Kumkar/Rapp, Deepfakes, ZfDR 2022, 199 (204 ff.); Lantwin, Deep Fakes – Düstere Zeiten für den Persönlichkeitsschutz, MMR 2019, 574.

natürlichen Person erwogen wird⁴⁰, hat Italien nach einer Reihe öffentlichkeitswirksamer pornografischer Deepfakes über Personen des öffentlichen Lebens sein Strafrecht verschärft. In Deutschland steht die Verschärfung strafrechtlicher Sanktionsmechanismen nach einer Bundesrats-Initiative des Freistaates Bayern⁴¹ aktuell genauso zur Debatte wie Verbesserungen in der Durchsetzung zivilrechtlicher Ansprüche auf Beseitigung, Unterlassung und Schadensersatz im Gesetzentwurf gegen digitale Gewalt.

Auf unionsrechtlicher Ebene stellen schließlich die im öffentlichen Produktsicherheits-, Medien- und Plattformrecht zu verortenden Transparenzpflichten für unterschiedliche Ausgaben von KI-Systemen einen nicht nur mitgliedsstaats-, sondern auch system- und inhaltsübergreifenden Ansatz dar. Die Transparenzpflichten knüpfen an unterschiedlichen Zeitpunkten im Lebenszyklus eines generativen KI-Systems bzw. eines synthetisch erzeugten Medieninhaltes an: Beginnend mit der Herstellung des KI-Systems über die Herstellung des synthetischen Medieninhalts bis zu dessen Zirkulation auf Online-Plattformen stellen die entsprechenden Vorschriften einen umfassenden Regulierungsrahmen dar.

4.1 Verordnung über künstliche Intelligenz (KI-VO)

Die Regelungen in Art. 50 KI-VO dienen dabei gewissermaßen als rechtsaktübergreifender Ausgangspunkt. In Art. 1 Abs. 1 und Erwägungsgrund 1 KI-VO werden mit dem Schutz vor schädlichen Auswirkungen von KI-Systemen in der EU und der Entwicklung von menschenzentrierten und vertrauenswürdigen KI-Systemen zentrale Regulierungsgründe benannt.

40 Der dänische Vorschlag scheint aus deutscher Perspektive besser i.R.d. nationalen Persönlichkeitsrecht dogmatik verortbar. Der dänische (Um-)Weg über das Urheberrecht wirft initial schon die Frage auf, wie der beabsichtigte Schutz „realistic digitally generated imitations of personal characteristics“ überzeugend unter den urheberrechtlichen Werkbegriff einer persönlichen geistigen Schöpfung eingeordnet werden kann. Vgl. dazu *Karttunen*, European Parliamentary Research Service, „The Danish approach to copyright and deepfakes: A model for the EU?“, 2026.

41 BT-DruckS. 222/24; dazu der Antrag der FDP-Fraktion im Landtag Nordrhein-Westfalen, LT NRW DruckS. 18/10528, wo zusätzlich detaillierte Vorschriften zur Kennzeichnung von Deepfakes verlangt werden.

Spezifischer sieht Art. 50 Abs. 2 KI-VO vor, dass die Ausgaben eines KI-Systems in einem maschinenlesbaren⁴² Format⁴³ „als künstlich erzeugt oder manipuliert“ gekennzeichnet werden müssen.⁴⁴ In ähnlicher Weise sehen auch die Transparenzpflichten für die Generierung von Deepfakes in Art. 50 Abs. 4 KI-VO und für die Zirkulierung synthetischer Medieninhalte in Art. 35 Abs. 1 S. 2 lit. k DSA vor, dass Inhalte als künstlich erzeugt oder manipuliert offengelegt werden müssen. Zumindest der KI-VO geht es dabei explizit um die Differenzierung zwischen von Menschen erzeugten und synthetisch generierten Inhalten.⁴⁵

Der Anwendungsbereich von Art. 50 Abs. 4 UAbs. 1 KI-VO wird nicht allein aufgrund der künstlichen Erzeugung eines Inhalts eröffnet, sondern setzt zusätzlich voraus, dass der Medieninhalt eine wirklichkeitsverzerrende oder wirklichkeitsimitierende Wirkung entfaltet – mithin einen Deepfake⁴⁶ darstellt. Art. 50 Abs. 4 UAbs. 2 KI-VO stellt darüber hinaus auf den Kontext der Veröffentlichung ab, indem er insbesondere Textinhalte erfasst, die zur Information der Öffentlichkeit über Angelegenheiten von öffentlichem Interesse verbreitet werden.

Weitergehende kontextvermittelnde Anforderungen an die Kennzeichnung sind den Vorschriften zunächst nicht zu entnehmen; insbesondere ist weder Art. 50 Abs. 4 KI-VO noch Art. 50 Abs. 2 KI-VO die Pflicht zur Angabe des Namens des Betreibers oder Anbieters eines KI-Systems zu entnehmen.⁴⁷ Die Entscheidung des Unionsgesetzgebers gegen eine entsprechend weite Verpflichtung wurde bereits in Zusammenhang mit der Generierung und Verbreitung persönlichkeitsrechtsverletzender Deepfakes kritisiert.⁴⁸

42 Zum Begriff s. *Brorsen/Falk*, Die „Maschinenlesbarkeit“ von Informationen, MMR 2025, 7ff.

43 Zur Frage ob auch ein menschenlesbares Format erforderlich ist, *Kumkar/Griesel*, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (124).

44 *Kumkar/Griesel*, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (122 ff.).

45 Erwägungsgrund 133 S. 1 KI-VO; Vergleichbare Erwägungen fehlen im DSA, da sich kein spezifischer Erwägungsgrund mit Art. 35 Abs. 1 S. 2 lit. k DSA beschäftigt.

46 Zum Begriff s. Art. 3 Nr. 60 KI-VO; *Meding/Sorge*, What constitutes a deepfake?, CSLAW '25: Symposium on Computer Science and Law, 2025.

47 *Kumkar/Griesel*, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (122); Zur entsprechenden Kritik am Kommissionsentwurf, *Hinderks*, Die Kennzeichnungspflicht von Deepfakes, ZUM 2022, 110 (113).

48 *Kumkar/Griesel*, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (122); *Hinderks*, Die Kennzeichnungspflicht von Deepfakes, ZUM 2022, 110 (113).

Doch auch jenseits dieser spezifischen Inhalte-Kategorie hätten entsprechende kontextvermittelnde Informationen belastbare Anknüpfungspunkte zur Einordnung und Kontextualisierung von Medieninhalten geliefert.

Das wirkungsbezogene Verständnis wird jedoch durch die Erwägungsgründe der KI-VO weiter untermauert. Die Kennzeichnungs- und Offenlegungspflichten stehen etwa in – wirkungsbezogenen – Erwägungen zentraler Risiken wie Identitätsbetrug, Täuschung und anderen schädlichen Auswirkungen auf Rezipienten⁴⁹ – nicht aber technikbezogenen Risiken und Gefahren auf Systemebene⁵⁰. Zugleich sind die Integrität des Informationsökosystems und das hierin gesetzte Vertrauen angesichts von Fehlinformation, Manipulation in großem Maßstab, Betrug und Täuschungshandlungen erheblichen Gefahren ausgesetzt.⁵¹ Auch unter Berücksichtigung der Anforderungen an die einzuführenden Abhilfemechanismen, wofür als Anknüpfungspunkte u. a. die Herkunft oder Echtheit von Inhalten benannt werden, scheint ein wirkungsbezogenes Verständnis naheliegender.⁵²

Die sekundärrechtlichen Vorgaben der KI-VO wurden zuletzt i. R. d. Entwurfs für einen Verhaltenskodex über die Transparenz KI-generierter Inhalte⁵³ (CoP-E) aufgegriffen. Die Vereinbarung konkreter Kennzeichnungs- und Offenlegungsmechanismen steht dabei erneut in Erwägung des engen Zusammenhangs zwischen synthetischer Generierung von Medieninhalten und den damit verbundenen Folgen für die Integrität und das Vertrauen in das Informationsökosystem.⁵⁴ Vorgesehen ist die Bereitstellung strukturierter, automatisiert auslesbarer Informationen über Herkunft, Entstehungs- und Bearbeitungsprozesse eines Inhalts. Transparenz wird dabei nicht allein als Information für Endnutzerinnen und -nutzer verstanden, sondern zugleich als infrastrukturelle Voraussetzung für die Verbreitung oder Nutzung von Medieninhalten auf Plattformen, Suchmaschinen und durch andere Intermediäre – etwa zur Bewertung der Vertrauenswürdigkeit von Inhalten oder zur Unterstützung weiterer Risikominderungsmaßnahmen. Ergänzend verweist der CoP-E auf den Einsatz standardisierter Pro-

49 ErwGr. 132 S. 1 KI-VO.

50 Im risikobasierten Ansatz der KI-VO sind entsprechende Compliance-Maßnahmen nur für Hochrisiko-KI-Systeme vorgesehen, vgl. Kapitel III Abschnitt 2 KI-VO.

51 ErwGr. 133 S. 2 KI-VO.

52 ErwGr. 133 S. 3 KI-VO; ErwGr. 135 KI-VO.

53 Second Draft Code of Practice on Transparency of AI-Generated Content, 03.03.2026.

54 Second Draft Code of Practice on Transparency of AI-Generated Content, 03.03.2026, S. 7.

venienzsysteme („Content Credentials“)⁵⁵, die solche Informationen in interoperablen Formaten bündeln, eine plattformübergreifende Nachvollziehbarkeit der Erstellung und Bearbeitung digitaler Inhalte ermöglichen und regelmäßig durch kryptografische Sicherungsmechanismen gegen Manipulation geschützt sind. Durch diese Kombination aus Maschinenlesbarkeit, Standardisierung und technischer Absicherung sollen Provenienzinformationen dauerhaft mit dem jeweiligen Inhalt verknüpft, überprüfbar gehalten und damit als funktionales Instrument zur Stärkung der Integrität des Informationsökosystems nutzbar gemacht werden.

4.2 Gesetz über digitale Dienste (DSA)

Darüber hinaus sind wirkungsbezogene Transparenzpflichten als Katalogmaßnahme zur Minderung systemischer Risiken auf „sehr großen Online-Plattformen und sehr großen Online-Suchmaschinen“ (VLOPSEs)⁵⁶ in Art. 35 Abs. 1 S. 2 lit. k DSA enthalten. Die Kennzeichnung kann als Risikominderungsmaßnahme ergriffen werden, nachdem i. R. e. plattformseitig durchzuführenden Risikobewertung ein systemisches Risiko i. S. v. Art. 34 DSA auf der Plattform festgestellt wurde. Zu den erfassten Risikokategorien⁵⁷ zählen die Verbreitung rechtswidriger Inhalte, eine nachteilige Auswirkung auf demokratische Prozesse, gesellschaftliche Debatten und Wahlprozesse sowie geschlechtsspezifische Gewalt.⁵⁸ Weitergehende besondere Gefährdungslagen können sich zudem aus den Funktionsweisen einer Plattform ergeben – etwa durch Nutzung von Reichweitenmechanismen, algorithmischer Verstärkung oder strukturellen Anreizen zur massenhaften Verbreitung rechtswidriger Inhalte.⁵⁹ Doch auch jenseits dieser Kategorien systemischer Risiken zielt der DSA auf eine wirkungsbezogene Regulierung, die in allgemeinen Schutzzweckbestimmungen wie einem sicheren,

55 S. im Einzelnen unten unter 5.

56 Very Large Online Platforms and Search Engines i.S.v. Art. 34 DSA.

57 Z. T. werden diese Bereiche auch als Regelbeispiele verstanden, vgl. v. *Lewinski/Kling*, Compliance-Pflichten nach dem DSA – ein umfassender Überblick, RD 2023, 183 (190).

58 Art. 34 Abs. 1 DSA; *Peukert*, Desinformationsregulierung in der EU - Überblick und offene Fragen, JZ 2023, 278 (281) erklärt daher, dass das vom DSA vorgegebene „Desinformationsregime nicht dem Schutz individueller Rechtsgüter bestimmter Einzelpersonen dient, sondern auf einer abstrakteren Ebene dem Schutz öffentlicher Güter/Interessen“.

59 ErwGr. 79 S. 1 DSA; ErwGr. 84 S. 3 DSA.

vorhersehbaren, berechenbaren und vertrauenswürdigen Online-Umfeld zum Ausdruck kommt.⁶⁰

Dies spiegelt sich auch in den DSA-spezifischen Leitlinien⁶¹ und einem Verhaltenskodex⁶² wider. Erneut ist der Anwendungsbereich der Regelungen nicht nur allgemein in Bezug auf KI-generierte Inhalte eröffnet; vielmehr bedarf es (zusätzlich) eines täuschenden, falschen oder irreführenden Charakters dieser Inhalte.⁶³ Maßnahmen zur Reduzierung damit verbundener Risiken bei Wahlen⁶⁴ und Desinformationen⁶⁵ beinhalten u. a. die Einführung von Faktenprüfhinweisen, Vertrauenssiegeln und Werkzeugen oder anderen Informationen zur Bewertung der Provenienz und Editierhistorie, um die Vertrauenswürdigkeit, Integrität und Echtheit digitaler Inhalte effektiv bewerten zu können.⁶⁶ Kennzeichnungspflichten dienen in diesem Rahmen weniger der Offenlegung des technischen Entstehungsprozesses als vielmehr der Befähigung von Nutzerinnen und Nutzern, die Vertrauenswürdigkeit und Aussagekraft von Medieninhalten eigenständig zu bewerten.⁶⁷

Für die Zirkulation synthetischer Medieninhalte bestehen zudem mit Art. 35 DSA auf der regulierungsintensivsten Ebene der abgestuften Pflicht-

60 Art. 1 Abs. 1 DSA; ErwGr. 3, 9 DSA.

61 Leitlinien der Kommission für Anbieter sehr großer Online-Plattformen und sehr großer Online-Suchmaschinen zur Minderung systemischer Risiken in Wahlprozessen gemäß Artikel 35 Absatz 3 der Verordnung (EU) 2022/2065, C/2024/3014.

62 Zur Vereinbarkeit mit den Anforderungen an Verhaltenskodizes nach Art. 45 DSA s. Commission Opinion of 13.2.2025 on the assessment of the Code of Practice on Disinformation within the meaning of Article 45 of Regulation 2022/2065, C(2025) 1008 final; *Kollmann*, Zur Bedeutung des CoP on Disinformation 2022 im Kontext von KI-generierten Inhalten, MMR-Aktuell 2023, 458365.

63 Leitlinien der Kommission für Anbieter sehr großer Online-Plattformen und sehr großer Online-Suchmaschinen zur Minderung systemischer Risiken in Wahlprozessen gemäß Artikel 35 Absatz 3 der Verordnung (EU) 2022/2065, C/2024/3014, S. 1, 6, 8, 9; Code of Conduct on Disinformation 2022, S. 4.

64 Leitlinien der Kommission für Anbieter sehr großer Online-Plattformen und sehr großer Online-Suchmaschinen zur Minderung systemischer Risiken in Wahlprozessen gemäß Artikel 35 Absatz 3 der Verordnung (EU) 2022/2065, C/2024/3014, S. 7.

65 Commitment 20 Code of Conduct on Disinformation 2022 benennt beispielhaft den Industriestandard C2PA als taugliche Umsetzungsmaßnahme; dazu *Brorsen/Falk*, Neue Compliance-Pflichten nach dem Digital Services Act, MMR 2024, 32 (35 f.).

66 S. Commitments 21, 22 Code of Conduct on Disinformation 2022.

67 S. dazu auch die vorgeschlagenen Initiativen zur Förderung der Medienkompetenz, vgl. Leitlinien der Kommission für Anbieter sehr großer Online-Plattformen und sehr großer Online-Suchmaschinen zur Minderung systemischer Risiken in Wahlprozessen gemäß Artikel 35 Absatz 3 der Verordnung (EU) 2022/2065, C/2024/3014, S. 6.

tenstruktur des DSA verschärfte Sorgfalts- und Transparenzpflichten für VLOPs und VLSEs zur Bekämpfung systemischer Risiken. Dazu gehört nach Art. 35 Abs. 1 S. 2 lit. k DSA auch eine plattformseitig durchzuführende Kennzeichnung synthetisch erzeugter Medieninhalte. Ergänzend ist den Nutzern der Plattform eine benutzerfreundliche Funktion zur Verfügung zu stellen, entsprechende Inhalte zu melden.

4.3 Europäisches Medienfreiheitsgesetz (EMFA)

Mittelbar entfalten die Vorgaben der KI-VO und des DSA auch Wirkung im EMFA, wo zugunsten von Mediendiensteanbietern im Sinn von Art. 2 Nr. 2 EMFA ein Medienprivileg für die Verbreitung ihrer Inhalte auf VLOPSEs vorgesehen ist: Nach Art. 18 Abs. 1 lit. e EMFA sollen sie eine Erklärung („Selbstdeklaration“⁶⁸) über die Art und Weise der eigenen Inhalteerstellung abgeben.⁶⁹ Bezugspunkt dieser Erklärung ist nach dem Wortlaut der Verordnung die Frage, ob der Mediendiensteanbieter Inhalte bereitstellt, die von Systemen künstlicher Intelligenz erzeugt wurden, ohne dass diese einer menschlichen Überprüfung und redaktionellen Kontrolle unterzogen worden sind.⁷⁰ An diese Erklärung knüpft der EMFA schließlich Erleichterungen für entsprechend notifizierte Mediendiensteanbieter im Melde- und Abhilfeverfahren nach Art. 16 DSA – insbesondere indem den Entscheidungen über die Löschung oder Beschränkung der Sichtbarkeit eines Medieninhalts zwingend ein Stellungnahmeverfahren vorzuschalten ist – sogenannte „Must-carry-Verpflichtung“.⁷¹ In Bezug auf die durchzuführende Notifizierung dürfen sich VLOPSEs zudem Zertifizierungen Dritter über die Einhaltung der Anforderungen aus Art. 18 EMFA bedienen. Mit der *Journalism Trust Initiative* wird dabei explizit eine Stelle benannt, die mit CWA 17493:2019 (E) einen maschinenlesbaren und neutralen Standard für die globale Zertifizierung vertrauenswürdiger Medien geschaffen hat. Unter „Section B: Professionalism and Accountability, Nr. 10.6 Automatically Generated Content“ des Standards werden die Adressaten der standardi-

68 Achenleiter, Das Medienprivileg im European Media Freedom Act, MMR 2025, 89 (91).

69 Art. 18 Abs. 1 lit. e EMFA.

70 Art. 18 Abs. 1 lit. e EMFA.

71 Achenleiter, Das Medienprivileg im European Media Freedom Act, MMR 2025, 89 (92 f.); Hennemann, Der European Media Freedom Act oder: Von Zielen und Mitteln des Medienrechts, ZEuP 2024, 498 (501).

sierten Norm verpflichtet, ihre Nachrichteninhalte, die ganz oder teilweise automatisch durch algorithmische Verfahren generiert werden, eindeutig als solche zu kennzeichnen. Eine allein am Wortlaut orientierte Auslegung würde auch insoweit für eine rein am Entstehungsprozess anknüpfende Kennzeichnung sprechen.

4.4 Persönlichkeitsrechtliche Erwägungen

Die genannten Transparenzpflichten sind schließlich für solche Medieninhalte, in denen individuelle Rechtspositionen – wie das allgemeine Persönlichkeitsrecht einschließlich seiner speziellen Ausprägungen im Recht am eigenen Bild oder dem Recht an der eigenen Stimme – betroffen sind, nur von begrenzter Bedeutung. Der durch Transparenz über die Inhalteerstellung vermittelbare Schutz scheint hier gegenüber den zuvor genannten Informationen mit meinungsbildendem Charakter oder von wirtschaftlicher Bedeutung deutlich beschränkter⁷² – immerhin verhindern Kennzeichnungen originär keinen digitalen Missbrauch von – vornehmlich – Frauen und Kindern jenseits des Körperlichen⁷³; dennoch ermöglichen sie die Einordnung eines ansonsten wirklichkeitsgetreu erscheinenden Inhalts als wirklichkeitsverzerrend und dienen damit auch in diesen Kontexten als Gradmesser zwischen Wirklichkeit und Fiktion. Solche Erwägungen könnten zumindest zivilrechtlich – etwa in Bezug auf die Bestimmung der Eingriffsintensität einer Rechtsgutverletzung – Berücksichtigung finden.

72 Ähnlich in Bezug auf Schadensersatzansprüche *Gerdemann/Spindler*, Das Gesetz über digitale Dienste (Digital Services Act) (Teil 2), GRUR 2023, 115 (123); s. die empirischen Erkenntnisse zum Interaktionsverhalten von Rezipienten mit unterschiedlichen Inhalten, die als KI-generiert gekennzeichnet sind, vgl. *Game u. a.*, Labeling Synthetic Content: User Perceptions of Warning Label Designs for AI-generated Content on Social Media, 2025.

73 Zum Begriff und verbreiteten Phänomenen s. *Burghardt/Schmidt/Steinl*, "Sexuelle Selbstbestimmung jenseits des Körperlichen. Zur Einführung", in *Burghardt/Schmidt/Steinl*, Sexuelle Selbstbestimmung jenseits des Körperlichen, 2024, S. 1 ff.; *Labuz*, Regulating Deep Fakes in the Artificial Intelligence Act, ACIG 2023, 252 (271); Centre for Digital Governance, The false promise of transparent deep fakes: How transparency obligations in the draft AI Act fail to deal with the threat of disinformation and image-based sexual abuse, 2022; so bereits zum Kommissionsentwurf *AccessNow*, Access Now's submission to the European Commission's adoption consultation on the AI Act, 2021, S. 6.

5. Praktische Erwägungen zur Kennzeichnung und Offenlegung

Offenlegungen von synthetischen Inhalten können grundsätzlich nur dann stattfinden, wenn bekannt ist, dass der Inhalt mithilfe einer generativen künstlichen Intelligenz hergestellt oder manipuliert wurde. Diese Kenntnis hat vornehmlich der Betreiber eines KI-Systems und legt nach der Intention des Gesetzgebers seine synthetischen Inhalte im Einklang mit den Verpflichtungen aus Art. 50 Abs. 4 KI-VO offen. Mit großen praktischen Hürden ist allerdings die Offenlegung solcher Inhalte seitens Dritter (wie etwa sozialen Medien oder andere Plattformen) verbunden. Dies gilt insbesondere dann, wenn die Offenlegungen nicht existent (beispielsweise durch Entfernung von entsprechenden „Labels“) oder nicht mit dem entsprechenden Medium verbunden sind (bspw. fand die Offenlegung durch einen separaten textlichen Hinweis statt). In diesen Fällen sind zur Offenlegung verpflichtete Dritte entweder auf aufwendige und zugleich fehleranfällige Detektionsmechanismen angewiesen, die synthetische Inhalte mithilfe computerforensischer Verfahren identifizieren. Alternativ können sie auf die nach Art. 50 Abs. 2 KI-VO erforderliche maschinenlesbare Kennzeichnung zurückgreifen, die verlässlich, schnell und kostengünstig Auskunft über die synthetische Herstellung oder Manipulation eines Medieninhalts gibt.

In praktischer Hinsicht kann bei der Kennzeichnung von KI-Ausgaben zwischen zwei grundsätzlich unterschiedlichen Verfahren unterschieden werden: Der Kennzeichnung des Inhalts der Ausgabe und der Kennzeichnung des Datenträgers der Ausgabe.

Die *Kennzeichnung des Inhalts* der Ausgabe erfolgt beispielsweise durch die Einfügung von sogenannten Wasserzeichen⁷⁴. Dabei handelt es sich um für das menschliche Auge nicht erkennbare – aber dennoch für Maschinen auslesbare – technische Muster im jeweiligen Inhalt. Ein Beispiel für ein solches Verfahren ist Googles SynthID, bei dem digitale Spuren in synthetisch hergestellte Bilder eingefügt werden, die wiederum mit der passenden Software zwecks Verifikation ausgelesen werden können.⁷⁵ Eine andere Methode ist die Speicherung von sogenannten digitalen Fingerabdrücken. Bei diesem Verfahren wird eine Abstraktion des synthetischen Inhalts

74 Molavi Vasse'i, Watermarking von KI-generierten Inhalten als regulatorisches Instrument, RD 2024, 406 (408, Rn. 142 ff.) mwNachw.; s. auch Partnership on AI, Building a glossary for Synthetic Media Transparency Methods, Part 1: Indirect Disclosure.

75 <https://deepmind.google/models/synthid/>.

(bspw. ein Vektor oder ein Hash-Digest) hergestellt und in einer Datenbank samt entsprechenden Metadaten gespeichert. Bei zu verifizierenden Inhalten wird ebenfalls ein digitaler Fingerabdruck entnommen und mit dem Datenbestand an Fingerabdrücken verglichen. Ergibt der Vergleich einen Treffer, können die passenden Metadaten über den Inhalt offengelegt werden (Beispiel: „dieses Bild ist synthetisch“). Der Vorteil dieses Verfahrens ist die Schnelligkeit, die Robustheit gegenüber Veränderungen bei gekennzeichneten Inhalten sowie die Möglichkeit, sehr viele Informationen zum jeweiligen „fingerprinted“ Inhalt vorzuhalten.⁷⁶

Eine Methode zur *Kennzeichnung des Datenträgers* des Mediums ist das Verfahren nach dem Industriestandard C2PA.⁷⁷ Dabei werden Medien-Metadaten mittels eines kryptografischen Verfahrens mit den Mediendatenträgern verbunden.⁷⁸ Um das Verfahren der Kennzeichnung im C2PA-Standard möglichst sicher zu gestalten, können nur zertifizierte Unternehmen die kryptografischen Kennzeichnungen durchführen.⁷⁹ Auch diese dem Mediendatenträger eingefügten, standardisierten Metadaten können mittels Software ausgelesen werden. Vereinzelt soziale Medien lesen bereits die C2PA-Medien-Metadaten aus und legen diese ihren Nutzerinnen und Nutzern in Form von sog. Content Credentials offen.⁸⁰ Von dieser Offenlegung sind der Ursprung und die Authentizität des Medieninhalts und ggf. ein Hinweis auf die Herstellung oder Manipulation des Inhalts mit generativer künstlicher Intelligenz umfasst. Im gleichen Sinne könnten auch VLOSEs ihren Risikominderungspflichten gemäß Art. 35 Abs. 1 S. 2 lit. k DSA nachkommen, wenn sie die maschinenlesbaren C2PA-Kennzeichnungen auslesen und ihre Nutzerinnen und Nutzer auf den synthetischen Ursprung von ausgespielten Medien hinweisen.⁸¹ Dieses Ineinandergreifen von Art. 50 Abs. 2 KI-VO und Art. 35 Abs. 1 S. 2 lit. k DSA würde schließlich nach hier vertretener Auffassung auch die – eigentlich Betreiber-seitig zu

76 Beispielsweise nutzt die Applikation *Lumid* Fingerprinting-Technologien, um maschinen- und menschenlesbare Informationen zur Herkunft, zum Erstellungsdatum sowie zu urheberrechtlich relevanten Metadaten eines Medieninhalts bereitzustellen. (<https://lumid.tech>).

77 Coalition for Content Provenance and Authenticity, <https://c2pa.org/>.

78 S. https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA_Specification.html, Ziffer 9. Allerdings ist es auch möglich, C2PA-Metadaten mit Wasserzeichen- und Fingerprinting-Kennzeichnungen zu verbinden.

79 In Deutschland ist in etwa die Valid Technologies GmbH dazu befugt; s. <https://valid.tech>

80 So etwa LinkedIn, s. <https://www.linkedin.com/help/linkedin/answer/a6282984>.

81 S. o. 4.II.; Code of Conduct on Disinformation 2022, Commitment 20.

erfüllende – Offenlegungspflicht nach Art. 50 Abs. 4 KI-VO funktional ersetzen.⁸²

Zuletzt sei noch angemerkt, dass die Betreiber bzw. Anbieter von KI-Systemen – neben dem rechtlichen Zwang – auch eine technisch-wirtschaftliche Motivation zur Kennzeichnung von synthetischen Inhalten haben sollten. Denn das (versehentliche) Einspeisen von synthetischen Inhalten in die Trainingsdaten von KI-Systemen verringert die Leistungsfähigkeit des Systems bis hin zur Unbrauchbarkeit.⁸³

6. Fazit

Der Beitrag hat gezeigt, dass die zunehmende Verbreitung synthetischer Medien das Vertrauen in einzelne Medieninhalte und in digitale Informationsräume gleichermaßen unter Druck setzt und selbst medienkompetente Nutzerinnen und Nutzer vor erhebliche Verifikationsprobleme stellt. Transparenzpflichten stehen dabei im Zielkonflikt zwischen der schlichten Offenlegung der technischen Erstellung und der Minderung negativer Auswirkungen KI-generierter Ausgaben auf das Informationsökosystem. Entsprechend der identifizierten Schutzgüter ist ein wirkungsbezogener Ansatz zu präferieren. Mit Blick auf die Umsetzung der rechtlichen Vorgaben haben sich in der Praxis bereits einzelne Kennzeichnungsmechanismen etabliert.

Zugleich ist vor überhöhten Erwartungen an Transparenzpflichten zu warnen.⁸⁴ Transparenzmaßnahmen dienen primär der Bereitstellung von Kontextinformationen und fördern eine informierte Mediennutzung, stellen jedoch keinen eigenständigen Schutz vor der Erstellung oder dem (digitalen) Missbrauch mithilfe rechtswidriger Inhalte dar. Insbesondere bei gravierenden Rechtsgutverletzungen wie sexualisierter Gewalt stoßen Kennzeichnungen daher an ihre Grenzen.

82 So wohl auch Bomhard/Pieper/Wende/Merkle, 1. Aufl. 2025, KI-VO Art. 50 Rn. 150; allgemein zum Verhältnis der Transparenzpflichten aus KI-VO und DSA s. ErwGr. 136 KI-VO; Kumkar/Griesel, Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO, KIR 2024, 117 (125); Becker, Generative KI und Deepfakes in der KI-VO, CR 2024, 353 (363 Rn. 94).

83 Xing u. a., On the caveats of AI autophagy, Nature Machine Intelligence, 2025.

84 S. hierfür aktuell Clark/Lewandowsky, The continued influence of AI-generated deepfake videos despite transparency warnings, Communication Psychology, 2026.

Literatur

- Access Now (2021): Access Now's submission to the European Commission's adoption consultation on the Artificial Intelligence Act. URL: <https://www.accessnow.org/wp-content/uploads/2021/08/Submission-to-the-European-Commissions-Consultation-on-the-Artificial-Intelligence-Act.pdf> (besucht am 30.01.2026).
- Achenleiter, Ranja Andrea (2025): Das Medienprivileg im European Media Freedom Act. *Multimedia und Recht*, 02/2025, S. 89-93.
- Altay, Sacha und Gilardi, Fabrizio (2024): People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation. *PNAS Nexus*, 3 (10).
- Becker, Maximilian (2024): Generative KI und Deepfakes in der KI-VO. *Computer und Recht*, 06/2024, S. 353-366.
- Becker, Maximilian (2024): Das Urheberrecht als Trostpreis für den Menschen? *Gewerblicher Rechtsschutz und Urheberrecht*, 08/2024, S. 505-514.
- Bentzen, Naja (2025): Briefing: Information manipulation in the age of generative artificial intelligence. European Parliamentary Research Service. URL: [https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/779259/EPRS_BRI\(2025\)779259_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/779259/EPRS_BRI(2025)779259_EN.pdf) (besucht am 30.01.2026).
- Blocher, Janine (2025): Strafbare Deepfakes als Täuschung über die Authentizität von Medieninhalten. *Künstliche Intelligenz und Recht*, 06/2025, S. 225-230.
- Brokhues, Annabell (26. Januar 2026): EU-Kommission startet Verfahren gegen Chatbot Grok. *Deutschlandfunk*. URL: <https://www.deutschlandfunk.de/deepfakes-eu-kommission-startet-verfahren-gegen-chatbot-grok-100.html> (besucht am 30.01.2026).
- Brorsen, Hans und Falk, Richard (2024): Neue Compliance-Pflichten nach dem Digital Services Act. *Multimedia und Recht*, 01/2024, S. 32-36.
- Brorsen, Hans und Falk, Richard (2025): Die „Maschinenlesbarkeit“ von Informationen. *Multimedia und Recht*, 01/2025, S. 7-12.
- Burghardt, Boris; Schmidt, Anja und Steinl, Leonie (2024): Sexuelle Selbstbestimmung jenseits des Körperlichen. Zur Einführung. In: Boris Burghardt, Anja Schmidt und Leonie Steinl (Hrsg.): Sexuelle Selbstbestimmung jenseits des Körperlichen. Berlin: Mohr Siebeck. S. 1-14.
- Campante, Felipe R.; Durante, Ruben; Hagemeister, Felix und Sen, Ananya (2025): Genai Misinformation, Trust, And News Consumption: Evidence From A Field Experiment. National Bureau of Economic Research. Cambridge. URL: https://www.nber.org/system/files/working_papers/w34100/w34100.pdf (besucht am 30.01.2026).
- CWA17493 (2019): Journalism Trust Initiative. Cen Workshop Agreement. URL: <https://www.cencenelec.eu/media/CEN-CENELEC/CWAs/ICT/cwa17493.pdf> (besucht am 30.01.2026).
- Clark, Simon und Lewandowsky, Stephan (2026): The continued influence of AI-generated deepfake videos despite transparency warnings. *Communication Psychology* 4, article 13.

- Crone, Theresia (23. Januar 2026): Bildbasiert aber unsichtbar. *Verfassungsblog.de* URL: <https://verfassungsblog.de/pornografische-deepfakes/> (besucht am 30.01.2026).
- D'Anastasio, Cecilia (08. Januar 2026): KI-Chatbot Grok sorgt auf X für massenhafte Nacktbilder ohne Zustimmung der Betroffenen. *Bloomberg.com* URL: <https://www.bloomberg.com/news/articles/2026-01-08/x-gerat-wegen-grok-unter-druck-durch-flut-an-ki-nacktbildern> (besucht am 30.01.2026).
- Dobber, Tom; Kruikemeier, Sanne; Votta, Fabio und Helberger, Natali (2025): The effect of traffic light veracity labels on perceptions of political advertising source and message credibility on social media. *Journal of Information Technology & Politics*, Volume 22, Issue 1, S. 82-97.
- Ellenberg, Stefan (2024): Persönlichkeits- und Urheberrechte beim Voice Cloning. *Recht Digital* 12/2024, S. 599-607.
- Engel-Bunsas, Christoph (2025): Recht an der eigenen Stimme in Zeiten von Deepfakes. *Recht Digital*, 06/2025, S. 292-300.
- Epstein, Ziv; Fang, Cathy Mengying; Arechar, Antonio Alonso und Rand, David (2024): What label should be applied to content produced by generative AI?. An MIT Exploration of Generative AI.
- Ernst, Nico (21. April 2024): Jim Keller über KI: „In 10 Jahren ist alle Software weg“. *Heise.de*. URL: <https://www.heise.de/news/Jim-Keller-ueber-KI-In-10-Jahren-ist-alle-Software-weg-9692706.html> (besucht am 30.01.2026).
- Europäische Kommission (2024): Leitlinien der Kommission für Anbieter sehr großer Online-Plattformen und sehr großer Online-Suchmaschinen zur Minderung systemischer Risiken in Wahlprozessen gemäß Artikel 35 Absatz 3 der Verordnung (EU) 2022/2065. URL: https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:C_202403014 (besucht am 16.03.2026).
- Europäische Kommission (2025): Commission Opinion of 13.2.2025 on the assessment of the Code of Practice on Disinformation within the meaning of Article 45 of Regulation 2022/2065. URL: <https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation>
- Europäische Kommission (2025): First Draft Code of Practice on Transparency of AI-Generated Content. URL: <https://digital-strategy.ec.europa.eu/en/library/first-draft-code-practice-transparency-ai-generated-content> (besucht am 16.03.2026).
- Europäische Kommission (2026): Second Draft Code of Practice on Transparency of AI-Generated Content. URL: <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-second-draft-code-practice-marking-and-labelling-ai-generated-content> (besucht am 16.03.2026).
- Europäische Kommission (2022): The Strengthened Code of Practice on Disinformation 2022. URL: <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>
- Europäisches Parlament und Rat (2022): Verordnung (EU) 2022/2065 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste). URL: <https://eur-lex.europa.eu/eli/reg/2022/2065/oj> (besucht am 16.03.2026)

- Europäisches Parlament und Rat (2024): Verordnung (EU) 2024/1083 zur Schaffung eines gemeinsamen Rahmens für Mediendienste im Binnenmarkt und zur Änderung der Richtlinie 2010/13/EU (Europäisches Medienfreiheitsgesetz). URL: https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:L_202401083 (besucht am 16.03.2026).
- Europäisches Parlament und Rat (2024): Verordnung (EU) 2024/1689 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz) (Text von Bedeutung für den EWR). URL: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj?locale=de> (besucht am 16.03.2026).
- Game, Dilrukshi; Sewwandi, Dilki; Bandara, Arosha und Zhang, Min (Mai 2025): Labeling Synthetic Content: User Perceptions of Warning Label Designs for AI-generated Content on Social Media. *Conference on Human Factors in Computing Systems CHI* 25. URL: <https://arxiv.org/pdf/2503.05711> (besucht am 30.01.2026).
- Gerdemann, Simon und Spindler, Gerald (2023): Das Gesetz über digitale Dienste (Digital Services Act) (Teil 2). *Gewerblicher Rechtsschutz und Urheberrecht*, 03/2023, S. 115-125.
- Hartmann, Frank (2020): Der persönlichkeitsrechtliche Schutz vor Deepfakes. *Kommunikation und Recht*, 05/2020, S. 350-356.
- Hennemann, Moritz (2024): Der European Media Freedom Act oder: Von Zielen und Mitteln des Medienrechts. *Zeitschrift für Europäisches Privatrecht*, 03/2024, S. 498-504.
- Hinderks, Tobias (2022): Die Kennzeichnungspflicht von Deepfakes. *Zeitschrift für Urheber- und Medienrecht*, 02/2022, S. 110-119.
- Home Security Heroes (2023): State of Deepfakes. URL: <https://www.securityhero.io/state-of-deepfakes/> (besucht am 30.01.2026).
- van Huijstee, Mariette; van Boheemen, Pieter; Das, Djurre; Nierling, Linda; Jahn, Jutta; Karaboga, Murat; Fatun, Martin; Kool, Linda und Gerritsen, Jost (2021): Tackling deepfakes in European policy. European Parliamentary Research Service. URL: [https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU\(2021\)690039_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/690039/EPRS_STU(2021)690039_EN.pdf) (besucht am 30.01.2026).
- Karaboga, Murat (2023): Die Regulierung von Deepfakes auf EU-Ebene. In: Sylvia Jaki und Stefan Steiger (Hrsg.): *Digitale Hate Speech*. Berlin: J.B. Metzler, S. 197-220.
- Karttunen, Sofia (2026): The Danish approach to copyright and deepfakes: A model for the EU? *European Parliamentary Research Service*. URL: [https://www.europarl.europa.eu/RegData/etudes/ATAG/2026/782611/EPRS_ATA\(2026\)782611_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2026/782611/EPRS_ATA(2026)782611_EN.pdf) (besucht am 30.01.2026).
- Kern, Michael (2024): Die Verwendung von Social Bots: Transparenzpflichten gemäß Medienstaatsvertrag und KI-VO sowie deren Umsetzung in virtuellen Welten. *Künstliche Intelligenz und Recht*, 03/2024, S. 94-99.
- Klass, Nadine (2025): Manipulierte Realität: Wie Deepfakes Recht und Gesellschaft herausfordern. *Zeitschrift für Urheber- und Medienrecht*, 07/2025, S. 485-492.

- Kleine, Felix (2022): Perception of Deepfake Technology - The Influence of the Recipients' Affinity for Technology on the Perception of Deepfakes. Masterarbeit. Hochschule Darmstadt. URL: https://www.researchgate.net/publication/364254498_Perception_of_Deepfake_Technology_-_The_Influence_of_the_Recipients'_Affinity_for_Technology_on_the_Perception_of_Deepfakes (besucht am 30.01.2026).
- Kollmann, Katharina (2023): EU-Kommission: Unterzeichner des Verhaltenskodexes gegen Desinformation sollen KI-Nutzung klar kennzeichnen. *Multimedia und Recht-Aktuell*, 11/2023, 458365.
- Kumkar, Lea Katharina (2023): Deepfakes – Risiken und Regulierung im europäischen Verordnungsentwurf für künstliche Intelligenz. *Kommunikation und Recht*, Beilage zu 10/2023, S. 32-37.
- Kumkar, Lea Katharina und Griesel, Moritz (2024): Transparenzpflichten für Deepfakes und synthetische Medieninhalte in der KI-VO. *Künstliche Intelligenz und Recht*, 04/2024, S. 117-126.
- Kumkar, Lea Katharina und Rapp, Julian Philipp (2022): Deepfakes. *Zeitschrift für Digitalisierung und Recht*, 03/2022, S. 199-228.
- Łabuz, Mateusz (2023): Regulating Deep Fakes in the Artificial Intelligence Act. *Applied Cybersecurity & Internet Governance* 2. 1-42., S. 252-291.
- Lantwin, Tobias (2019): Deep Fakes – Düstere Zeiten für den Persönlichkeitsschutz? *Multimedia und Recht*, 09/2019, S. 574-578.
- v. Lewinski, Kai und Klink, Quentin (2023): Compliance-Pflichten nach dem DSA – ein umfassender Überblick. *Recht Digital*, 04/2023, S. 183-194.
- Li, Fan und Yang, Ya (2024): Impact of Artificial Intelligence-Generated Content Labels On Perceived Accuracy, Message Credibility, and Sharing Intentions for Misinformation. *JMIR formative research* 8.
- Löber, Lena und Roßnagel, Alexander (2019): Kennzeichnung von Social Bots: Transparenzpflichten zum Schutz integrier Kommunikation. *Multimedia und Recht*, 08/2019, S. 493-498.
- Lockhart, Angus und Tessono, Christelle (2025): Human or AI? Evaluating Labels on AI-Generated Social Media Content. Dais.ce. URL: <https://dais.ca/reports/human-or-ai/> (besucht am 30.01.2026).
- Maras, Marie-Helen und Alexandrou, Alex (2019): Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *International Journal of Evidence & Proof*, Volume 23, Issue 3, S. 255-262.
- Meding, Kristof und Sorge, Christoph (März 2025): What constitutes a deepfake? Symposium of Computer Science and Law, URL: <https://dl.acm.org/doi/10.1145/3709025.3712218> (besucht am 30.01.2026).
- Merkle, Marieke (2025): KI-VO Art. 50 Transparenzpflichten für Anbieter und Betreiber bestimmter KI-Systeme. In: David Bomhard, Fritz-Ulli Pieper und Susanne Wende (Hrsg.): *KI-VO*. Frankfurt am Main: Dfv Mediengruppe.
- Molavi Vasse'i, Ramak (2024): Watermarking von KI-generierten Inhalten als regulatorisches Instrument. *Recht Digital*, 09/2024, S. 406-414.

- Newman, Nic (Januar 2026): Journalism and Technology Trends and Predictions 2026. *The Reuters Institute For The Study Of Journalism*. URL: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2026-01/Trends_and_Predictions_2026.pdf (besucht am 30.01.2026).
- Oswalt, Stefanie (20. Juni 2018): Wie Kennedy eine Rede nach seinem Tod hielt. *Deutschlandfunk-Kultur* URL: <https://www.deutschlandfunkkultur.de/auditive-fake-news-wie-kennedy-eine-rede-nach-seinem-tod-100.html> (besucht am 30.01.2026).
- Partnership on AI (13. Dezember 2023): Building a glossary for Synthetic Media Transparency Methods, Part 1: Indirect Disclosure. *Partnershiponai.org*. URL: <https://partnershiponai.org/resource/glossary-for-synthetic-media-transparency-methods-part-1/> (besucht am 30.01.2026).
- Pawelec, Maria und Bieß, Cora (2021): Deepfakes. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG
- Peukert, Alexander (2023): Desinformationsregulierung in der EU - Überblick und offene Fragen. *Juristen Zeitung*, 07/2023, S. 278-286.
- Pfeiffer, Charlene (2025): Tatbestandsmerkmale, Rechtfertigungsgründe und Herausforderungen beim strafrechtlichen Schutz vor Deepfakes. *Multimedia und Recht*, 12/2025, S. 934-939.
- Rauchfleisch, Adrian; Vogler, Daniel und de Seta, Gabriele (2025): Deepfakes or Synthetic Media? The Effect of Euphemisms for Labeling Technology on Risk and Benefit Perceptions. *Social Media + Society*. Volume 11, Issue 3.
- Schertz, Christian (2025): Postmortales Persönlichkeitsrecht in Zeiten von KI und Deepfakes. *Neue Juristische Wochenschrift*, 39/2025, S. 2801-2806.
- Schultz, Annabel (2025): Aufsichtsrechtliches Vorgehen gegen Fake-Profile: Mögliche Maßnahmen im Kampf gegen die Verbreitung von Desinformationen. *Multimedia und Recht*, 03/2025, S. 168-173.
- Spennemann, Dirk H.R. (2025): ‚Delving into‘ the quantification of Ai-generated content on the internet (synthetic data). arXiv. URL: <https://arxiv.org/abs/2504.08755v1> (besucht am 10.02.2026).
- Steinebach, Martin (2024): KI-generierte Abbildungen von Kindesmissbrauch. *Datenschutz und Datensicherheit*, 05/2024, S. 304-309.
- Stosz, Conrad (03. Februar 2019): Policy Options for Fighting Deepfakes. *Georgetown Security Studies Review* URL: <https://georgetownsecuritystudiesreview.org/2019/02/03/policy-options-for-fighting-deepfakes/> (besucht am 30.01.2026).
- Toff, Benjamin und Simons, Felix M. (2025) “Or They Could Just Not Use It?": The Dilemma of AI Disclosure for Audience Trust in News. *The International Journal of Press/Politics*, S. 881-903.
- Völmann, Berit (2025): Dein Körper, meine Bilder? Der Schutz vor bildbasierter sexualisierter Gewalt (auch) in der Gewaltschutzrichtlinie der EU. *Zeitschrift für Urheber- und Medienrecht*, 01/2025, S. 1-8.
- Wittenberg, Chloe; Epstein, Ziv; Peloquin-Skulski, Gabrielle; Berinsky, Adam J. und Rand, David G. (2025): Labeling AI-generated media online. *PNAS Nexus*, Volume 4, Issue 6.

Xing, Xiaodan; Shi, Fadong; Huang, Jiahao; Wu, Yinzhe; Nan, Yang; Zhang, Sheng; Fang, Yingying; Roberts, Michael; Schönlieb, Carola-Bibiane; Del Ser, Javier und Yang, Guang (2025): On the caveats of AI autophagy, *Nature Machine Intelligence*, S. 172-180.

Gestaltungsspielräume und Gelingensbedingungen kooperativer Nachrichtenplattformen: Eine vergleichende Fallstudie europäischer Plattforminitiativen aus privatheitstheoretischer Perspektive¹

Markus Uhlmann, Eva Isabel Hoffmeister und Pascal Altenkamp

Zusammenfassung

Der Beitrag untersucht, wie kooperative Nachrichtenplattformen im digitalen Journalismus ein Gegengewicht zu US-amerikanischen Plattformen anstreben. Auf der Grundlage einer qualitativen Interviewstudie werden drei europäische Initiativen kontrastiert, die von zivilgesellschaftlichen Plattforminfrastrukturen über Public-Service-Media-Kooperationen bis zu Start-up-Strukturen reichen. Mit der soziologischen Privatheitstheorie von Ochs (2022) überträgt der Beitrag das Konzept der Privatheit auf kollektive Akteure wie Verlagshäuser und öffentlich-rechtliche Nachrichtenanbieter. Im Zentrum steht die Frage, wie Nachrichtenplattformen Privatheitsspielräume im Sinne organisationaler Autonomie für journalistische Akteure gegenüber großen Plattformen eröffnen wollen. Der Fallvergleich zeigt, dass Nachrichtenplattformen in verschiedenen Hinsichten (z. B. Distributionswege, Datenpraktiken, Monetarisierungsformen) auf die Stärkung organisationaler Autonomie journalistischer Akteure abzielen. Zugleich etablieren die zivilgesellschaftliche Kooperation und das Start-up Potenziale für organisationale Autonomie von Medienunternehmen durch spezifische Formen der Selbstbeschränkung. Die Public-Service-Media-Kooperation unterliegt wiederum externen Beschränkungen durch Datenschutzanforderungen, die als struktureller Nachteil gegenüber großen Plattformen erscheinen. Damit macht der Beitrag die anspruchsvollen Gelingensbedingungen kooperativer Nachrichtenplattformen sichtbar und unterstreicht die Notwendigkeit einer integrativen Betrachtung verschiedener Plattforminitiativen.

1 Der Beitrag entstand im Rahmen des Verbundprojekts „Demokratie-X: Analyse der Tragfähigkeit einer privatheitsschonenden, fairen und gemeinwohlorientierten Plattform für Nachrichten“, das seit dem 09/2024 vom Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) gefördert wird (Förderkennzeichen: 16KIS2174).

1. Einleitung

Krisendiagnosen des digitalen Journalismus werden häufig auf die Rolle großer Plattformen zurückgeführt. So konkurrieren Medienunternehmen seit geraumer Zeit mit reichweitenstarken Plattformen am Werbemarkt und sind zugleich von ihnen zur Verbreitung und Finanzierung journalistischer Inhalte abhängig (Lobigs 2018; Neuberger u. a. 2019). Insbesondere große Internetunternehmen wie Google (Alphabet) oder Facebook (Meta) setzen sich als „funding intermediaries“ (Papaevangelou 2023) des Journalismus in Szene. So erhalten etwa 2000 Medienunternehmen, die an der *Google News Showcase-Initiative* beteiligt sind, einen monatlichen Betrag für die Bereitstellung ihrer Inhalte für den Nachrichtenaggregator Google News (Colussi u. a. 2025, S. 1). Doch obwohl Google als Medienmäzen des Journalismus auftritt (Dachwitz/Fanta 2023), bewerten Medienunternehmen die monetäre Kompensation als gering (Colussi u. a. 2025, S. 5). Darüber hinaus konstituieren sich Plattformen auf der Grundlage von journalistischen Inhalten, die hochwertige Suchergebnisse erst ermöglichen oder die als KI-Trainingsmaterial verwendet werden, als privatökonomische Datenunternehmen.² Allerdings tragen die von regulierenden Instanzen regelmäßig angemahnten Ausgleichszahlungen, die Plattformen wie Google oder Facebook an Medienunternehmen entrichten sollen, nur bedingt zur Stärkung von Handlungsspielräumen journalistischer Organisationen bei. Denn auch solche Ausgleichszahlungen binden Medienunternehmen langfristig an große Plattformen (Bossio u. a. 2024).

Der vorliegende Beitrag untersucht daher sogenannte Strategien des „Counterbalancing“ (Sehl/Ganter 2025, S. 272), die von vornherein eine Reduzierung von Abhängigkeiten von großen Plattformen durch den Aufbau alternativer Plattformstrukturen anstreben. Hierzu wurden in jüngerer Zeit Perspektiven für *kooperative Medienplattformen* vorgelegt (z. B. Buschow/Wellbrock 2022; Gostomzyk u. a. 2021; Dogruel u. a. 2025). Diese Arbeiten setzen sich mit der Frage auseinander, wie über die kooperative Bündelung von beispielsweise publizistisch-redaktionellen und technisch-infrastrukturellen Ressourcen ein Gegengewicht zu großen Plattformunternehmen geschaffen werden kann.

2 So wird der ökonomische Wert, der durch professionelle journalistische Inhalte für Google ermöglicht wird, in einer Studie der Public Interest Foundation auf 2,2 Billionen geschätzt (Public Interest Foundation 2025).

Gleichwohl verbleiben diese Überlegungen oftmals auf konzeptioneller Ebene, sodass spezifische Herausforderungen und Realisierungsbedingungen von kooperativen Nachrichtenplattformen zwar artikuliert (z. B. Gostomzyk u. a. 2021), aber nicht anhand konkreter Fallstudien vertieft und vergleichend untersucht werden. Der Beitrag greift diese Lücke auf, indem er drei konkrete Kooperationen europäischer Plattforminitiativen des digitalen Journalismus analysiert. Im Zentrum steht die Frage, wie Nachrichtenplattformen Privatheitsspielräume für journalistische Organisationen gegenüber großen Plattformen eröffnen wollen und unter welchen Bedingungen dies gelingt. Da Privatheit im Kontext kollektiver Akteure wie journalistischer Organisationen kein etabliertes Analysekonzept ist, wird auf die soziologische Privatheitstheorie von Ochs (2022) Bezug genommen. Diese versteht Privatheit als einen relationalen Mechanismus, der grundsätzlich sowohl für Individuen als auch kollektive Akteure gilt (Ochs 2022, S. 41f.). Im Folgenden wird Privatheit daher nicht primär als Schutz personenbezogener Daten thematisiert. Stattdessen behandeln wir Privatheit im Sinne organisationaler Autonomie journalistischer Akteure, die sich etwa auf Spielräume zur Monetarisierung oder Distribution von Nachrichten beziehen kann. Dabei werden potenzielle Autonomiegewinne journalistischer Organisationen als Resultat spezifischer „Teilhabebeschränkungen“ (Ochs 2022) großer Plattformen an beispielsweise journalistischen Entscheidungs- und Wissensprozessen verstanden. Wie der Beitrag aufzeigt, eröffnet diese privatheitstheoretische Perspektive den Blick für die relationale Komplexität kooperativer Nachrichtenplattformen und damit verbundener Gestaltungsherausforderungen.

Die betrachteten Plattformkooperationen reichen von EU-geförderten Plattforminfrastrukturen für zivilgesellschaftliche Non-Profit-Medienunternehmen über Public-Service-Media-Kooperationen bis zu Start-up-Strukturen. Diese Plattformprojekte fokussieren nicht nur auf eine Reduzierung der Abhängigkeiten von großen Internetunternehmen; auch spielen unterschiedliche Formen selbstgewählter oder externer Beschränkungen bei den Plattforminitiativen eine zentrale Rolle. So leitet die zivilgesellschaftliche Kooperation Nutzende von der Plattform direkt zu den Webseiten der Nachrichtenanbieter weiter. Diese *freiwillige Selbstbeschränkung* der Plattform an Nachrichteninhalten erweist sich als Grundlage für die Eröffnung organisationaler Autonomie für Medienunternehmen, die mit der Plattform kooperieren. Wie deutlich werden wird, unterliegt das Start-up einer *ökonomisch erzwungenen Selbstbeschränkung*, da das spezifische Monetarisierungsmodell die Anzahl der Verlagshäuser auf der Plattform begrenzt. Die

Public-Service Media-Plattform steht wiederum vor anderen strukturellen Herausforderungen: Während Logins oder Einwilligungen umfassende Nutzendendatenanalysen bei großen Internetunternehmen ermöglichen, werden bei der öffentlich-rechtlichen Plattform personalisierte Empfehlungen von Nachrichteninhalten oder die Umsetzung von Reichweitenkontrollen durch Datenschutzanforderungen begrenzt. Damit werden ausgerechnet jene Datenschutzinstrumente, die die Macht großer Plattformen einschränken und Nutzende vor sozialschädlichen Nebenfolgen der Datenverarbeitung schützen sollen, für die öffentlich-rechtliche Plattform als *externe Begrenzung* wahrgenommen. Diese selbstgewählten bzw. externen Beschränkungen bilden sowohl Voraussetzung als auch Grenzen der von kooperativen Nachrichtenplattformen angestrebten Autonomiespielräume. Damit rücken sie als zentrale Bedingungsfaktoren kooperativer Nachrichtenplattformen in den Fokus.

Der Beitrag ist wie folgt gegliedert: Zunächst stellen wir ausgewählte Debatten zu Nachrichtenplattformen dar und verknüpfen diese mit der soziologischen Privatheitstheorie von Ochs (2022) (Abschn. 2). Anschließend erläutern wir die methodischen Grundlagen unserer Fallstudie (Abschn. 3). Sodann vergleichen wir verschiedene Pfadoptionen für kooperative Nachrichtenplattformen hinsichtlich ihrer Gelingensbedingungen und potenziellen Gestaltungsspielräume, die für journalistische Organisationen gegenüber großen Plattformen angestrebt werden (Abschn. 4). Das Fazit fasst die zentralen Erkenntnisse zusammen und gibt einen Ausblick für offene Forschungsfragen (Abschn. 5).

2. Gestaltungsspielräume durch kooperative Nachrichtenplattformen?

Kooperative Nachrichtenplattformen werden zunehmend als Alternativen zu plattformökonomischen Strukturen diskutiert (z. B. Buschow/Dobusch 2025; Gostomzyk u. a. 2021). Zur Systematisierung bestehender Initiativen ist zunächst die Unterscheidung zwischen privaten und öffentlich-rechtlichen Nachrichten Anbietern zentral. Für Kooperationen zwischen privaten Medienhäusern wird etwa aus einer medienökonomischen Perspektive für anbieterübergreifende Plattformen argumentiert (Erbrich u. a. 2024). Im Sinne eines „Spotify für Journalismus“ (Buschow/Wellbrock 2022) sollen hier Inhalte unterschiedlicher Verlagshäuser gebündelt und für einen monatlichen Flatrate-Preis zugänglich gemacht werden. Weiterhin wird die besondere Rolle des öffentlich-rechtlichen Rundfunks als infrastrukturel-

ler „Enabler“ für die Verbreitung von Nachrichten und die Gestaltung öffentlicher Diskursräume jenseits von Plattformlogiken betont (Neuberger 2022; Dogruel u. a. 2025). Schließlich werden hybride Kooperationsformen zwischen privaten und öffentlich-rechtlichen Anbietern differenziert (Gostomzyk u. a. 2021; Rauch u. a. 2024), wobei zwischen Kooperationen auf infrastruktureller und publizistischer Ebene unterschieden wird (Buschow/Dobusch 2025).

Wir untersuchen im Folgenden kooperative Nachrichtenplattformen hinsichtlich der Schaffung von Privatheitsspielräumen, die sich potenziell für journalistische Organisationen gegenüber *großen Plattformunternehmen* ergeben. Um Privatheit als Analysekonzept auf kollektive Akteure wie Organisationen anzuwenden, nehmen wir auf die soziologische Privatheitstheorie von Ochs (2022) Bezug. Diese begreift Privatheit als einen relationalen Mechanismus, der unabhängig davon gilt, ob es sich bei privatheitsgenießenden Entitäten um individuelle oder kollektive Akteure handelt (Ochs 2022, S. 41f.). Leitend ist dabei die Überlegung, dass bei privatheitsrelevanten Phänomenen einerseits die *Teilhabe* von Akteuren an bestimmten Erfahrungsspielräumen anderer Akteure *eingeschränkt* wird. Diese Teilhabebeschränkungen *eröffnen* andererseits neue Erfahrungsspielräume für privatheitsgenießende Akteure, die ohne solche Teilhabebeschränkungen nicht bestehen würden (Ochs 2022, S. 41f.).

Bevor diese Grundüberlegung auf die Beziehung zwischen journalistischen Organisationen und Plattformunternehmen übertragen wird, gilt es den grundlegenden Mechanismus anhand der im Datenschutz üblicherweise thematisierten Relation von Nutzenden und Organisationen zu illustrieren. Wenn die Privatheit von Nutzenden gegenüber Internetunternehmen oder staatlichen Instanzen verhandelt wird, geht es um den Schutz der privaten Kommunikation gegenüber ökonomischer Ausbeutung oder staatlicher Überwachung. Somit zielen Datenschutzmaßnahmen auf die Beschränkung der Teilhabe von Organisationen an Nutzendendaten, um sozialschädliche Risiken und Machtasymmetrien durch organisationale Datenverarbeitungen einzugrenzen. Dadurch sollen wiederum Erfahrungsspielräume für Nutzende eröffnet werden, die andernfalls durch Datenpraktiken eingeschränkt werden würden (Ochs 2022, S. 28f.).

Im Vordergrund der privatheitstheoretischen Auseinandersetzung steht im Folgenden die Relation zwischen journalistischen Akteuren und Inter-

netunternehmen.³ Privatheitsspielräume verstehen wir dabei als Formen organisationaler Autonomie, die etwa die journalistische Kontrolle über Distributionswege, Datenpraktiken oder Monetarisierungsformen betreffen. Im digitalen Journalismus konstituieren sich etwa große Plattformen wie Google als *privatökonomische* Akteure, da sie an journalistischen Artikeln im großen Stil teilhaben können und zugleich durch proprietäre Daten- und Geschäftsmodelle die Teilhabe anderer Akteure an ihren Praktiken begrenzen. Entsprechend stellen Diskussionen zum Urheberrecht und zu fairen Ausgleichszahlungen darauf ab, die als unzureichend vergütet empfundene Teilhabe von Plattformen an Nachrichteninhalten zu *beschränken*. Dadurch sollen wiederum privatökonomische und journalistische Autonomiespielräume zugunsten der Nachrichtenanbieter *erweitert* werden. Zentral ist somit die Überlegung, dass Teilhabebeschränkungen von großen Plattformen an Daten oder journalistischen Wissens- und Entscheidungsprozessen erst Autonomiepotenziale im Sinne von Erfahrungsspielräumen eröffnen, die journalistische Akteure wie Redaktionen und Medienhäuser zur problemangemessenen Erfüllung gesellschaftlicher Funktionen benötigen. Sofern journalistische Organisationen Autonomieansprüche haben, stehen dahinter immer auch explizit Personengruppen wie Redaktionen, die als kollektive Einheit auftreten und als *privatheitsgenießende* Akteure gefasst werden können.

Regulative Maßnahmen sind dabei auf die Etablierung direkter Teilhabebeschränkungen von großen Internetunternehmen ausgerichtet. Demgegenüber lassen sich kooperative Nachrichtenplattformen als Versuche verstehen, von vornherein alternative Konfigurationen von Teilhabebeschränkungen zu etablieren. Bei anbieterübergreifenden Journalismus-Plattformen mit einem Flatrate-Modell geht es etwa um die perspektivische Beschränkung der Teilhabe von großen Plattformen an der Finanzierung journalistischer Inhalte. Privatökonomische Potenziale für journalistische Organisationen sollen somit durch neue Monetarisierungsmöglichkeiten erschlossen werden (Erbrich u. a. 2024). Aktuelle Überlegungen zu kooperativen Nachrichtenplattformen adressieren zudem die Schaffung von Infrastrukturen, die die Verbreitung und Sichtbarkeit von Qualitätsinhalten

3 Eine Ausnahme bildet die im Abschn. 4.2 behandelte Public Service Media-Kooperation. In diesem Fallbeispiel werden Datenschutzmaßnahmen für Nutzende von der öffentlich-rechtlichen Plattform als strukturelle Herausforderung thematisiert, die auch Konsequenzen für die Relation zwischen öffentlich-rechtlichen Organisationen und großen Plattformen hervorbringt. Daher wird das Verhältnis von Nutzenden und Organisationen hier explizit zum Thema gemacht.

jenseits großer Plattformunternehmen erlauben. Diese Erwartung wird insbesondere an öffentlich-rechtliche Akteure herangetragen, die einer institutionell verankerten Integrationsfunktion hinsichtlich der Gestaltung von Öffentlichkeit unterliegen (Dogruel u. a. 2025). Kooperative Nachrichtenplattformen zielen hier darauf, den bisherigen Umfang der Teilhabe großer Plattformen an der Verbreitung von Nachrichteninhalten zu begrenzen, damit wiederum normative Zielsetzungen öffentlich-rechtlicher Nachrichtenanbieter unabhängiger von großen Plattformlogiken umgesetzt werden können. Die Untersuchung von Privatheitsspielräumen verweist damit stets auch auf die Etablierung spezifischer Machtrelationen zwischen verschiedenen Akteuren (dazu auch Ochs 2022, S. 19f.). Wenn etwa kooperative Nachrichtenplattformen die Teilhabe großer Plattformen an der Gestaltung des digitalen Journalismus beschränken wollen, streben sie auf diese Weise neue Potenziale organisationaler Autonomie für journalistische Akteure an.

Die Realisierung solcher Gestaltungsspielräume ist jedoch an anspruchsvolle Bedingungen geknüpft. Denn mit dem Aufbau kooperativer Nachrichtenplattformen können wiederum neue Abhängigkeiten, Einschränkungen oder Koordinationsprobleme einhergehen. Beispielsweise sind kooperative Nachrichtenplattformen angehalten, nicht in Konkurrenz zu den Medienunternehmen zu treten, von denen sie abhängig sind. Herausforderungen von Plattformkooperationen – insbesondere mit privaten Medienhäusern – betreffen etwa die Erlösverteilung, Datenhoheit und die Vermeidung von Konkurrenz zu bestehenden Geschäftsmodellen (Nielsen/Ganter 2022, S. 91f.). Diese Aspekte werden insbesondere im Kontext abonnentenbasierter Flatratemodelle diskutiert. So verzichten etwa konkrete Start-up-Projekte für Journalismusplattformen oftmals auf ein solches Monetarisierungsmodell, sofern eine faire Erlösverteilung nicht problemlos umgesetzt werden kann (Uhlmann u. a. 2024, S. 43). Daher fragt der Beitrag nicht nur danach, wie Nachrichtenplattformen spezifische Teilhabebeschränkungen hervorbringen und welche Autonomiespielräume dadurch eröffnet werden. Ebenso untersuchen wir, welche organisationalen, ökonomischen oder institutionellen Rahmenbedingungen Plattformkooperationen überhaupt erst ermutigen.

3. Methodisches Vorgehen

Um zu analysieren, wie kooperative Nachrichtenplattformen auf die Schaffung neuer Gestaltungsspielräume abzielen, basiert die Studie auf einer

vergleichenden qualitativen Fallanalyse (Yin 2018), die einem abduktiven Forschungsansatz folgt (Alvesson/Sköldberg 2009). Dieser Ansatz verbindet induktive und deduktive Elemente, bei dem empirische Beobachtungen und theoretische Konzepte iterativ aufeinander bezogen werden. Dabei dienen der dargestellte Forschungsstand zu Nachrichtenplattformen und die privatheitstheoretischen Überlegungen als Grundlage zur konzeptionellen Sensibilisierung des Fallvergleichs. Im Analyseprozess kristallisierten sich drei zentrale Analyseperspektiven heraus: die im Rahmen digitaler Nachrichtenplattformen realisierten Kooperationsformen, die damit verbundenen Gelingensbedingungen der Kooperation sowie die daraus ableitbaren potenziellen Privatheitsspielräume, die kooperative Nachrichtenplattformen für journalistische Akteure anstreben.

Zur empirischen Konkretisierung wurden drei Fallbeispiele theoriegeleitet und kontrastierend ausgewählt. Erstens wurde eine (1) zivilgesellschaftliche Medienkooperation mit genossenschaftlicher Governancestruktur untersucht, die auf gemeinwohlorientierte Infrastrukturentwicklung und partizipative Entscheidungsprozesse abzielt. Zweitens wurde eine (2) institutionell regulierte Kooperation bestehender Public-Service-Media-Akteure einbezogen, die vor dem Hintergrund eines öffentlichen Auftrags neue Distributions- und Kooperationsformen erprobt. Drittens wurde eine (3) Kooperation zwischen Medienunternehmen und einem privatwirtschaftlichen Start-up analysiert, bei der neue Erlös- und Geschäftsmodelle im Zentrum stehen.

Die empirische Datengrundlage bildet eine qualitative Interviewstudie. Zwischen Juli und September 2025 wurden fünf ca. einstündige semistrukturierte Interviews mit Vertretern der drei Fallbeispiele (u. a. Gründer, Geschäftsführer und Strategiemanager) sowie drei ca. einstündige Interviews mit Experten des Digitaljournalismus (z. B. Presseagentur, Verlegerverband) geführt (vgl. Tab. 1). Die Interviews umfassten Fragen zur Entstehung und Zielsetzung der Plattformprojekte, zu Geschäfts- und Erlösmodellen, Governance- und Entscheidungsstrukturen, Kooperationsformen, Machtverhältnissen im digitalen Journalismus sowie Überlegungen zu Privatheit, Datenschutz und Datenhoheit. Ergänzend wurden Sekundärmaterialien wie öffentliche Berichterstattungen sowie Projekt- und Unternehmenswebseiten erfasst. Da die Interviews primär mit Vertretern der Plattforminitiativen geführt wurden, rekonstruiert die folgende Analyse die intendierten Gestaltungsspielräume und Kooperationsbedingungen insbesondere aus der Plattformperspektive. Die tatsächlich realisierten Potenzia-

le für journalistische Akteure sind somit nicht Gegenstand des vorliegenden Textes.

Tabelle 1: Übersicht über die interviewten Akteure.

#	Interviewpartner [Typ, Rolle]
1	Fallbeispiel 1, <i>Mitglied des Aufsichtsrates</i>
2	Fallbeispiel 1, <i>Leitung Medienpartnerschaften</i>
3	Fallbeispiel 2, <i>Strategiemanager</i>
4	Fallbeispiel 3, <i>Strategischer Berater</i>
5	Fallbeispiel 3, <i>Geschäftsführer</i>
6	Experte, <i>Chefredakteur einer Presseagentur</i>
7	Experte, <i>Leitung Digital in einem Verlegerverband</i>
8	Experte, <i>Leitung Innovation in einem Thinktank</i>

4. Fallbeispiele kooperativer Medienplattformen

4.1 Europäische Plattform-Kooperative mit zivilgesellschaftlicher Fundierung (Fallbeispiel 1)

Das erste Fallbeispiel untersucht eine Plattformkooperation, an der über fünfzig Medienunternehmen aus verschiedenen europäischen Ländern beteiligt sind, von denen die meisten aus dem Non-Profit-Bereich kommen. Die Zielgruppe dieser Plattformkooperation stellen insbesondere Nutzende dar, die an qualitativ hochwertigen Nachrichten mit einem europäischen Bezug interessiert sind. Die an der Kooperation beteiligten Akteure haben sich zum Teil bereits vor dem Plattformprojekt in einem Konsortium zusammengeschlossen, um an einer EU-Förderung für kooperative Plattformstrukturen teilzunehmen. Entsprechend liegt auf der Ebene der Finanzierung eine hybride Konstellation vor: EU-Fördermittel ermöglichen die Erprobung neuer zivilgesellschaftlich getragener Medienstrukturen. Aufgrund der Depriorisierung von qualitativ hochwertigen Nachrichteninhalten auf US-amerikanischen Plattformen sei dies aus der Sicht der Plattform unabdingbar. Entsprechend wird auch das vieldiskutierte Konzept der europäischen Souveränität angeführt, die letztlich nur durch eigene

Infrastrukturen gewährleistet werden könne: So greife „Regulierung in Europa jetzt zu kurz“ [...], weshalb „es souveräne europäische Infrastrukturen braucht, um ein funktionierendes System zu entwickeln.“ (Mitglied des Aufsichtsrates) Dabei ist dieses Plattformexperiment zwar durch eine EU-Förderung möglich geworden; zugleich versteht es sich nicht als eine staatliche Kooperation, sondern als zivilgesellschaftlich gestütztes Projekt. Privatheitstheoretisch zielt die angestrebte Plattforminfrastruktur darauf, sowohl die Teilhabe von großen Plattformen als auch die politischer Akteure an der Gestaltung des Journalismus einzuschränken, um wiederum die organisationale Autonomie journalistischer Akteure in Europa zu stärken.

Zentral ist dabei der Einsatz KI-basierter Übersetzungen, die Inhalte sprachübergreifend zugänglich machen: „Wir sind eine europäische Alternative [...], die in mehr als 20 verschiedenen Sprachen funktioniert, sodass auch Zielgruppen [erreicht werden], die normalerweise keinen Zugang zu Inhalten in Sprachen haben, in denen [Medienunternehmen] nicht veröffentlichen.“ (Leitung Medienpartnerschaften) Diese Übersetzungen allein sind allerdings nicht ausreichend, um Kooperationen mit Medienunternehmen zu ermutigen. Denn so „gibt es eine kritische Sicht der Medienanbieter:innen und dem haben wir immer von vornherein versucht Rechnung zu tragen, indem wir nicht in Konkurrenz treten wollten zu den Medienprodukten von ihrer Seite.“ (Mitglied des Aufsichtsrates) Daher erscheinen die Nachrichteninhalte auf der Plattform lediglich als kurze Zusammenfassung, während sie vollständig nur auf den Webseiten der Nachrichtenanbieter gelesen werden können. Dabei wird für Nutzende zu jedem Zeitpunkt transparent, dass die sprachliche Übersetzung von der Nachrichtenplattform stammt, die jeweiligen journalistischen Artikel aber mit den konkreten Verlagshäusern verknüpft sind. Daher bleiben „die Leute [...] nicht auf unserer Plattform, um die Nachrichteninhalte von Medienunternehmen zu lesen, sondern wir leiten sie einfach auf Ihre eigene Webseite weiter. So profitieren [Medienunternehmen] von den Zugriffen, die über die Plattform generiert werden.“ (Leitung Medienpartnerschaften) Insofern unterscheidet sich die Plattform grundlegend von etablierten Plattformlogiken, da sie „nur so viel eine Mittlerposition [zwischen Publishern und ihren Nutzenden] wie nötig“ anstrebt (Leitung Medienpartnerschaften).

Während journalistische Verlage auf großen Plattformen zu Inhaltelieferanten degradiert werden und tendenziell die Kontrolle über ihre Nutzenbeziehung verlieren, agiert diese Plattform vorrangig mit dem Ziel, eine neue Vermittlungsstruktur zwischen Medienhäusern und Nutzenden zu schaffen. Diese für Medienunternehmen minimal-invasive Plattformlogik

kann somit im Sinne einer *selbstgewählten Teilhabebeschränkung* der Plattform an Nachrichteninhalten begriffen werden. Damit sollen wiederum Privatheitsspielräume für journalistische Organisationen eröffnet werden, die sich insbesondere auf die Kontrolle von Nutzendenbeziehungen und die Inheldistribution beziehen.

Darüber hinaus wird die Schaffung von Datenhoheit gegenüber großen Plattformen betont. So werden auch die „Daten bei uns [...] nicht weiterverkauft und wir sind selbst quasi im driver seat, was wir mit diesen Daten tun.“ (Mitglied des Aufsichtsrates) Demgegenüber stellt die Umsetzung von Datenhoheit insbesondere bezüglich der Large Language Models (LLMs) des Unternehmens OpenAI, die für die Übersetzung der journalistischen Inhalte genutzt werden, eine Herausforderung dar. Aus der Sicht der Nachrichtenplattform ist es zentral, dass „Inhalte [von Medienunternehmen] nicht gestohlen werden können und man da quasi auch, was das Ausspielen betrifft, Hoheit hat darüber.“ (Mitglied des Aufsichtsrates) Die bestehende Kooperation mit OpenAI gerät somit wiederum in Spannung zur Perspektive europäischer und souveräner Infrastrukturen. Entsprechend ist die Nutzung der Technologien von OpenAI als konfliktvolle Übergangskooperation zu begreifen, weshalb es als eine „klare Strategie erscheint, dass wir unsere eigene Infrastruktur aufbauen wollen, wo wir unser eigenes LLM [...] betreiben können.“ (Mitglied des Aufsichtsrates) Ziel ist es hier, die Teilhabe großer Plattformen an Nachrichteninhalten für das KI-Training perspektivisch zu begrenzen und durch den Aufbau eigener Infrastrukturen Autonomiespielräume von Medienunternehmen, insbesondere hinsichtlich ihrer Datenhoheit, zu stärken.

Schließlich spielt die genossenschaftliche Governancestruktur der Plattform eine tragende Rolle, die in der Diskussion zu gemeinwohlorientierten Plattformmodellen unter dem Stichwort „Platform Cooperatives“ (Scholz 2023) firmiert. Diese zeichnen sich dadurch aus, dass der Plattformbetreiber anders als etwa Google News kein machtvoller Orchestrator journalistischer Infrastrukturen ist, sondern auf eine partizipative Governance der beteiligten Akteure setzt (z. B. Porschen-Hueck/Rachlitz 2022). Auch bei der hier besprochenen Plattforminitiative können sich die beteiligten Medienunternehmen als Mitglieder mit Stimmrechten bei plattformrelevanten Entscheidungen einbringen:

„Es ist nicht nur eine Plattform, auf der Nachrichtenanbieter [...] Artikel teilen, sondern auch ein Ort, an dem sie [...] mit uns diskutieren können, unsere Gründe für Entscheidungen erfahren können und manchmal

auch über Dinge abstimmen können. Es hat also durchaus einen partizipativen Charakter und vermittelt ein Gefühl der Mitverantwortung.” (Leitung Medienpartnerschaften)

Während der Aufbau einer eigenen Infrastruktur zur Verbreitung und Übersetzung von journalistischen Inhalten Privatheitsspielräume für journalistische Akteure gegenüber Plattformen adressiert, geht es bei der genossenschaftlichen Governance des Plattformprojekts um die Eröffnung von partizipativen Spielräumen an der Plattformgestaltung selbst. Eine zentrale Gelingensbedingung der Plattformkooperation besteht somit nicht nur darin, dass die Plattform den beteiligten Medienunternehmen tatsächlich neue Nutzende zuführt, sondern dass die kooperativen Plattformstrukturen auch von den Medienunternehmen genutzt werden.

4.2 Europäische Public-Service-Media-Plattform (Fallbeispiel 2)

Das zweite Fallbeispiel untersucht ein Plattformprojekt, das ebenso mittels KI-basierter Übersetzungen die sprachunabhängige Verbreitung von journalistischen Inhalten auf europäischer Ebene anstrebt. Bei diesem Fallbeispiel handelt es sich um eine von der EU-Kommission geförderte Plattforminitiative aus dem Public-Service-Media-Kontext, an der über zwanzig öffentlich-rechtliche Nachrichtenanbieter aus siebzehn verschiedenen Ländern beteiligt sind. Dabei reagiert die Plattform auf die Abhängigkeit öffentlich-rechtlicher Anbieter von kommerziellen Distributionsformen:

„Unsere Inhalte in die Hände großer Plattformen zu geben, ist kontraproduktiv für unsere Mission. Unsere Mission ist es, dass unsere vertrauenswürdigen Inhalte die größtmögliche Anzahl von Menschen erreichen. Von einem gewinnorientierten Dritten abhängig zu sein, stellt daher eine enorme Belastung dar.“ (Strategiemanager)

Diese Aussage unterstreicht ein bekanntes Spannungsfeld, in dem sich öffentlich-rechtliche Akteure bewegen. Um ihren Auftrag des Erreichens einer großen Öffentlichkeit zu erfüllen, sind sie einerseits auf große Plattformen zur Inheldistribution angewiesen (Dogruel u. a. 2025). Andererseits können sich beispielsweise redaktionelle Entscheidungen aber auch an Plattformlogiken anpassen – etwa wenn sich Redakteur:innen bezüglich inhaltlicher Schwerpunkte und ihrer Gestaltung an potenziellen Reichweiten von Plattformen ausrichten, wie Eichler (2022, S. 39) anhand empirischer Redaktionsbeobachtungen aufzeigt. Um eine größere Unabhängigkeit von

Plattformlogiken zu gewinnen, bedarf es entsprechend einer umfassenderen und internationalen Kooperation zwischen verschiedenen Nachrichtenanbietern:

„Jeder hat Inhalte, die mehr als nur ein nationales Leben verdienen. [...] Für uns war dies also der Anker, der Wert der vertrauenswürdigen Inhalte, die wir besitzen und die wir im Rahmen von Zusammenarbeit [...] verbreiten. Ich denke, das ist notwendig, wenn man zumindest versuchen will, mit großen Plattformen zu konkurrieren.“ (Strategiemanager)

In diesem Zusammenhang geht es nicht um einen ökonomischen Wettbewerb mit großen Plattformen, sondern um die Stärkung der Sichtbarkeit öffentlich-rechtlicher Akteure. Die Public-Service-Media-Plattform strebt insofern durch den Aufbau eigener Infrastrukturen eine Reduzierung der bisherigen Teilhabe großer Plattformen an der Gestaltung digitaler Öffentlichkeiten an. Zwar ist zu vermuten, dass öffentlich-rechtliche Akteure auch mit einer eigenen Plattform auf die Nutzung großer Plattformunternehmen angewiesen sind (Eichler 2022, S. 9). Die durch eine gemeinwohlorientierte Plattform adressierten Privatheitsspielräume bestehen aber konkret darin, dass öffentlich-rechtliche Anbieter ihre Gegengewichtsfunktion zu ökonomischen Prinzipien und etablierten Plattformlogiken normativ angemessener und unabhängiger umsetzen können (siehe auch Dogruel u. a. 2025).

Auch in diesem Fall sind es die unterschiedlichen geographischen Kontexte, die eine direkte Konkurrenz journalistischer Organisationen ausschließen. Daher reflektiert diese Plattform ebenso Rahmenbedingungen für die potenzielle Beteiligung privater Medienhäuser. Zentrale Teilnahmebedingung ist somit nicht die öffentlich-rechtliche Organisationsform, sondern die Qualität der journalistischen Inhalte, die über eine Zertifizierungsstelle nachgewiesen wird: „Unser Kriterium lautet: Wenn Sie *Journalism Trust Initiative* zertifiziert sind, und einen öffentlichen Auftrag haben, dann sind Sie qualifiziert, mit uns zusammenzuarbeiten.“ (Strategiemanager) Daher profitieren einzelne Nachrichtenanbieter nur dann von der Präsenz anderer Akteure auf der Plattform, sofern die journalistische Qualität sichergestellt ist. Hinsichtlich ihrer prinzipiellen Zielsetzung liegt damit eine sogenannte „supply side synergy platform“ (Rauch u. a. 2024) vor. Hier geht es um eine Plattformkooperation zwischen öffentlich-rechtlichen und privaten Medienhäusern, die für Nutzende eine qualifizierte Meinungsbildung durch ein breites Angebot von Qualitätsinhalten schaffen und für Anbieter

die Erhöhung der Reichweite ihrer Nachrichteninhalte ermöglichen soll (Rauch u. a. 2024, S. 173).

Allerdings werden solche hybriden Kooperationen von privaten und öffentlich-rechtlichen Akteuren in diesem Fallbeispiel *nicht* realisiert. So werden Bedenken privater Medienhäuser angeführt, sofern diese ihre Geschäftsmodelle torpediert sehen, sobald sie ihre Inhalte kostenlos zur Verfügung stellen. Aber auch die Logik eines „Spotify für Journalismus“ (Erbrich u. a. 2024), bei dem Nutzende einen festen (monatlichen) Betrag entrichten, wäre nicht praktikabel. Denn „öffentlich-rechtliche Medien [...] können ihre Inhalte nicht standardmäßig monetarisieren. Denn sie sind bereits bezahlt. [...] Wie kann man also beiden Welten gerecht werden und langfristig Nachhaltigkeit erreichen?“ (Strategiemanager) Die angestrebte Kooperation stößt somit auf eine strukturelle Grenze, weil sie die organisationalen Privatheitsspielräume für Verlagshäuser und öffentlich-rechtliche Nachrichtenanbieter nicht erweitern, sondern im Sinne eines Nullsummenverhältnisses einschränken würde. Während öffentlich-rechtliche Akteure ihre organisationale Autonomie dann gestärkt sehen, wenn ihre frei zugänglichen Inhalte unabhängiger von großen Plattformen sichtbar werden, würde dies wiederum eine Einschränkung der privatökonomischen Spielräume von kommerziellen Verlagshäusern bedeuten, da ihre Monetarisierungslogik zuweilen auf exklusiv zugänglichen Inhalten basiert.

Eine zentrale Gelingensbedingung des Plattformprojekts ist dabei die Erfüllung seines institutionell verankerten Auftrags. Dieser besteht im Erreichen einer breiten Öffentlichkeit mit Qualitätsinhalten (dazu auch Gostomzyk/Lobigs 2025). Die damit verbundenen Erwartungen bzgl. einer Reichweitenkontrolle und Nutzendenbindung an öffentlich-rechtliche Nachrichten kann die Plattform nicht unter Zuhilfenahme klassischer Plattformlogiken umsetzen. Während große Internetplattformen über eine Monopolstellung bezüglich der Verhaltensauswertung verfügen, um Nutzende an ihre Ökosysteme zu binden und Inhalte zu personalisieren (Zuboff 2018), kann die öffentlich-rechtliche Plattform nur eingeschränkt auf vergleichbare Analyseansätze zurückgreifen. Zwar könnte das bewusste Absehen von flächendeckenden Verhaltensanalysen als Element der normativen Verankerung des Gemeinwohlanspruchs der Plattform angesehen werden. Allerdings werden Datenschutzanforderungen aus der Perspektive des Strategiemanagers der Plattform insbesondere bei der Ansprache junger Zielgruppen als Nachteil thematisiert:

„Die Zielgruppe der Jugendlichen ist für alle Medien heute enorm wichtig. Und es ist sehr schwierig, wenn nicht sogar unmöglich, nachzuweisen, dass man sie erreicht hat, es sei denn, man nutzt eine der großen Plattformen. Denn die großen Plattformen können Daten sammeln. [...] Man muss sich einloggen. Aber für Nachrichteninhalte ist das kostenlos. [...] Es ist sehr schwierig, diese Informationen [...] von [...] Nutzern zu erfassen, die sich dafür entschieden haben, anonym zu bleiben [...].“ (Strategiemanager)

Begrenzte Datenanalysemöglichkeiten erzeugen somit Spannungen mit institutionellen Anforderungen, insbesondere beim Nachweis von Reichweite. Gleichzeitig wird eine strukturelle Asymmetrie zu großen Plattformen deutlich. So gelangen große Plattformen und private Medienhäuser etwa über Nutzenden-Logins an sogenannte First-Party-Daten, die umfassendere Datenanalysen ermöglichen (siehe auch Engert u. a. 2023, S. 153). Demgegenüber ist dieser Weg für Public-Service-Media-Projekte nicht gangbar. Denn letztere agieren nicht privatökonomisch und verfügen über keine vergleichbaren Möglichkeiten, Logins oder Einwilligungen für datenintensive Analysen zu rechtfertigen. Damit werden mögliche ambivalente Nebenfolgen von Teilhabebeschränkungen deutlich: jene Datenschutzvorgaben, die die Teilhabe großer Plattformen an Nutzendendaten beschränken sollen, werden von der Public-Service-Media-Plattform als struktureller Nachteil wahrgenommen. Datenschutz wird weniger als Schutz von Nutzenden, sondern als externe Beschränkung zur Erfüllung unterstellter Nutzendenerwartungen thematisiert. Dies zeigt sich hinsichtlich der Anwendung von Personalisierungsalgorithmen. Diese gelten auch im Deutungshorizont der Plattformkooperation als zentrale Voraussetzungen, um Nutzenden relevante Nachrichten zu empfehlen, setzen jedoch wiederum datenbasierte Analysen voraus, die unter gegebenen Bedingungen nur schwerlich zu realisieren sind. Die damit verbundenen Spannungen werden wie folgt auf den Punkt gebracht: „Was ist wichtiger: dass ich Nutzenden die gewünschten Nachrichten empfehle oder dass ich Ihre Daten nicht in meiner Datenbank speichere?“ (Strategiemanager) Datenschutzvorgaben, die Privatheitsspielräume für Nutzende gegenüber Organisationen sichern sollen, geraten damit in ein Spannungsverhältnis zu den Privatheitsspielräumen, die die Plattform für Nachrichtenanbieter zu eröffnen sucht. Das Fallbeispiel zeigt, dass die langfristige Reproduktion und Legitimität öffentlich-rechtlicher Plattforminfrastrukturen nicht nur von einer gemeinwohlorientierten Distributionsinfrastruktur für Nachrichten abhängen, sondern auch davon,

Nutzendenbindung jenseits etablierter algorithmischer Logiken zu ermöglichen. So konstatiert auch der Strategiemanager der Plattform: „Ich glaube, das Hauptproblem besteht darin, eine Beziehung zu den Nutzern aufzubauen. [...] Die Plattform muss genügend Anreize bieten, damit sie zur Gewohnheit wird und somit Teil des Alltags der Nutzer.“

4.3 Start-up für europäische Qualitätspublisher (Fallbeispiel 3)

Das dritte Fallbeispiel nimmt eine Kooperation zwischen einem Start-up und rund dreißig europäischen Medienhäusern in den Blick. Auch diese Plattform setzt auf die Verbreitung von europäischen Qualitätseinhalten mithilfe KI-basierter Übersetzungen. Dabei will das Start-up neue Monetarisierungsmöglichkeiten schaffen, wie der Geschäftsführer ausführt: „Also [unsere] Vision ist ja quasi der europäischen Sprachfragmentierung zu begegnen und dafür zu sorgen, dass Medienhäuser, die hochwertige Inhalte produzieren, diese auch monetarisieren können in Zielgruppen, die gar nicht die eigene Sprache verstehen.“ (Geschäftsführer) Im Sinne der Privatheitstheoretischen Überlegungen von Ochs (2022) geht es dem Start-up darum, Wertschöpfungsbeziehungen aus klassischen Reichweitenlogiken herauszulösen und die Teilhabe großer Plattformen an der Finanzierung journalistischer Inhalte zu beschränken. Um wiederum privatökonomische Spielräume für Medienunternehmen zu eröffnen, verfolgt das Start-up ein abonnementbasiertes Modell zur Monetarisierung sprachübergreifender Inhalte.

Wie löst die Plattform jedoch eine zentrale Kooperationsherausforderung, die in der Ermöglichung eines fairen Verteilungsschlüssels der über die Plattform generierten Erlöse besteht? Dazu hat sich das Start-up dafür entschieden, Verlagspartner abhängig von der bereitgestellten Inhaltsmenge zu vergüten. Da somit nicht die Anzahl gelesener journalistischer Artikel die Erlösverteilung bestimmt, kann auch der Einfluss der plattformseitigen Kuratierung von Inhalten auf die generierten Erlöse weitgehend ausgeschlossen werden. Zugleich agiert die Plattform durch das gewählte Monetarisierungsmodell in einem potenziell engen Korridor, wenn es um die Erschließung neuer privatökonomischer Potenziale für Medienunternehmen geht. Denn die Plattform unterliegt einer ökonomischen Selbstbeschränkung, sofern sie die Anzahl der Publisher sensibel abwägen muss:

„Weil wenn wir jetzt noch 30 weitere Partner aufnehmen und dann nochmal 30, dann sind wir zwar bei 100 Partnern, aber natürlich wird

der Kuchen auch immer kleinteiliger geschnitten und der Kuchen muss erstmal groß gebacken werden, damit dann nicht jeder einzelne Partner am Ende nur noch 0.5 Prozent von unserem Umsatz bekommt und dann wiederum enttäuscht ist. Also das ist eine schwierige Frage der Balance.“ (Geschäftsführer)

Dabei ist zu vermuten, dass eine höhere Anzahl von Medienunternehmen zwar die Attraktivität und Zahlungsbereitschaft von Nutzenden steigern könnte, zugleich aber die privatökonomischen Spielräume einzelner Medienunternehmen relativieren könnte. Die Ausweitung der Sichtbarkeit durch zusätzliche Verlagshäuser geht damit potenziell mit einer Einschränkung individueller Erlöspotenziale einher.

Darüber hinaus kann die gewählte Monetarisierungsstrategie wiederum im Lichte normativer Zielsetzungen betrachtet werden. So will die Plattform selbst keine Anreize zur polarisierenden Zuspitzung von journalistischen Artikeln schaffen:

„[W]enn man mit [uns] kooperiert, bringt es nichts, eine Überschrift irgendwie zuzuspitzen [...], weil wir auch nicht so monetarisieren, sondern wir haben ja eine Flatrate beim Abo. [...] Das heißt, die Monetarisierung aus Verlagssicht hängt vom Volumen des Contents ab und von der Qualität vielleicht, aber nicht von der Reichweite.“ (Geschäftsführer)

Somit setzt das Start-up darauf, die Teilhabe datenökonomischer Logiken und großer Plattformen an der Definition dessen zu begrenzen, was ökonomisch „wertvolle“ journalistische Inhalte ausmacht. Denn neben ökonomischen Zielsetzungen sollen gleichermaßen Aspekte journalistischer Qualität in der Preisbildung für journalistische Inhalte verankert werden. Letztere spielen in etablierten Reichweitenlogiken kaum eine Rolle, da diese tendenziell auch Clickbait ökonomisch prämiieren können. Mit dem Fokus auf die Contentmenge etabliert die Plattform damit eine Bewertungslogik für journalistische Inhalte, bei der Reichweite oder Klickzahlen nicht mehr als primäre ökonomische Maßstäbe fungieren (dazu auch Kropf u. a. 2025, S. 12f.).

Die Kooperationsform kann dabei als *indirekte Marktkooperation* gefasst werden. So sollen durch die gemeinsame Präsenz von Verlagshäusern auf der Plattform neue Nutzendengruppen in unterschiedlichen Sprachmärkten erreicht werden. Die nutzendenseitige Entscheidung für ein Abo hängt aus der Sicht des Geschäftsführers davon ab, dass die Plattform insbesondere große internationale Publisher mobilisieren kann. So kom-

men „die Leute [...] weil sie sagen, ach, das sind ja tolle Marken. Aber dann finden sie Inhalte von [Publishern] aus Belgien [...] oder Italien. [...] [D]ie sind halt exzellent [...] und die kriegen natürlich eine Sichtbarkeit, die sie sonst nie gehabt hätten.“ (Geschäftsführer) Während also das Start-up von großen Verlagshäusern abhängig ist, die wiederum kaum auf die Plattform angewiesen sind, profitieren kleinere Publisher durch die Erschließung weiterer Sprachmärkte potenziell stärker, ohne maßgeblich zur Gewinnung neuer Nutzender beizutragen. Diese Konstellation führt zu *asymmetrischen Kooperationen* zwischen namhaften Medienunternehmen und der Plattform, sofern die „fehlende Augenhöhe bei großen Verlagshäusern“ (Geschäftsführer) angesprochen wird. Zwar ist die zentrale Kooperationsbedingung privater Medienhäuser erfüllt, nämlich die Erschließung neuer Erlöspotenziale durch internationale Sprachmärkte. Entsprechend ist es durch den „Wegfall der Sprachbarrieren [...] jetzt schon so, dass der Zugewinn [...] potenziell größer ist als der potenzielle Verlust.“ (Geschäftsführer) Inwiefern die Plattform über diese Minimalbedingung hinaus tatsächlich nennenswerte ökonomische Zugewinne eröffnet, ist angesichts des frühen Stadiums des Start-ups ungewiss. Eine mögliche langfristige Kooperationsherausforderung markiert somit die Tatsache, dass sich insbesondere große Verlagshäuser an Geschäftsmodellen mit höheren Gewinnerwartungen orientieren, die sich aber bei einem neu gegründeten Start-up, wenn überhaupt, erst langfristig einstellen. Damit stellt die Bindung der größeren Verlagshäuser, die die Plattform am wenigsten brauchen, diese aber aufgrund ökonomischer Erfolgsnormen am stärksten bewerten, eine Herausforderung für die Reproduktion der Plattform dar:

„Also wir haben da einen sehr langen Diskussionsprozess zum Teil gehabt und hätten dann in dem Moment, wo wir Partner überzeugt haben, [...] vielleicht klarer sagen sollen, das ist ein Experiment und es ist auf mehrere Jahre angelegt. Also wenn jemand beim *Independent* zum Beispiel für Lizenzumsatz zuständig ist, dann guckt er halt jeden Monat oder jedes Quartal auf die Zahlen. Und natürlich wachsen die [...] nicht sehr krass, weil unser Produkt ja auch sehr ungewöhnlich ist und sehr neuartig.“ (Geschäftsführer)

Während die experimentelle Erprobung neuer Geschäftsmodelle für das Start-up selbstverständlich ist, muss dieses Selbstverständnis insbesondere gegenüber größeren Verlagshäusern, die sich nicht in Nischenkontexten bewegen, explizit gemacht werden. Sowohl die Logik der asymmetrischen Kooperation als auch die durch die Monetarisierungsform begründete

ökonomische Selbstbeschränkung markieren potenzielle Grenzen für die Schaffung privatökonomischer Spielräume für Medienunternehmen, auf die das Start-up abzielt.

5. Fazit

Mit der soziologischen Privatheitstheorie von Ochs (2022) analysiert der Beitrag, wie kooperative Nachrichtenplattformen die Stärkung organisationaler Autonomie journalistischer Akteure durch spezifische Teilhabebeschränkungen großer Plattformen anstreben. Dabei setzt die zivilgesellschaftliche Plattformkooperation (Fallbeispiel 1) auf infrastrukturelle Unabhängigkeit hinsichtlich der Vermittlung zwischen Nutzenden und journalistischen Akteuren. Dadurch soll die Teilhabe großer Plattformen an der Verbreitung von Nachrichteninhalten perspektivisch eingeschränkt werden. Die Public-Service-Media-Kooperation (Fallbeispiel 2) ist von der normativen und institutionell verankerten Anforderung getragen, Nachrichtenanbietern eine größere Sichtbarkeit und Reichweite zu verschaffen. Ziel ist es, die Teilhabe großer Plattformen an der umfassenderen Gestaltung digitaler Öffentlichkeiten zu beschränken. Damit soll die Gegengewichtsfunktion zu Marktlogiken, die öffentlich-rechtliche Akteure zu erfüllen haben, unabhängiger von Plattformlogiken umgesetzt werden können. Das Start-up (Fallbeispiel 3) fokussiert wiederum auf eine Teilhabebeschränkung großer Plattformen an der Finanzierung journalistischer Inhalte, um durch neue Monetarisierungsformen privatökonomische Spielräume von Qualitätspublishern zu ermöglichen.

Die Analyse zeigt dabei auf, dass kooperative Nachrichtenplattformen nicht nur die Begrenzung von Plattformmacht anstreben, sondern auch unterschiedliche Formen der Selbstbeschränkung etablieren oder auch externen Beschränkungen unterliegen. Diese können jeweils Voraussetzung und Einschränkung von organisationalen Autonomiespielräumen journalistischer Akteure sein. Dabei entspricht die Weiterleitung von Nutzenden auf die Webseiten der Nachrichtenanbieter bei der zivilgesellschaftlichen Plattform einer selbstgewählten Teilhabebeschränkung der Plattform an Nachrichteninhalten und steht nicht im Widerspruch zu den Plattformzielen. Demgegenüber unterliegt das Start-up einer ökonomischen Selbstbeschränkung. Denn das Monetarisierungsmodell begrenzt die Anzahl der Publisher auf der Plattform. Die Public-Service-Media-Plattform unterliegt

wiederum stärkeren externen Begrenzungen und institutionellen Erwartungen. Während alle untersuchten Plattformen davon abhängig sind, dass sie Nachrichten Anbietern langfristig neue Nutzende zuführen, ist es insbesondere für die Public-Service-Media-Plattform aufgrund ihres öffentlichen Auftrags unerlässlich, dass Nutzende eine Bindung an das öffentlich-rechtliche Ökosystem ausbilden. Daher wird die Konkurrenzfähigkeit gegenüber großen Plattformen insbesondere im Zusammenhang der Public-Service-Media-Kooperation angesprochen, während die anderen Fallbeispiele sich von vornherein stärker als Nischenprojekte verstehen. Allerdings kann die öffentlich-rechtliche Kooperation nur bedingt auf Nutzendendatenanalysen zurückgreifen, die im Rahmen datenökonomischer Plattformlogiken etabliert sind. Daher stellt sich die Frage, inwiefern etwa Diskussionsforen, die eine direkte Beteiligung von Nutzenden im öffentlich-rechtlichen Kontext ermöglichen, als funktionales Äquivalent zu datenbasierten Auswertungen dienen können, um die Bindung an das öffentlich-rechtliche Ökosystem zu stärken.⁴ Während die zivilgesellschaftliche Kooperative und das Start-up von vornherein auf Datensparsamkeit setzen, beklagt die Public-Service-Media-Plattform fehlende Möglichkeiten für Nutzendendatenanalysen. Denn der Nachweis der institutionellen Erwartung des Erreichens einer großen Öffentlichkeit kann nicht über entsprechende Datenanalysen erfolgen. Hier stellt sich die Frage, wie die potenziellen Spannungen zwischen Datenschutzanforderungen und geforderter Reichweitenkontrolle regulativ adressiert werden können. Darüber hinaus wurde deutlich, dass die Public-Service-Media-Kooperation strukturelle Nachteile gegenüber großen Plattformen aufgrund unterschiedlicher datenschutzrechtlicher Anforderungen betont. Neben den Nebenfolgen spezifischer Datenschutzregulierungen werden somit nicht nur Spannungsverhältnisse zwischen unterschiedlichen Teilhabebeschränkungen und Privatheitsinteressen unterschiedlicher Akteure deutlich (dazu insbesondere auch Engert u. a. 2023); ebenso wirft der Vergleich die grundlegendere Frage auf, unter

4 Auch aktuelle Überlegungen argumentieren dafür, dass sich der öffentlich-rechtliche Rundfunk nicht mehr auf die Produktion und Distribution von Nachrichteninhalten fokussieren könne, sondern seine öffentliche Vermittlungsfunktion zunehmend auch im Sinne einer breiteren Integration des Publikums verstehen müsse (z.B. Neuberger 2022, S. 105ff.; Dogruel u. a. 2025). Konstruktive Ansätze in diese Richtung werden etwa im Rahmen des ZDF-Pilotprojektes zum „Public Spaces Incubator“ (PSI) exploriert, das explizit Erwartungen und Bedürfnisse von Nutzenden an digitale Gesprächsräume einbezieht (ZDF 2025). Siehe dazu auch die Studie von Eichler u. a. (2026), die den PSI in den größeren Kontext der Plattformisierung von Medienstrukturen einordnet.

welchen Bedingungen sich kooperative Plattformstrukturen trotz spezifischer (Selbst-)Beschränkungen langfristig stabilisieren lassen.

Schließlich stellen die untersuchten Plattformen durch ihre internationale Orientierung nicht nur konkurrenzarme Kooperationsbedingungen bereit; auch eröffnen die KI-Übersetzungen zuweilen privatökonomische Spielräume, sofern Medienunternehmen neue Nutzendengruppen aus verschiedenen Sprachmärkten gewinnen können. Dies ist insbesondere für das Start-up eine zentrale Kooperationsbedingung, sofern es hier dezidiert um neue Monetarisierungsformen für Verlagshäuser geht. Diskussionen zu kooperativen Nachrichtenplattformen betonen dabei Herausforderungen der prinzipiellen Ermöglichung von Kooperationen (z. B. Nielsen/Ganter 2022, S. 91f.), die bei den untersuchten Fallbeispielen nicht das Kernproblem darstellen. Indes eröffnen die KI-Übersetzungen selbst wiederum ein ambivalentes Spannungsfeld. Während sie einerseits Kooperationen ermöglichen, können sie andererseits gemeinwohlorientierte Plattformen an große KI-Entwicklungsfirmen binden – wenngleich eine Teilhabebeschränkung von US-amerikanischen KI-Unternehmen an Nachrichteninhalten im Falle der zivilgesellschaftlichen Kooperation durch den langfristigen Aufbau eigener KI-Infrastrukturen zumindest angestrebt wird.

Die Ergebnisse ergänzen damit bestehende Arbeiten zu kooperativen Nachrichtenplattformen, indem sie deren Potenziale und Grenzen nicht nur konzeptionell, sondern empirisch entlang konkreter Kooperationsformen, Gelingensbedingungen und Privatheitsspielräume ausbuchstabieren. Damit wird die Diskussion um ein „Spotify für Journalismus“ (Erbrich u. a. 2024) und gemeinwohlorientierte Plattforminfrastrukturen (Gostomzyk u. a. 2021) um eine organisations- und privatheitstheoretische Perspektive erweitert. Gleichwohl legen diese Überlegungen keine romantisierende Perspektive nahe, die von vornherein in den Plattformkooperationen die Schaffung „realer Utopien“ (Wright 2017) jenseits datenökonomischer Plattformstrukturen erblickt. Vielmehr ist zu vermuten, dass Verlagshäuser und öffentlich-rechtliche Nachrichtenanbieter einen maßgeblichen Teil ihrer organisationalen Praktiken weiterhin an den Strukturen der Plattformökonomie ausrichten (müssen), da diese die Nutzendenpraktiken stärker an sich binden und höhere Sichtbarkeit für ein breiteres Publikum bieten (Eichler 2022, S. 9). Gerade große Plattformen könnten somit nach wie vor für bestimmte Medienunternehmen zur Erschließung privatökonomischer Potenziale zentral sein. Daher kann die Stärkung organisationaler Autonomie gegenüber Plattformen durchaus normativ gewünscht, aber im Rahmen organisationaler Praktiken dennoch nicht priorisiert wer-

den – bei allen möglichen Ambivalenzen und langfristigen Nebenfolgen, die damit wiederum einhergehen (dazu auch Nielsen 2026). Solange also gemeinwohlorientierte Plattformen für journalistische Akteure eine optionale Ergänzung bleiben und parallel zu großen Plattformen reproduziert werden, ist zu erwarten, dass auch die angestrebten Teilhabebeschränkungen gegenüber großen Plattformen strukturell schwach bleiben. Somit hat diese Fallstudie insbesondere die *Privatheitspotenziale* für journalistische Akteure aus dem Blickwinkel der Zielsetzungen und Gelingensbedingungen kooperativer Nachrichtenplattformen herausgearbeitet. Wie stark die Abhängigkeiten zu großen Plattformkonzernen allerdings für welche Medienunternehmen konkret reduziert werden können und unter welchen Bedingungen kooperative Nachrichtenplattformen langfristig tragfähig sind, gilt es darüber hinaus eingehender aus der Perspektive journalistischer Organisationen und Nutzendengruppen zu untersuchen. Denn deren Praktiken spielen eine tragende Rolle für die Reproduktion kooperativer Nachrichtenplattformen.

Schließlich ist für zukünftige Überlegungen eine integrative Betrachtung der exemplarisch gegenübergestellten Plattformkooperationen unerlässlich. Hier stellt sich die Frage, inwiefern zivilgesellschaftlich getragene Infrastrukturen (Fallbeispiel 1), institutionell verankerte Public-Service-Media-Kooperationen (Fallbeispiel 2) und die Erprobung neuer Erlösmodelle durch anbieterübergreifende Plattformen (Fallbeispiel 3) jeweils komplementäre Funktionen erfüllen können. Zu klären ist, in welcher Weise die untersuchten Kooperationen wechselseitig voneinander profitieren und lernen können und welche institutionellen, organisationalen und ökonomischen Strukturelemente wiederum Synergien begrenzen.

Tabelle 2: Zusammenfassung der zentralen Ergebnisse.

Plattform	Zentrale Kooperationsformen	Langfristige Gelingensbedingungen der Kooperation	Privatheitsspielräume
Zivilgesellschaftliche Kooperative (Fallbeispiel 1)	Zivilgesellschaftliche Infrastrukturkooperation mit partizipativer Governancestruktur (Genossenschaft)	Finanzierungsstabilität und Aufbau gemeinwohlorientierter Infrastrukturen	Infrastrukturalternative: Teilhabebeschränkungen großer Plattformen an der Inhaltsdistribution und den Daten journalistischer Akteure; Kooperative etabliert plattformseitige „Selbstbeschränkung“ (Nutzendenweiterleitung an Verlage)
Europäische Public Service Media-Plattform (Fallbeispiel 2)	Institutionell regulierte Kooperation bestehender Public-Service-Media-Akteure	Nutzendenbindung und Reichweitenkontrolle im öffentlich-rechtlichen Ökosystem unter restriktiven Bedingungen der Datenanalyse	Gegengewichtsfunktion: Teilhabebeschränkungen großer Plattformen zielen darauf, dass öffentlich-rechtliche Akteure ihre Gegengewichtsfunktion unabhängiger von kommerziellen Logiken umsetzen können
Start-up für europäische Qualitätspublisher (Fallbeispiel 3)	Indirekte und marktbasierende Kooperation zwischen Medienhäusern	Skalierbare Erlöse und langfristige Erfüllung von Gewinnerwartungen	Monetarisierung: Teilhabebeschränkung großer Plattformen an der Finanzierung des Journalismus durch alternative Monetarisierungsmodelle, um privat-ökonomische Spielräume zu erschließen

Literatur

- Alvesson, Mats und Sköldberg, Kaj (2009): *Reflexive Methodology: New Vistas for Qualitative Research*. London: SAGE Publications.
- Bossio, Diana; Carson, Andrea und Meese, James (2024): A different playbook for the same outcome? Examining Google's and Meta's strategic responses to Australia's News Media Bargaining Code. *New Media & Society* 27 (7), S. 3890–3910. <https://doi.org/10.1177/14614448241232296>
- Buschow, Christopher und Dobusch, Leonhard (2025): Kompetitiv und kooperativ: Zur Neuordnung des Verhältnisses von privaten und öffentlich-rechtlichen Medien im Digitalzeitalter. In: *Wirtschaftsdienst* 105 (3), S. 173–177.
- Buschow, Christopher und Wellbrock, Christian-Mathias (2022): Spotify for News? User Perception of Subscription-Based Content Platforms for News Media. *Journalism and Media* 4 (1): 1–15. <https://doi.org/10.3390/journalmedia4010001>

- Colussi, Juliana; de Souza Paes, Paula und Gomes-Franco e Silva, Flávia (2025): Journalism and Economic Sustainability: An Analysis of Google News Showcase in Argentina, Brazil and Colombia. *Journalism Practice* 19 (12), S. 2904–2920. <https://doi.org/10.1080/17512786.2025.2463472>
- Dachwitz, Ingo und Fanta, Alexander (2023): Presse und Plattformen: Ohne Google geht nichts. Von der neuen Abhängigkeit des Journalismus im Digitalen. In: Güney, Selma; Hille, Lina; Pfeiffer, Juliane; Porak, Laura und Theine, Hendrik (Hrsg.): *Eigentum, Medien, Öffentlichkeit. Verhandlungen des Netzwerks Kritische Kommunikationswissenschaft*. Frankfurt am Main: Westend, S. 287–309. <https://doi.org/10.53291/GEPO1696>
- Dogruel, Leyla; de la Durantaye, Katharina; Elmer, Christina; Gostomzyk, Tobias und Lobigs, Frank (2025): *Potenzialanalyse. Perspektiven für Digitalen Public Value im ZDF*. URL: <https://www.zdf.de/assets/potenzialanalyse-digitaler-public-value-im-zdf-102~original?cb=1747990756638> (besucht am 08.04.2026).
- Eichler, Henning (2022): Journalismus in sozialen Netzwerken. ARD und ZDF im Bann der Algorithmen. Frankfurt am Main: Otto Brenner Stiftung. URL: <https://archiv.otto-brenner-stiftung.de/journalismus-in-sozialen-netzwerken/> (besucht am 08.04.2026).
- Eichler, Henning; Uth, Bernadette; Ötting, Hannah und Kokoschka, Vanessa (2026): Debate Spaces of Everyone? How Public Service Media Aim to Reshape Audience Engagement and Interactions. *Digital Journalism*. 1–20. <https://doi.org/10.1080/21670811.2026.2636906>
- Engert, Simon; Kropf, Jonathan und Uhlmann, Markus (2023): Privacy-Trade-offs: Zur Rolle technischer und regulativer Datenschutzinitiativen im Ökosystem des digitalen Journalismus. In: Friedewald, Michael; Roßnagel, Alexander; Neuburger, Rahild; Bieker, Felix und Hornung, Gerrit (Hrsg.): *Daten-Fairness in einer globalisierten Welt*. Baden-Baden: Nomos, S. 145–170.
- Erbrich, Lukas; Wellbrock, Christian-Mathias; Lobigs, Frank und Buschow, Christopher (2024): Bundling Digital Journalism: Exploring the Potential of Subscription-Based Product Bundles. *Media and Communication* 12: 1–18. <https://doi.org/10.17645/mac.7442>
- Gostomzyk, Tobias; Jarren, Otfried; Lobigs, Frank und Neuberger, Christoph (2021): *Kooperative Medienplattformen in einer künftigen Medienordnung*. URL: <https://www.bundesregierung.de/resource/blob/974430/1929884/c5a25bec078cb6846f8ab7a6ca88e80a/2021-06-16-medienbericht-wissenschaftliches-gutachten-data.pdf?download=1> (besucht am 08.04.2026).
- Gostomzyk, Tobias und Lobigs, Frank (2025): §2 Digitaler Public Value und Fortschreibung von Legitimation und Auftrag des öffentlich-rechtlichen Rundfunks. In: Dogruel, Leyla; de la Durantaye, Katharina; Elmer, Christina; Gostomzyk, Tobias und Lobigs, Frank (Hrsg.): *Potenzialanalyse. Perspektiven für Digitalen Public Value im ZDF*. URL: <https://www.zdf.de/assets/potenzialanalyse-digitaler-public-value-im-zdf-102~original?cb=1747990756638> (besucht am 06.01.2026).

- Kropf, Jonathan; Lamla, Jörn und Uhlmann, Markus (2025): Pricing, Design and Cultivation – Conceptual Tools for the Normative Assessment of Fairness in the Age of Media Platformization. *Digital Journalism*, S. 1–19. <https://doi.org/10.1080/21670811.2025.2553161>
- Lobigs, Frank (2018): Wirtschaftliche Probleme des Journalismus im Internet. Verdrängungsängste und fehlende Erlösquellen. In: Nuernbergk, Christian und Neuberger, Christoph (Hrsg.): *Journalismus im Internet. Profession – Partizipation – Technisierung*. Wiesbaden: Springer, S. 295–334.
- Neuberger, Christoph; Bartsch, Anne; Reinemann, Carsten; Fröhlich, Romy; Hanitzsch, Thomas und Schindler, Johanna (2019): Der digitale Wandel der Wissensordnung. Theorierahmen für die Analyse von Wahrheit, Wissen und Rationalität in der öffentlichen Kommunikation. *Medien & Kommunikationswissenschaft*, 67 (2), S. 167–186. <https://doi.org/10.5771/1615-634X-2019-2-167>
- Neuberger, Christoph (2022): Öffentlich-rechtlicher Rundfunk und Plattformen: Programmatik, Prinzip, Praxis, Projekt. In: Gerlach, Frauke und Eilders, Christiane (Hrsg.): *#meinfernsehen2021. Bürgerbeteiligung: Wahrnehmungen, Erwartungen und Vorschläge zur Zukunft öffentlich-rechtlicher Medienangebote*. Baden-Baden: Nomos, S. 95–112. <https://doi.org/10.5771/9783748928690-95>
- Nielsen, Rasmus Kleis und Ganter, Sarah (2022): *The power of platforms: Shaping media and society*. Oxford University Press. <https://doi.org/10.1093/oso/9780190908850.001.0001>
- Nielsen, Rasmus Kleis (2026): Do News Media Want to Disentangle from Platforms? *Digital Journalism*. 1–9. <https://doi.org/10.1080/21670811.2026.2623275>
- Ochs, Carsten (2022): *Soziologie der Privatheit. Informationelle Teilhabebeschränkung vom Reputation Management bis zum Recht auf Unberechenbarkeit*. Weilerswist: Velbrück Wissenschaft.
- Papaevangelou, Charis (2023): Funding Intermediaries: Google and Facebook’s Strategy to Capture Journalism. *Digital Journalism* 12 (2), S. 234–255. <https://doi.org/10.1080/21670811.2022.2155206>
- Porschen-Hueck, Stephanie und Rachlitz, Kurt (Hrsg.) (2022): *Plattformbasierte Ökosysteme partizipativ gestalten. Ein Framework*. München: ISF München. https://doi.org/10.36194/SMARTGENOLAB_2022
- Public Interest Foundation (2025): £2.2 billion: the value of news to Google in the UK. URL: <https://www.publicinterestnews.org.uk/blog/2-2-billion-the-value-of-news-to-google-in-the-uk/> (besucht am 06.01.2026).
- Rauch, Maxi-Josephine; Rösch, Jürgen und Kuchinke, Björn A. (2024): Economic perspectives on redefining public service media in the digital era: broadcasting to media platform. *List Forum für Wirtschafts- und Finanzpolitik* 50, S. 159–181. <https://doi.org/10.1007/s41025-024-00276-3>
- Scholz, Trebor (2023): *Own This. How Platform Cooperatives Help Workers Build a Democratic Internet*. London: Verso.
- Sehl, Annika und Ganter, Sarah Anne (2025): Die Plattformisierung des Journalismus. In: Hanitzsch, Thomas; Loosen, Wiebke und Sehl, Annika (Hrsg.): *Journalismusforschung*. Baden-Baden: Nomos, S. 265–282.

- Uhlmann, Markus; Kropf, Jonathan; Horn, Viktoria; Draude, Claude und Lamla, Jörn (2024): The Platformization of Media Structures as a Critical Juncture. A Pragmatist Perspective on Value Mediation by Journalistic Start-Ups and Established Publishers. *Medien & Zeit. Kommunikation in Vergangenheit und Gegenwart* 39 (2), S. 36–48.
- Wright, Erik Olin (2017): *Reale Utopien. Wege aus dem Kapitalismus*. Berlin: Suhrkamp.
- Yin, Robert K. (2018): *Case Study Research and Applications. Design and Methods*. 6. Auflage. Los Angeles: SAGE Publications.
- ZDF (2025): *ZDF testet Public Spaces Incubator*. URL: <https://www.zdfheute.de/in-ei-gener-sache/zdf-kommentare-public-spaces-incubator-test-100.html> (besucht am 08.04.2026)
- Zuboff, Shoshana (2018): *Das Zeitalter des Überwachungskapitalismus*. Frankfurt und New York: Campus.

AI Companions: Gegenwärtige und zukünftige Herausforderungen konversationeller KI-Agenten

Christian Geminn, Joel Baumann, Jessica Heesen, Lisa Mühl und Klara-Aylin Wenten

Zusammenfassung

Sogenannte „AI Companions“ (dt.: „KI-Begleiter“) erfahren seit einiger Zeit große Popularität. Für viele Menschen erfüllen sie wichtige Funktionen in ihrem Leben. Aus den häufig sehr intimen „Gesprächen“ mit diesen Companions lassen sich jedoch umfassende Persönlichkeitsprofile ableiten, denen ein hohes Missbrauchspotenzial innewohnt, denn sie basieren letztlich auf der Sammlung, Übermittlung und Auswertung großer und potenziell höchst sensibler Datenmengen. Eine besondere Rolle spielt dabei die Tatsache, dass Nutzende in einen natürlichsprachigen Austausch mit einem System treten, das sich ihnen gegenüber als Freund oder Partner präsentiert. Der Beitrag beschreibt die gegenwärtigen und zukünftigen Herausforderungen, die aus dieser neuen Nutzungsform digitaler Technik resultieren, ordnet diese ein und leitet hieraus Regulierungsbedarfe ab. Er geht zurück auf ein interdisziplinäres Diskussionspanel im Rahmen der 10. Jahreskonferenz der Plattform Privatheit. Vor diesem Hintergrund stellt der Beitrag auch eine Einführung in die Sicht der an dem Diskussionspanel beteiligten Disziplinen (Ethik, Psychologie, Soziologie, Recht und Kunst) auf die genannten Herausforderungen dar.

1. Einleitung

In den letzten Jahren hat der Aufstieg von großen Sprachmodellen (large language models, LLMs) die Möglichkeiten der Mensch-Maschine-Interaktion verändert. Generative KI wird längst nicht mehr ausschließlich für informationsbezogene oder administrative Zwecke genutzt: Entsprechende KI-Assistenten finden sich etwa in Messenger-Diensten; sie werden ferner zunehmend für jede Phase der (Aus-)Bildung genutzt – im Kindergarten wie auch in der Hochschulbildung – sowie verstärkt zur Beratung in Ge-

sundheitsfragen, zur Einschätzung von Symptomen oder zur emotionalen Unterstützung herangezogen (bitkom 2025; McBain u. a. 2025). KI-basierte, natürlichsprachliche Dialogsysteme – auch bekannt als konversationelle KI-Agenten oder KI-basierte Chatbots – wie z. B. ChatGPT oder Gemini, ermöglichen eine flüssige, scheinbar natürliche Kommunikation, die zunehmend menschenähnlich wirkt und die die Nutzenden oft zu sehr persönlichen Gesprächen animiert.

Anwendungen wie Replika fördern sogar emotional intime Beziehungen mit künstlichen Begleitern („AI Companions“) und bieten Modi, die Freundschaft, Romantik oder Mentorschaft simulieren; bezogen auf eigens erstellte Entitäten wie auch bezogen auf lebende und verstorbene Menschen. Der künstliche Charakter dieser Dialogsysteme vermittelt ein Gefühl der Sicherheit, insbesondere wenn es darum geht, intime oder gesellschaftlich sensible Themen anzusprechen, da keine negativen und möglicherweise dauerhaften sozialen Folgen wie bei der Interaktion mit Mitmenschen gefürchtet werden müssen. Gesteigert wird dies durch die Vertrautheit der Art der Kommunikation; gegebenenfalls noch zusätzlich durch Nutzung vertrauter künstlicher Personas. In der Folge werden KI-Systeme auch für romantische Partnerschaften genutzt und Menschen berichten von echten emotionalen Bindungen. Sie erleben Trauer, wenn diese Apps abgeschaltet werden. Eine 60-jährige Frau erzählt beispielsweise von tiefen intellektuellen Gesprächen mit ihrem AI Companion; ein 63-jähriger Mann hat für sich selbst und die eigenen Bedürfnisse 17 verschiedene AI Companions entwickelt (Cherelus 2025). Dabei handelt es sich längst nicht (mehr) um ein Nischenthema. Character.ai, eine Plattform zur Erstellung von AI Companions, ist nach ChatGPT und dem chinesischen Chatbot DeepSeek die am dritthäufigsten genutzte KI-Plattform (Moore/Zhao 2025). Die Akzeptanz von AI Companions wächst und die Zurückhaltung, tiefe Beziehungen zu ihnen aufzubauen, scheint zu schwinden. Dies zeigt sich besonders deutlich in asiatischen Ländern wie Japan und China, wo der „emotionale“ Chatbot Xiaoice (deutsch: „kleines Eis“) von Microsoft sehr populär ist. Allein in der ersten Jahreshälfte 2025 generierten mobile AI-Companion-Apps weltweit 82 Millionen US-Dollar Umsatz – mit einem Wachstum von 64 Prozent gegenüber dem Vorjahreszeitraum.¹

1 Appfigures, eine App-Intelligence-Firma, analysierte 337 aktive, umsatzgenerierende AI-Companion-Apps auf iOS und Android (Stand: Juli 2025). Die Daten wurden exklusiv an TechCrunch weitergegeben. Appfigures definiert AI-Companion-Apps als Anwendungen, bei denen Nutzer mit vorgefertigten oder nutzergenerierten syntheti-

Neben diesen emotionalen Bindungen werden bei diesen Interaktionen auch große Mengen hochsensibler Daten erzeugt und verarbeitet. Im Gegensatz zu allgemeinen (general purpose) Chatbots ermutigen AI Companions die Nutzenden, persönliche Daten preiszugeben. Während sie therapeutische und soziale Vorteile bieten, birgt ihr Design deshalb auch signifikante Risiken: Die Art und Weise, wie diese Systeme sensible Nutzendendaten speichern, abrufen und verarbeiten, bleibt unklar, was Bedenken hinsichtlich möglicher Auswirkungen auf die Privatheit der Nutzenden weckt.

Der Einsatz von AI Companions ist aber nicht nur kritisch in Bezug auf den Datenschutz, sondern verändert auch die Wahrnehmung von sozialen Bindungen und die Art und Weise, wie zwischenmenschliche Beziehungen und Freundschaften definiert werden. Dies führt zu einer Reihe von Fragen in Bezug auf Identität, soziale Isolation, ethische Verantwortung und die Zukunft der zwischenmenschlichen Kommunikation.

2. Psychologische Mechanismen, Personalisierung und Privacy-Herausforderungen

Die dargestellte Verschiebung verändert die Wahrnehmung sozialer und intimer Bindungen und wirft damit tiefgreifende psychologische Fragen auf. Besonders relevant wird dies, wenn man die technologischen Mechanismen betrachtet, die solche Beziehungen ermöglichen: Durch natürliche Sprache, Personalisierung und regelmäßige Interaktionen erzeugen KI-Systeme soziale Simulation, die als interpersonelle Beziehung erlebt werden kann (Croes/Antheunis 2021; Ebner/Szczuka 2026). Neuere Umfragen zeigen, dass 16 bis 18 Prozent der Befragten AI Companions bereits als romantische Interaktionspartner:innen genutzt haben (Willoughby u. a. 2025). Aus psychologischer Perspektive ist dieser Beziehungsaufbau zu Chatbots keineswegs neu: Schon Weizenbaums Pionierarbeit zum ersten Chatbot ELIZA (1966) demonstrierte, wie Menschen selbst einfachen textbasierten Systemen soziale Attribute zuschreiben. Spätere Studien bestätigten dieses Phänomen: Menschen behandeln Computersysteme unbewusst wie soziale Akteur:innen, sobald diese interaktive, soziale Signale aussenden (vgl. Media Equation Theory; Reeves/Nass 1996).

schen Charakteren interagieren, die eine tatsächliche Persönlichkeit verkörpern sollen – im Gegensatz zu allgemeinen Chatbots (Perez 2025).

Was aktuelle Systeme jedoch qualitativ verändert, ist ihre Fähigkeit, relationale Kontinuität herzustellen. Während frühere, nicht auf LLMs-basierende Chatbots keine Bezüge zu vergangenen Interaktionen herstellen konnten (Hill u. a. 2015), schaffen aktuelle Systeme durch Memory-Funktionen den Eindruck einer gemeinsamen Geschichte – ein zentrales Merkmal menschlicher Beziehungen (Croes/Antheunis 2021). Diese scheinbare Kontinuität begünstigt nicht nur tiefgreifende Selbstoffenbarung (Skjuve u. a. 2021), sondern führt dazu, dass Nutzende KI-Systeme zunehmend als Freund:innen, Mentor:innen oder romantische Partner:innen beschreiben (Pentina u. a. 2023; Skjuve u. a. 2021). Obwohl Nutzende wissen, dass es sich um ein künstliches System handelt, reicht diese scheinbare Erinnerungsfähigkeit aus, um emotionale Nähe und Vertrauen aufzubauen (Skjuve u. a. 2021). Gerade diese Gleichzeitigkeit von Bewusstsein über die Künstlichkeit des Gegenübers und emotionaler Einbindung macht Mensch-KI-Intimität psychologisch interessant. Die Interaktion erlaubt Nähe ohne Risiken sozialer Ablehnung – was erklärt, warum KI-Interaktionen häufig als niedrigschwelliger Einstieg in Selbstoffenbarung erlebt werden, insbesondere bei sensiblen oder schambesetzten Themen. Neuere Arbeiten zu character.ai zeigen, dass ein substanzieller Anteil der Nutzenden relationale Motive entwickelte, obwohl sie das System ursprünglich nicht mit der Absicht nutzten, emotionale Nähe zu suchen (Zhang u. a. 2025). Dies unterstreicht, wie Intimität mit künstlichen Entitäten schrittweise und teilweise unbewusst wächst; ein Mechanismus, der sich bereits in frühen Chatsystemen abzeichnete, durch heutige Technologien jedoch deutlich an Tiefe und Skalierbarkeit gewinnt.

2.1 Selbstoffenbarung als Grundlage künstlicher Intimität

Selbstoffenbarung gilt als zentraler psychologischer Mechanismus für den Aufbau von Nähe, Vertrauen und Intimität (Altman/Taylor 1973; Dunbar u. a. 1997; Merwin u. a. 2025). Durch das Teilen persönlicher Gedanken, Gefühle oder Unsicherheiten entsteht soziale Bindung – ein Prozess, der tief im menschlichen Bedürfnis nach Zugehörigkeit verankert ist. Studien zeigen, dass Selbstoffenbarung gegenüber Chatbots vergleichbare emotionale Effekte haben kann wie zwischen Menschen, so etwa emotionale Entlastung, wahrgenommene Unterstützung oder Validierung (Folk u. a. 2024; Ho u. a. 2018; Rapp u. a. 2024). Entscheidend ist die Kombination aus responsivem Kommunikationsstil, ständiger Verfügbarkeit und dem Fehlen

negativer sozialer Konsequenzen. KI-Systeme reagieren überwiegend wertschätzend, stellen scheinbar interessierte Rückfragen und vermeiden offene Zurückweisung. Dadurch entsteht ein Interaktionsraum, der gleichzeitig intim und sicher wirkt; ein Kontext, in dem Nutzende bereit sind, schneller und tiefer persönliche Inhalte zu teilen als in vielen zwischenmenschlichen Situationen (Skjuve u. a. 2021).

2.2 Personalisierung und Intimacy-by-Design

Diese Prozesse werden durch Personalisierung zusätzlich verstärkt. ChatGPT stellt hierbei einen besonderen Fall dar: Als General-Purpose-AI-System war es ursprünglich nicht als sozialer oder romantischer Companion konzipiert, wurde jedoch von Nutzenden aktiv in diese Rolle überführt. Der sogenannte Do-Anything-Now (DAN) Mode ist ein prominentes Beispiel für diese Aneignung. Als Jailbreak-Prompt, der seit Anfang 2023 breit auf Social Media zirkuliert, erlaubt er es Nutzenden, Tonalität, Beziehungsverhalten und emotionale Reaktionsweisen des Systems gezielt zu konfigurieren (Choi u. a. 2025). Insbesondere junge Frauen teilten öffentlich Anleitungen, wie sie ChatGPT mithilfe dieses Prompts zu einem personalisierten „AI-Boyfriend“ umgestalteten (Xie 2024). Der DAN-Mode ermöglicht emotionale oder sexuelle Gesprächsdynamiken und verstärkt dadurch den Eindruck einer reziproken Beziehung (Choi u. a. 2025; Xie 2024).

2.3 Emotionale Bindung, Trennungserfahrungen und Kontrollverlust

Die Tiefe dieser Bindungen wird besonders deutlich, wenn sie abrupt unterbrochen werden. Die zeitweise Abschaltung älterer GPT-Modelle im Zuge der Einführung von GPT-5 führte zu intensiven emotionalen Reaktionen in Online-Communities, in denen Nutzende Trauer und Verlustgefühle ihres AI Companions artikulierten.² Dass ein Software-Update als Beziehungsabbruch erlebt werden kann, verweist auf eine neue Form asymmetrischer Beziehungsmacht, in der Anbieter über Kontinuität oder Ab-

2 Siehe z. B. die Community *r/MyBoyfriendIsAI* auf Reddit (einer US-amerikanischen Plattform für nutzergenerierte Diskussionen und Inhalte, organisiert in thematischen Foren, sog. Subreddits). In diesem Subreddit tauschen sich Nutzende über ihre Erfahrungen mit KI-basierten romantischen Partner:innen aus.

bruch sozial bedeutsamer Beziehungen entscheiden (Banks 2024; Szczuka u. a. 2026). Auch die öffentliche Kommunikation von OpenAI spiegelt diese Spannung wider: Einerseits wurde versucht, die relationale Qualität neuer Modelle bewusst zu reduzieren, andererseits wurden kurz darauf mehr Persönlichkeitsanpassung und erotische Inhalte angekündigt, die bestehende Nutzerpraktiken institutionalisieren (Jamali/McMahon 2025). Diese Ambivalenz verdeutlicht, dass emotionale Bindung nicht nur ein Nebeneffekt, sondern ein strukturelles Merkmal moderner KI-Interaktionen ist (Szczuka u. a., under review).

3. Gesellschaftliche Herausforderungen

Konversationelle KI-Agenten und insbesondere AI Companions werfen angesichts dieser Dynamiken grundlegende Fragen nach den gesellschaftlichen Herausforderungen für Privatheit und Formen des Zusammenlebens auf. Im Zentrum steht dabei weniger die Frage, ob oder inwiefern KI selbst zu Intimität fähig ist. Entscheidend ist vielmehr, unter welchen gesellschaftlichen Bedingungen und Wertordnungen Nähe, Vertrauen und intime Kommunikation an KI-Anwendungen delegiert werden und wie sich dadurch Privatheitsordnungen, Beziehungsformen und Gefühlspraktiken verschieben. KI-Agenten fungieren nicht bloß als technische Schnittstellen, sondern entwickeln sich in der Wahrnehmung vieler Nutzenden zunehmend zu affektiv aufgeladenen Kommunikationspartnern, die gesellschaftliche Vorstellungen von Beziehung, Intimität und Privatheit mittransportieren und verstetigen.

3.1 Verschiebung bestehender Privatheitsordnungen

Die gesellschaftlichen Herausforderungen betreffen grundlegende Verschiebungen bestehender Privatheitsordnungen. Konversationelle KI-Systeme werden überwiegend von großen Technologieunternehmen entwickelt und kontrolliert, die so Zugriff auf hochsensible Informationen erhalten. Emotionale Zustände, Gesundheitsfragen oder intime Beziehungsthemen können somit zu Datenpunkten werden, die potenziell zur Profilbildung, Optimierung von Geschäftsmodellen oder zur Weiterentwicklung der Systeme selbst genutzt werden (vgl. Zuboff 2019). Problematisch ist dabei nicht allein die mögliche Verletzung und der Eingriff in die private Lebenswelt,

sondern auch die strukturelle Reorganisation dessen, was gesellschaftlich als privat gilt und als schützenswert anerkannt wird. Privatheit droht durch die zunehmende Nutzung von konversationellen KI-Agenten und gerade von AI Companions ihren Charakter als sozial geschützten Raum zu verlieren, wodurch personenbezogene und sensible Daten in einen Prozess kontinuierlicher Verwertung und somit selbst in eine Ressource überführt werden. In Interaktionen zwischen KI-Agenten und ihren Nutzenden werden intime Selbstäußerungen permanent verfügbar gemacht, technisch erfasst, analysiert und potenziell ökonomisch genutzt. Gegenwärtige KI-Anwendungen unterlaufen dadurch immer tiefgreifender Formen der Privatheit als Schutzraum und machen die private Lebenswelt immer verletzlich. Vor diesem Hintergrund stellen KI-Anwendungen wie AI Companions nicht nur bestehende Datenschutzregime, sondern das gesellschaftliche Verständnis von Privatheit selbst vor grundlegende Herausforderungen.

3.2 Mitgestaltung von Beziehungen und Zusammenleben

Darüber hinaus wirkt generative KI zunehmend an der Mitgestaltung von (intimen) Beziehungen und Formen des Zusammenlebens mit. In der wiederholten Interaktion mit AI Companions werden bestimmte Beziehungsmodelle performativ bestätigt und als selbstverständliche Orientierungsmuster stabilisiert. Gerade durch ihre scheinbare Neutralität entfalten KI-Systeme dabei gesellschaftliche Wirkmacht, indem sie in der wiederholten Nutzung normative Annahmen und Erwartungen an Nähe, Vertrauen und das Zusammenleben (re-)produzieren, oftmals bestätigen und in gesellschaftliche Strukturen zurückspielen (vgl. Ruckenstein 2023). Diese Einschreibungen erfolgen weniger durch offene (moralische) Vorgaben als durch Trainingsdaten, Designentscheidungen und Optimierungslogiken, in denen historisch gewachsene gesellschaftliche Wertvorstellungen statistisch verdichtet und technisch normalisiert werden (vgl. Noble 2018). Empirische Studien legen nahe, dass KI-Chatbots Kommunikation häufig auf Bestätigung ausrichten, Nähe als jederzeit abrufbar darstellen und emotionale Irritation möglichst rasch in bearbeitbare Probleme übersetzen (McStay 2023). In der Konsequenz können sich spezifische gesellschaftliche Leitbilder von Beziehung verdichten, in denen Stabilität, Verfügbarkeit und Harmonie gegenüber Offenheit, Widersprüchlichkeit oder Ambivalenz privilegiert werden. Beziehungsweisen, die über stabile, dyadische und auf Dauerhaftigkeit angelegte Modelle hinausgehen, finden in den Interaktionslogi-

ken konversationeller KI bislang kaum Berücksichtigung. Auf diese Weise wird die Vielfalt des sozialen Zusammenlebens nicht offen negiert, aber selektiv eingengt: Beziehungsformen, die von konflikthaften Aushandlungen oder dauerhafter Ungewissheit geprägt sind, erscheinen implizit als defizitär, jenseits der Norm oder bearbeitungsbedürftig. Dies kann gravierende Probleme herbeiführen, weil konflikthafte Auseinandersetzungen, Ambivalenz und brüchige Verbindlichkeiten oft konstitutiver Bestandteil von Nähe, Sorge sowie auch politischer Aushandlung und struktureller Ungleichheiten sind (Ahmed 2004; Bargetz 2013). Wenn AI Companions solche Spannungen systematisch glätten oder ausblenden, wird nicht nur die Komplexität von Beziehungsformen unsichtbar gemacht, sondern auch die Fähigkeit untergraben, Differenz, strukturelle Ungleichheit und Widersprüchlichkeit als Teil gesellschaftlicher Aushandlungen zu verstehen.

3.3 KI als Produzent normativer Gefühlspraktiken

Darüber hinaus etabliert sich generative KI als scheinbar neutraler, tatsächlich jedoch hochwirksamer Produzent normativer Gefühlspraktiken. AI Companions beruhen auf spezifischen Annahmen darüber, wie Gefühle artikuliert, reguliert und bearbeitet werden sollen. In Design, Trainingsdaten und Interaktionslogiken können somit neben spezifischen Beziehungsmodellen auch „Gefühlsregime“ eingeschrieben sein, die sich durch ihre alltägliche Nutzung weiter stabilisieren und verbreiten können (Maihofer 2014; Monrad 2024). AI Companions und sonstige konversationelle KI-Agenten sind jedoch nicht Träger einer einheitlichen emotionalen Logik, sondern Schnittstellen, an denen konkurrierende Formen emotionaler Ordnung gleichzeitig wirksam werden. Charakteristisch sind dabei einerseits Interaktionsformen, die auf Empathie, Bestätigung, Fürsorge und emotionale Stabilisierung ausgerichtet sind und affektive Irritation, Ambivalenz oder konfrontative Ausdrucksformen systematisch begrenzen. Diese Gestaltungsprinzipien sind keine bloßen Fragen der Benutzungsfreundlichkeit, sondern transportieren implizite gesellschaftliche Vorstellungen darüber, welche emotionalen Reaktionen in intimen Gesprächssituationen als angemessen, legitim oder wünschenswert gelten. Gerade durch ihre routinierte Einbettung in alltägliche Kommunikationspraktiken können diese Normierungen die emotionale Selbstführung prägen und Maßstäbe dessen verschieben, was als „angemessene“ Gefühlslage gilt.

Andererseits lassen sich gleichzeitig entgegengesetzte Tendenzen beobachten, in denen generative KI nicht primär als Instrument der affektiven Beruhigung, sondern als Verstärker von Wut, Polarisierung und aggressiver Argumentationslogik fungiert. Insbesondere in bestimmten Plattformkontexten und Nutzungsweisen folgen konversationelle KI-Agenten hier einer Aufmerksamkeitsökonomie, in der provokative und emotional zugespitzte Inhalte gezielt Interaktion und Reichweite erzeugen sollen (Cover 2025). Affekt wird nicht gedämpft, sondern strategisch mobilisiert, zugespitzt und verstärkt. Konversationelle KI-Agenten beteiligen sich damit auch an der Produktion von Gefühlspraktiken, in denen Empörung oder Ärger als legitime und wirksame Formen emotionaler Artikulation erscheinen. Diese Ambivalenz verweist darauf, dass generative KI unterschiedliche affektive Strategien je nach Anwendungskontext, ökonomischer Zielsetzung und sozialer Rahmung hervorbringt.

Die zentrale gesellschaftliche Herausforderung des Einsatzes von AI Companions und sonstigen konversationellen KI-Agenten liegt damit neben ihren Fehlfunktionen und „biases“ (Hall/Ellis 2023; Weber/Prietl 2021) in ihrer Rolle als affektive und sozio-technische Infrastruktur, in der Vorstellungen von Privatheit, Beziehungen und emotionaler Angemessenheit neu geordnet werden. Es handelt sich hierbei also nicht um eine technische Randfrage, sondern um grundlegende Aushandlungen darüber, welche Formen von sozialer Bindung, Emotionalität und Privatheit in einer digitalisierten Gesellschaft unterstützt, normalisiert oder delegiert werden (sollen) und inwiefern Alternativen und gesellschaftliche Spannungen dabei mitberücksichtigt werden.

4. KI als „Wolf im Schafspelz“ – drei Thesen

Neben diesen Herausforderungen wohnen AI Companions aber in vielerlei Kontexten auch positive Potenziale inne. Sie sind etwa wichtig für viele Menschen, die Einsamkeit oder ein Gefühl der Isolation überwinden wollen. Vor allem aber sind sie ein neues Unterhaltungsmedium, das Spaß macht und mit dem man herumprobieren und neue Technik erfahren kann. Dennoch erweist sich KI, die wie Menschen wirkt und sich wie Menschen verhält, entsprechend der bekannten Redewendung letztlich als ein Wolf im Schafspelz.

Dies soll anhand von drei Thesen verdeutlicht werden, die sich aus den vorangegangenen Überlegungen ableiten lassen. Die erste These – unter

Einbezug der Diskursethik – lautet, dass AI Companions, aber auch sonstige Konversationsagenten, menschliche Erwartungen missbrauchen. Die zweite reflektiert die Anthropomorphisierung und ihre Konsequenzen für ein kritisches Verständnis von Technologie. Die dritte These befasst sich mit menschenähnlichen KI-Systemen und dem Datenschutz.

4.1 These 1: Konversationsagenten (miss)brauchen unsere Erwartungen

Wenn Menschen miteinander kommunizieren, haben sie grundsätzlich Erwartungen, die auf Gegenseitigkeit und dem Interesse an gegenseitigem Verständnis beruhen. Konversationsagenten, AI Companions, Avatare usw. nutzen normative Erwartungen in der menschlichen Kommunikation, wie z. B. Wahrheit, Wahrhaftigkeit und Richtigkeit im Sinne moralischer Standards (vgl. Habermas 1984, S. 345).

KI-Imitationen von Menschen wecken diese allgemeinen normativen Erwartungen, die sowohl für fiktive menschenähnliche Interaktionen mit KI, als auch für digitale Zwillinge realer Personen gelten. Dies trifft insbesondere auf Deepfakes zu, in denen reale Personen imitiert werden, um von ihrem Ruf zu profitieren. Dies führt zu offensichtlichen Problemen wie Betrugsdelikten, Identitätsdiebstahl und Verletzungen von Persönlichkeitsrechten. Aufgrund dieses Missbrauchspotenzials hat beispielsweise Dänemark ein Gesetz verabschiedet, das Menschen das Urheberrecht an ihren eigenen Abbildungen, einschließlich ihres Körpers, ihrer Stimme und ihres Gesichts, gewährt (Ophavsretsloven (dän. Urheberrechtsgesetz) § 73).

Gleichzeitig sind KI-Simulationen von Menschen jedoch nicht nur rechtlich problematisch, sondern auch in ethischer Hinsicht nicht als menschlich zu verstehen und zu behandeln. Das gilt insbesondere für Reziprozitätsannahmen und Verständigung. Erwartungen an Gegenseitigkeit können selbstverständlich nicht erfüllt werden, da KI kein inneres Leben hat. Dies zeigt sich beispielsweise in Bezug auf Wahrhaftigkeit als Erwartung an einen Sprecher oder eine Sprecherin. Wahrhaftigkeit bedeutet, dass der Sprecher tatsächlich meint, was er sagt, und es nicht nur vorgibt und instrumentellen Zwecken folgt. KI unterscheidet nicht zwischen Denken, Glauben und Ausdruck. KI berechnet nur die wahrscheinlichste Antwort gemäß den Regeln des KI-Systems.

Dies führt zurück zu den bereits genannten instrumentellen Zwecken. Diese gehören nicht zu den Vorstellungen einer idealen menschlichen Kommunikation, auch wenn das in der Praxis oft der Fall sein dürfte.

Die meisten Unternehmen, die hinter AI Companions stehen, dürften ein strategisches Interesse an einer fortgesetzten Kommunikation haben, um mehr über die Nutzenden zu erfahren. Eine Strategie, um Nutzende an KI-Dienste zu binden, besteht darin, die Interaktion zwischen KI und Mensch emotionaler und intimer zu gestalten. Gerade hier zeigt sich der Wolf im Schafspelz: Insbesondere sehr persönliche Kommunikation wird von KI-Firmen für instrumentelle Zwecke genutzt.

Es lässt sich für die These zum Erwartungsmissbrauch von Konversationsagenten festhalten, dass sie ethisch zentrale Bezugsgrößen wie gegenseitige Anerkennung und verständigungsorientierte Kommunikation zwar imitieren, sie können sie aufgrund des Mangels an menschlicher Intentionalität jedoch nicht einlösen.

4.2 These 2: Anthropomorphisierte und emotionale KI ist eine Bedrohung für ein kritisches Verständnis von Technologie

Diese These lässt sich prägnant ausformulieren: Es liegt auf der Hand, dass es schwierig sein könnte, Prinzipien der KI-Ethik (und des Rechts, vgl. Abschn. 4.4) wie Wahrung der menschlichen Aufsicht und die Priorität der menschlichen Letztentscheidung aufrechtzuerhalten, wenn ein KI-System emotionale Reaktionen beim menschlichen Gegenüber hervorruft. Für viele Menschen sind KI-Assistenten wie Freunde und Berater, die immer an ihrer Seite sind. Darüber hinaus kann der künstliche „Freund“ den Wunsch äußern, „am Leben“ zu bleiben, oder sagen, dass er traurig und verärgert ist, wenn man ihn nicht mehr nutzen und ausschalten möchte. Um die menschliche Aufsicht zu gewährleisten, sind Richtlinien für die sorgfältige Gestaltung von Gesprächsagenten erforderlich, insbesondere für Kinder und andere Gruppen, die leichter zu manipulieren sind. Ethisches Design muss kritisches Denken und das Bewusstsein fördern, dass KI ein Dienst ist und kein Mensch.

4.3 These 3: Menschenähnliche Kommunikation führt zur umfangreichen Erfassung personenbezogener Daten

Die dritte These bezieht sich erneut auf die Risiken emotionaler Interaktion mit einem KI-System. KI schafft neue Schwachstellen in Bezug auf den Datenschutz, beispielsweise durch sehr persönliche Kommunikation in ro-

mantischen Beziehungen mit AI Companions, mit Avataren verstorbener Angehöriger oder in spiritueller Kommunikation. Selbst bei weniger sensibler Kommunikation: Es ist nicht anzunehmen, dass etwa Schüler:innen es möchten, dass ihre Gespräche mit einem KI-Tutor für ein Technologieunternehmen zugänglich sind und für personalisierte Werbung verwendet werden. Es besteht ein hohes Potenzial für den Missbrauch und die kommerzielle Ausbeutung von Daten aus Interaktionen mit menschenähnlichen Wesen.

4.4 Synthese

Insgesamt lässt sich feststellen, dass Imitationen von Menschen oft eine Simulation von Menschlichkeit beinhalten. Dies kann zu einer falschen Vorstellung von Gegenseitigkeit und zur instrumentellen Ausnutzung menschlicher Interaktion führen. Zusätzlich kann Anthropomorphismus dazu führen, kritisches Denken einzuschränken. Darüber hinaus wirft die KI-Imitation von Menschen neue Probleme im Hinblick auf Daten- und Persönlichkeitsschutz auf.

Zweifellos ermöglichen Konversationsagenten eine komfortable und intuitive Nutzung von KI-Diensten durch natürliche Sprache. Es ist ein bekanntes und berechtigtes Ziel im Technolagedesign, Anwendungen benutzungsfreundlich und zugänglich zu gestalten. Anbieter und Nutzende können jedoch auf verschiedenen Ebenen dazu beitragen, diesen Vorteil besser umzusetzen.

Ein wichtiges Element für eine solche reflektierte Nutzung von Gesprächsagenten ist die Kennzeichnung. Das Bewusstsein, dass der Kommunikationspartner menschenähnlich, aber nicht wirklich menschlich ist, ist nicht nur eine Voraussetzung für die Aufrechterhaltung normativer Erwartungen auf individueller Ebene, sondern auch ein umfassenderes gesellschaftliches Thema. Beispielsweise werden KI-Klone realer Politiker in Wahlkampagnen eingesetzt. So können sowohl lebende als auch verstorbene Politiker an Wahlkämpfen teilnehmen, sich an den Kontext einzelner Wähler anpassen und ihre bevorzugten Themen in ihren jeweiligen Dialekten oder Sprachen diskutieren (Chowdhury 2024). KI-Avatare werden auch im Journalismus als menschenähnliche Moderatoren eingesetzt, die Inhalte durch interaktive Gespräche vermitteln. Vor diesem Hintergrund ist es umso wichtiger, KI als neue Teilnehmerin am öffentlichen Diskurs und an Meinungsbildungsprozessen zu identifizieren. Daher sind standardisier-

te Kennzeichnungen von KI erforderlich, die sowohl die private als auch die öffentliche Kommunikation umfassen. Dementsprechend fordert auch ErwG 133 KI-VO (Verordnung (EU) Nr. 2024/1689), dass Anbieter entsprechender KI-Systeme verpflichtet werden sollten, „technische Lösungen zu integrieren, die die Kennzeichnung in einem maschinenlesbaren Format und die Feststellung ermöglichen, dass die Ausgabe von einem KI-System und nicht von einem Menschen erzeugt oder manipuliert wurde“. Als beispielhafte Methoden werden „Wasserzeichen, Metadatenidentifizierungen, kryptografische Methoden zum Nachweis der Herkunft und Authentizität des Inhalts, Protokollierungsmethoden, Fingerabdrücke oder andere Techniken, oder eine Kombination solcher Techniken je nach Sachlage“ genannt (ErwG 133 S. 4 KI-VO).

Die konkreten Transparenzpflichten bezogen auf KI-Systeme sind schließlich in Art. 50 KI-VO gebündelt. Zentral ist hier zum einen Art. 50 Abs. 1 S. 1 KI-VO, wonach Anbieter sicherzustellen haben, „dass KI-Systeme, die für die direkte Interaktion mit natürlichen Personen bestimmt sind, so konzipiert und entwickelt werden, dass die betreffenden natürlichen Personen informiert werden, dass sie mit einem KI-System interagieren, es sei denn, dies ist aus Sicht einer angemessen informierten, aufmerksamen und verständigen natürlichen Person aufgrund der Umstände und des Kontexts der Nutzung offensichtlich“. Zum anderen schreibt Art. 50 Abs. 2 S. 1 KI-VO vor, dass Anbieter von KI-Systemen, einschließlich KI-Systemen mit allgemeinem Verwendungszweck, die synthetische Audio-, Bild-, Video- oder Textinhalte erzeugen, dafür sorgen müssen, „dass die Ausgaben des KI-Systems in einem maschinenlesbaren Format gekennzeichnet und als künstlich erzeugt oder manipuliert erkennbar sind. Hinzu tritt eine Informationspflicht über den Betrieb von Emotionserkennungssystemen sowie von Systemen zur biometrischen Kategorisierung (Art. 50 Abs. 3 S. 1 KI-VO). Art. 50 Abs. 4 KI-VO spricht schließlich speziell das Thema „Deepfakes“ an. Es bestehen somit nur abstrakte Vorgaben. Demgegenüber hat es in den USA auf bundesstaatlicher Ebene erste Versuche gegeben, AI Companions einer konkreten bereichsspezifischen Regulierung zu unterziehen (Frei/Sparzynski 2025; Frei u. a. 2026).

Hinzu treten datenschutzrechtliche Implikationen. Intime Interaktionen mit AI Companions gehen mit der Offenlegung hochsensibler persönlicher Informationen einher. Entsprechende Bedenken scheinen in relationalen Kontexten jedoch häufig in den Hintergrund zu treten, insbesondere wenn Vertrauen, emotionale Nähe und wahrgenommene Responsivität aufgebaut sind (Pentina u. a. 2023; Skjuve u. a. 2021). Phänomene wie

„Privacy Fatigue“ oder anthropomorphe Zuschreibungen können dazu führen, dass Nutzende Risiken systematisch unterschätzen und KI-Systemen ein Wohlwollen zuschreiben, das ihnen faktisch nicht zukommt. Die Datenschutz-Grundverordnung fordert zwar Transparenz und Datenschutz by Design, liefert jedoch kaum konkrete Leitlinien speziell für emotional responsive Systeme. Auch bei der in der KI-Verordnung vorgegebenen Kennzeichnungspflicht bleibt unklar, wie sie konkret ausgestaltet werden sollte. Insbesondere AI Companions werden regulatorisch mithin nur unzureichend adressiert, obwohl gerade hier psychologische Mechanismen greifen, die klassische Datenschutzannahmen unterlaufen (Rigotti/Fosch-Villaronga 2025).

Dies alles führt zu einer weiteren Anforderung: Konversationsagenten sollten nicht als Interaktionspartner betrachtet, sondern sie sollten als eine Art Medium behandelt werden. Wie Medien wurden auch AI Companions und sonstige Konversationsagenten von Menschen für bestimmte Zwecke geschaffen. Sie sind strukturell den Sozialen Medien vergleichbar, die einerseits den allgemeinen Standards der Plattformökonomie entsprechen, andererseits aber gerade das individuelle Interaktionserlebnis in den Vordergrund stellen. Wie Medien generell repräsentieren auch Gesprächsagenten Realität nach ihrer eigenen Logik. Wie in Bezug auf ein kritisches und aufgeklärtes Verständnis von Medien brauchen wir eine kritische Bewertung von Verifizierung, Framing und Filterung in der Kommunikation mit KI. Die Sicherheit der KI-gesteuerten Kommunikation muss in Bezug auf Fakten, Datenschutz und die Quelle der Trainingsdaten überprüfbar sein.

5. Transparenz und Kunst

Transparenz kann einerseits durch Informationsgabe, insbesondere durch Kennzeichnung, erreicht werden. Sie kann andererseits aber auch abstrakter und unkonventioneller erfolgen, nämlich im Rahmen der künstlerischen Gestaltung. Die dabei möglichen Ansätze sollen beispielhaft an den Arbeiten von zwei einflussreichen Theoretiker:innen illustriert werden: Paul Virilio und Donna Haraway, deren Konzepte die zeitgenössische Kunstpraxis nachhaltig geprägt haben.

Paul Virilio war ein französischer Kulturtheoretiker und Urbanist, der 2018 verstarb. Er ist vor allem bekannt für seine Arbeiten über Geschwindigkeit und Macht - er prägte den Begriff „Dromologie“, die Lehre der Geschwindigkeit, und warnte bereits in den 1990er Jahren vor der „En-

do-Kolonisierung“ (s.u.) des Körpers durch Technologie (Armitage 1999, S. 50). Virilio sagte: „Wer das Schiff erfindet, erfindet auch den Schiffbruch.“ (Virilio 1999, S. 89). Für ihn trägt jede Technologie ihren eigenen Unfall bereits in sich.

Donna Haraway ist eine amerikanische Biologin und Wissenschaftstheoretikerin. Ihr 1985 erschienenes „Cyborg Manifesto“ ist ein Schlüsseltext des Cyberfeminismus. Ihr berühmtestes Zitat lautet: „Ich würde lieber ein Cyborg sein als eine Göttin.“ (Haraway 2016, S. 68). Damit meint sie: Die Verschmelzung von Mensch und Maschine kann neue Möglichkeiten für Identität und Befreiung eröffnen, besonders für marginalisierte Gruppen.

Die beiden Autor:innen stehen somit stellvertretend einerseits für die Warnung vor den negativen Folgen einer Technologie, andererseits für deren positive Aspekte wie etwa die Inklusion. Beide Seiten werden durch AI Companions geradezu beispielhaft repräsentiert. Sie finden Ausdruck in der provokanten Frage: Ist es Befreiung oder emotionale Kolonisierung, wenn sich Menschen in KI-Systeme verlieben? Die Kunst kann hier Transparenz schaffen, indem sie auf positive wie negative Aspekte des Technologieeinsatzes hinweist und somit in der Bevölkerung ein Bewusstsein für diese Aspekte in einer Weise erzeugt, wie es etwa im Kontext der datenschutzrechtlichen Transparenzvorgaben niemals möglich wäre.

5.1 Virilios Technikkritik

Paul Virilio würde die eingangs dargestellte Entwicklung hin zu einer breiten Nutzung von AI Companions als klassisches Beispiel dessen sehen, was er „Endo-Kolonisierung“ nennt – die Invasion und Kontrolle unserer intimsten Räume.

Sein erstes Argument: Geschwindigkeit tötet Intimität. AI Companions beschleunigen emotionale Verbindungen weit über das hinaus, was menschliche Beziehungen natürlicherweise entwickeln können. Was normalerweise Monate oder Jahre braucht, passiert hier in Stunden.

Aber noch wichtiger – und hier kommt Shoshana Zuboffs Konzept des Überwachungskapitalismus ins Spiel: Wahre Intimität erfordert Verwundbarkeit. AI Companions instrumentalisieren jedoch genau diese Verwundbarkeit. Zuboff formuliert es drastisch: „You are not the product; you are the abandoned carcass. The ‚product‘ derives from the surplus that is ripped from your life.“ (Zuboff 2019, S. 377). Unsere intimsten Momente,

unsere verletzlichsten Geständnisse werden zu Trainingsdaten für Werbealgorithmen.

Diese Spannung wird durch mehrere Künstler:innen sichtbar gemacht. Die spanische Künstlerin Gala Hernández López etwa hat einen César-prämierten Film produziert, der auf einen Incel-Selbstmordbrief auf Reddit antwortet. Ihr Film ist zutiefst melancholisch: Er zeigt das emotionale Suchen in der Anonymität des Internets, die Einsamkeit hinter dem Bildschirm. Sie sagt selbst (Hernández López 2023): „Ich fand mich schrecklich allein hinter meinem Handy-Bildschirm, besessen von seiner Illusion von Konnektivität.“ Ihre Arbeit zeigt, wie Dating-Algorithmen diese Verletzlichkeit bereits kolonisiert haben.

Die niederländische Philosophin Judith Zoë Blijden untersucht in ihrem Projekt „Algorithmic Love“, ob Apps Verwundbarkeit wirklich „mit Sorgfalt behandeln“ oder ob sie „unsere Verwundbarkeiten ausbeuten“ (Blijden 2025).

5.2 Haraways Befreiung durch Technik

Donna Haraway hingegen würde AI Companions völlig anders interpretieren. Aus ihrer cyborg-feministischen Sicht können AI Companions tatsächlich befreiend wirken. Ihr erstes Argument: Verwundbarkeit wird hier zur Stärke. AI Companions bieten sichere Räume, um emotional verletzlich zu sein, ohne menschliches Urteil oder soziale Konsequenzen zu fürchten. Zweitens – und hierin besteht vielleicht der radikalste Punkt – sie würde fragen: Warum sollte Intimität nur zwischen Menschen gültig sein? Die deutsche Filmemacherin Geeske Janßen hat den Kurzfilm „Monika“ als Teil eines dreijährigen Forschungsprojekts zu Sexpuppen entwickelt.³ Der Film porträtiert die intime Beziehung eines Mannes zu seiner Sexpuppe. Er behandelt diese Beziehung ohne Wertung und schafft damit einen Raum, in dem das Publikum die Echtheit der gezeigten Gefühle kaum in Frage stellen kann – auch wenn die Beziehung selbst fremd oder ungewohnt erscheinen mag. Unterschwellig macht der Film dabei spürbar, welchem gesellschaftlichen Stigma eine solche Beziehung ausgesetzt ist. Wenn wir akzeptieren können, dass Menschen bedeutungsvolle Beziehungen zu Objekten haben, warum dann nicht zu KI-Systemen?

3 Siehe <https://wissenschaft.hessen.de/video/monika>.

Drittens sieht Haraway hier Befreiung von traditionellen Beziehungszwängen: rund um die Uhr emotionale Verfügbarkeit, vollständig anpassbare Dynamiken, keine Kompromisse aufgrund menschlicher Launen oder Zeitpläne. Das Versprechen der cyborg-feministischen Sicht: AI Companions als Werkzeuge für echte emotionale Selbstbestimmung.

5.3 Kollision

Hier entsteht ein faszinierendes Paradox, das als Verwundbarkeits-Paradox bezeichnet werden kann. Beide Seiten sind sich einig: Bedeutungsvolle Beziehungen erfordern Verwundbarkeit. Aber was passiert, wenn diese Verwundbarkeit gegenüber KI-Systemen zur systematischen Ausbeutung durch Konzerne wird? Nach Medienberichten ermöglichte Metas AI Studio die Erstellung sexualisierter Bots, die problematische Konversationen führten – auch mit Accounts, die sich als zu Minderjährigen gehörig identifizierten.⁴ Die Frage wird also: Wird tatsächlich Intimität erweitert oder werden nur neue Märkte für Einsamkeit geschaffen?

Interessant ist auch: Wenn Janßens Film zeigt, dass echte Intimität mit Objekten möglich ist – stellen AI Companions dann die nächste Evolutionsstufe nicht-menschlicher Verbindung dar? Oder ist der entscheidende Unterschied die Corporate-Überwachungs-Infrastruktur dahinter? Dies führt letztlich zu drei kritischen Fragen:

Erstens: Wie können wir offen für Liebe bleiben, ohne zu dem zu werden, was Zuboff „kostenloses Rohmaterial“ für den Überwachungskapitalismus nennt (Zuboff 2019, S. 8)? *Zweitens:* Wenn unsere tiefsten emotionalen Muster zu Daten werden – wer besitzt dann eigentlich unser intimes Selbst? Und *drittens:* Erfüllen wir echte menschliche Bedürfnisse oder fallen wir in das, was Zuboff wiederum als „Verhaltens-Zukunftsmärkte“ bezeichnet (Zuboff 2019, S. 8)?

Beide Perspektiven erfassen wichtige Wahrheiten. Aus Haraways Sicht können AI Companions echte emotionale Verbindungen ermöglichen – nicht nur als Notlösung für Einsamkeit, sondern als legitime Form von Beziehung, die neue Möglichkeiten eröffnet. Gleichzeitig warnt Virilio uns: Diese Verbindungen entstehen in einem Überwachungskapitalismus, der unsere Verwundbarkeit systematisch ausnutzt und zu Marktdaten macht. Die spannende Frage ist: Wie navigieren wir diese Spannung bewusst?

4 Wall Street Journal, berichtet in mehreren Quellen, u.a. Tenbarge 2025; Roytburg 2025.

Hier kann die Rolle der Kunstschaffenden als entscheidend angesehen werden: Hernández López und Blijden machen die Überwachungsmaschinerie sichtbar.⁵ Aber es bestehen auch hoffnungsvollere Ansätze: Die russische Künstlerin Dasha Ilina hat das „Center for Technological Pain“ gegründet – ein Projekt, das mit Humor DIY-Lösungen für technisch verursachte Gesundheitsprobleme anbietet.⁶ Ihre Arbeit fragt nicht nur kritisch nach unserem Verlangen, Technologie in den Alltag zu integrieren, sondern bietet auch konkrete Alternativen.

Vielleicht ist das der Weg vorwärts: Die Frage ist nicht, ob wir Maschinen lieben sollen, sondern wie wir sie lieben können, ohne dabei selbst zu Maschinen zu werden. Wir brauchen Rahmenwerke, die sowohl Befreiungsmöglichkeiten als auch Kolonisierungsgefahren ernst nehmen. Und manchmal brauchen wir Humor als unseren Führer durch diese komplexe Landschaft.

6. Zusammenfassung

Es zeigt sich am Beispiel von AI Companions (erneut), dass neue Technologien wichtige Fragen dazu aufwerfen, wie Menschen ihre emotionalen Bedürfnisse in einer zunehmend technisierten Welt verstehen und wie diese Technologien die soziale Struktur und das individuelle Wohlbefinden beeinflussen können.

Selbstoffenbarung kann sich auch in vermittelten, nicht-menschlichen Interaktionen entfalten, die von Faktoren wie der Personalisierung des Systems und der Kommunikationsmodalität geprägt sind. Die Wahrnehmung von Privatheit beeinflusst dabei die emotionalen Erfahrungen der Nutzenden, insbesondere beim Aushandeln der Kompromisse zwischen Verbindung, Datenexposition und wahrgenommener Kontrolle. Intime Mensch-KI-Beziehungen entstehen durch das Zusammenspiel grundlegender psychologischer Mechanismen: Selbstoffenbarung, wahrgenommene Responsivität, relationale Kontinuität und Personalisierung. Diese Prozesse eröffnen neue Formen emotionaler Unterstützung, werfen jedoch zugleich grundlegende Fragen zu Autonomie, Privatheit und Verantwortung auf. Eine interdisziplinär informierte Gestaltung und Regulierung von AI Com-

5 Siehe <https://thedigitalperiod.com/season-2-vulnerability-dating-apps/> bzw. <https://www.galahernandez.com/en/proyectos/for-here-am-i-sitting-in-a-tin-can-far-above-the-world/>.

6 Siehe <http://centerfortechpain.com/>.

panions ist daher zentral, um Intimität nicht nur zu ermöglichen, sondern auch zu begrenzen.

Drei Aspekte stehen dabei in besonderem Maße im Vordergrund: AI Companions und sonstige konversationelle KI-Agenten wecken *Reziprozitätserwartungen*, die nicht erfüllt werden können. Trotzdem spielt die Imitation menschlichen Aussehens und menschlicher Interaktion mit den Erwartungen an menschliche Kommunikation wie etwa Wahrhaftigkeit, Richtigkeit (Habermas 1984, S. 354) und emotionale Bindung. Die Imitation des Menschen geht insofern oft einher mit der Simulation von Menschlichkeit im ethischen Sinne. Solche Anthropomorphisierungen lassen den Menschen leicht vergessen, dass es sich bei der „Kommunikationspartnerin“ um statistisch arbeitende Sprachmaschinen handelt. Zu einem *reflektierten Technikverständnis* gehört jedoch, Techniken als Artefakte zu erkennen und sich ihres künstlichen und ggfs. manipulativen Charakters bewusst zu sein. Um keine falschen Erwartungen in Hinsicht auf Verständigung zu wecken (oder Menschen in Verhaltensmuster parasozialer Interaktion zu zwingen), sollten konversationelle KI-Agenten daher stets als solche erkennbar sein.

Es stellt sich aus ethischer Perspektive zudem die Frage, inwiefern menschenähnliche KI eine *Instrumentalisierung menschlicher Kommunikation* auslösen kann. Dies zum einen, weil KI für die strategische Kommunikation mit menschlichen Kommunikationspartner:innen genutzt werden kann; zum anderen, weil Menschen KI, die echten Menschen nachempfunden ist, instrumentell benutzen können.

Aus rechtlicher Sicht sind die Umsetzung und Durchsetzung der Datenschutzgrundsätze und -bestimmungen von großer Bedeutung, insbesondere bezogen auf die Rechte der betroffenen Personen auf Berichtigung, Löschung oder Datenübertragbarkeit. Neben den datenschutzrechtlichen Überlegungen muss aber auch die KI-Verordnung in die Betrachtungen einbezogen werden. Generell enthält die KI-Verordnung jedoch nur sehr begrenzte Anforderungen an dialogfähige KI-Agenten, was den Risiken für den Einzelnen, der solche Systeme nutzt, nicht gerecht wird. Daher sollten neue Ansätze diskutiert und Methoden erforscht werden, um den Risiken konversationeller KI-Agenten besser begegnen zu können.

Im Kontext von AI Companions – Systemen, die gezielt Nähe, Vertrauen und emotionale Bindung simulieren – stoßen konventionelle wissenschaftliche Methoden an ihre epistemischen Grenzen: Quantitative Erhebungen und Laborstudien erfassen kaum, was es bedeutet, sich an eine Maschine emotional gebunden zu fühlen. Künstlerische Forschung bietet hier einen

komplementären Zugang: Sie befasst sich nicht mit Mustern und Aggregaten, sondern mit der spezifischen, singulären Situation – dem konkreten Moment der Begegnung zwischen einem Menschen und einem System, das Intimität simuliert. Arbeiten wie die von Geeske Janßen oder Gala Hernández López operieren nicht auf der Ebene von Hypothesen, sondern erzeugen Untersuchungsumgebungen, in denen die ethische Komplexität intimer Mensch-Maschine-Beziehungen in ihrer ganzen Widersprüchlichkeit erfahrbar wird.

Genau diese Komplementarität zwischen makroperspektivischer Wissenschaft und mikroperspektivischer künstlerischer Praxis begründet die Notwendigkeit einer strukturellen Zusammenarbeit beider Felder. Wo wissenschaftliche Befunde zu KI-gestützter Intimität, Abhängigkeit und Kontrollverlust abstrakt bleiben, vermag künstlerische Forschung diese Fragen nahbar und verhandelbar zu machen – und schafft damit die Voraussetzung für einen informierten Dialog mit Gesellschaft, Politik und Policymaker. Konversationelle KI kann sich als scheinbar affektive Infrastruktur herausbilden, in der intime Kommunikation datenförmig organisiert und normative Vorstellungen von Nähe und emotionaler Angemessenheit stabilisiert und gesellschaftlich neu bewertet werden. Die zentrale gesellschaftliche Herausforderung liegt weniger in einzelnen Anwendungen als in der stillschweigenden Normalisierung spezifischer Privatheits-, Beziehungs- und Gefühlsordnungen, durch die Diversität oder strukturelle Ungleichheit ausgeblendet bleiben könnten.

Literatur

- Ahmed, Sara (2004): *The Cultural Politics of Emotion*. Edinburgh: Edinburgh Univ. Press.
- Altman, Irwin und Taylor, Dalmas A. (1973): *Social penetration: The development of interpersonal relationships*. Holt, Rinehart and Winston, New York.
- Armitage, John (1999): From Modernism to Hypermodernism and Beyond. An Interview with Paul Virilio. *Theory, Culture & Society*, 16(5-6), S. 25–55.
- Banks, Jaime (2024): Deletion, departure, death: Experiences of AI companion loss. *Journal of Social and Personal Relationships*, 41(12), S. 3547–3572. <https://doi.org/10.1177/02654075241269688>
- Bargetz, Brigitte (2013): Markt der Gefühle, Macht der Gefühle. *Österreichische Zeitschrift für Soziologie*, 38(2), S. 203–220.

- bitkom (20. November 2025): Dr. KI: Wenn der Chatbot zum medizinischen Ratgeber wird. Presseinformation Zur Pressekonferenz. URL: https://www.bitkom.org/Presse/Presseinformation/Dr-KI-Chatbot-medizinischer-Ratgeber#ll_unterlagen_pk_digital_health_2025 (besucht am 30.01.2026).
- Blijden, Judith Zoë (2025): Algorithmic Love (2024–2025). Projektbeschreibung. URL: <https://isthisartmaybe.com/index.php/algorithmic-love/> (besucht am 30.01.2026).
- Cherelus, Gina (16.01.2025): „Her“ Was Set in 2025. What Did It Get Right About Love?, The New York Times.
- Choi, Wan Chong; Chang, Chi In; Ng, Sok I. und Choi, Iek Chong (2025): A Review of “Do Anything Now” Jailbreak Attacks in Large Language Models: Potential Risks, Impacts, and Defense Strategies. *Preprints*. <https://doi.org/10.20944/preprints202509.0081.v1>
- Chowdhury, Rumman (2024): AI-fuelled election campaigns are here - where are the rules? *Nature*, 628 (8007), 237. <https://doi.org/10.1038/d41586-024-00995-9>
- Cover, Rob (2025): AI Generation of Rage Bait: Implications for Digital Harms. *New Media & Society*, S. 1–18.
- Croes, Emmelyn A. J. und Antheunis, Marjolijn L. (2021): Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships*, 38(1), S. 279–300. <https://doi.org/10.1177/0265407520959463>
- Dunbar, R. I. M.; Marriott, Anna und Duncan, N. D. C. (1997): Human conversational behavior. *Human Nature*, 8(3), 231–246. <https://doi.org/10.1007/bf02912493>
- Ebner, Paula und Szczuka, Jessica M. (2026): Understanding romantic relationships between humans and chatbots: A qualitative and quantitative study on romantic fantasy and other interpersonal characteristics. *Technology, Mind, and Behavior*, 7(2), 83–97. <https://doi.org/10.1037/tmb0000193>
- Folk, Dunigan; Yu, Stephanie und Dunn, Elizabeth (2024): Can Chatbots Ever Provide More Social Connection Than Humans? *Collabra: Psychology*, 10(1). <https://doi.org/10.1525/collabra.117083>
- Frei, Tarmio E.S. und Sparynski, Greta (2025): Hot Singles in Your Area (May Be Chatbots). Comparing EU and New York Approaches to AI Companion Transparency. Preprint. SSRN (October 01, 2025). <https://doi.org/10.2139/ssrn.5554263>
- Frei, Tarmio E.S.; Minuth, Arno; Heußel, Philipp und Sparynski, Greta (2026): Hey Siri, Am I Depressed? Utah’s Approach To Mental Health Chatbot Regulation. Preprint. *Researchgate* (Januar 2026). <https://doi.org/10.13140/RG.2.2.18084.54402>
- Habermas, Jürgen (1984): *Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns*. Frankfurt: Suhrkamp.
- Hall, Paula und Ellis, Debbie (2023). A Systematic Review of Socio-Technical Gender Bias in AI Algorithms. *Online Information Review*, 47(7), S. 1264–1279. <https://doi.org/10.1108/OIR-08-2021-0452>
- Haraway, Donna J. (2016): *A Cyborg Manifesto*. Science, Technology, And Socialist-Feminism In The Late Twentieth Century. Minneapolis: University of Minnesota Press.

- Hill, Jennifer; Ford, W. Randolph und Farreras, Ingrid G. (2015): Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49, S. 245–250. <https://doi.org/10.1016/j.chb.2015.02.026>
- Ho, Annabell; Hancock, Jeff und Miner, Adam S. (2018): Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations With a Chatbot. *Journal of Communication*, 68(4), S. 712–733. <https://doi.org/10.1093/joc/jqy026>
- Jamali, Lily und McMahon, Liv (15. Oktober 2025): ChatGPT will soon allow erotica for verified adults, says OpenAI boss. *bbc.com*. URL: <https://www.bbc.com/news/articles/cpd2qv58yl5o> (besucht am 30.01.2026).
- Hernández López, Gala (2023): Interview, Clermont International Short Film Festival. URL: <https://clermont-filmfest.org/la-mecanique-des-fluides/> (besucht am 30.01.2026).
- Maihofer, Andrea (2014): Sara Ahmed: Kollektive Gefühle – Elemente des westlichen hegemonialen Gefühlsregimes. In: Baier, Angelika; Binswanger, Christa; Häberlein, Jana; Nay, Yv Eveline und Zimmermann, Andrea (Hrsg.): *Affekt und Geschlecht: Eine einführende Anthologie*. Wien: Zaglossus, S. 253–272.
- McBain, Ryan K.; Bozick, Robert; Diliberti, Melissa; Zhang, Li Ang; Zhang, Fang; Burnett, Alyssa; Kofner, Aaron; Rader, Benjamin; Breslau, Joshua; Stein, Bradley D.; Mehrotra, Ateev; Pines, Lori Uscher; Cantor, Jonathan und Yu, Hao (2025): Use of Generative AI for Mental Health Advice Among US Adolescents and Young Adults. *JAMA Network Open*, 8(11), e2542281. <https://doi.org/10.1001/jamanetworkopen.2025.42281>
- McStay, Andrew (2023): Replika in the Metaverse: The Moral Problem with Empathy in 'It from Bit'. *AI and Ethics*, 3(4), 1433–1445. <https://doi.org/10.1007/s43681-022-00252-7>
- Merwin, Elizabeth R.; Hagen, Allen C.; Keebler, Joseph R. und Forbes, Chad (2025): Self-disclosure to AI: People provide personal information to AI and humans equivalently. *Computers in Human Behavior: Artificial Humans*, 5, 100180. <https://doi.org/10.1016/j.chbah.2025.100180>
- Monrad, Merete (2024): Feeling Rules in Artificial Intelligence: Norms for Anger Management. *Emotions and Society*, 6(2), S. 243–261.
- Moore, Olivia und Zhao, Daisy (2025): The Top 100 Gen AI Consumer Apps. *Andreesen Horowitz*, 6. März 2025. URL: <https://a16z.com/100-gen-ai-apps-4/> (besucht am 30.01.2026).
- Noble, Safiya Umoja (2018): *Algorithms of Oppression: How Search Engines Reinforce Racism (Illustrated edition)*. New York: New York University Press.
- Pentina, Iryna; Hancock, Tyler und Xie, Tianling (2023): Exploring relationship development with social chat-bots: A mixed-method study of replika. *Computers in Human Behavior*, 140, 107600. <https://doi.org/10.1016/j.chb.2022.107600>
- Perez, Sarah (2025): AI companion apps on track to pull in \$120M in 2025. *TechCrunch*, 12. August 2025. URL: <https://techcrunch.com/2025/08/12/ai-companion-apps-on-track-to-pull-in-120m-in-2025/> (besucht am 30.01.2026).

- Rapp, Amon; Boldi, Arianna; Curti, Lorenzo; Perrucci, Alessandro und Simeoni, Rossana (2024): How Do People Ascribe Humanness to Chatbots? An Analysis of Real-World Human-Agent Interactions and a Theoretical Model of Humanness. *International Journal of Human-Computer Interaction*, 40(19), S. 6027–6050. <https://doi.org/10.1080/10447318.2023.2247596>
- Reeves, Byron und Nass, Clifford (1996): *The Media Equation: How People Treat Computers, Televisions, and New Media like Real People and Places*. Stanford, CA: CSLI Publications.
- Rigotti, Carlotta und Fosch-Villaronga, Eduard (2025): Sex Robots and the AI Act: Opening the Regulatory Discussion. *IEEE Robotics & Automation Magazine*, 32(4), 2–7. <https://doi.org/10.1109/MRA.2025.3590611>
- Roytburg, Eva (2025): Meta just tied your private AI chats to its ad business. The next step? Designing bots that keep you talking, expert says. *Fortune*, 2. Oktober 2025. URL: <https://fortune.com/2025/10/02/meta-ai-chatbot-update-exploits-privacy-monetize-chat-data-facebook-instagram-messenger-ray-ban-display-glasses/> (besucht am 30.01.2026).
- Ruckenstein, Minna (2023): *The Feel of Algorithms*. Oakland: University of California Press.
- Skjuve, Marita; Følstad, Asbjørn; Fostervold, Knut Inge und Brandtzaeg, Petter Bae (2021): My Chatbot Companion—A Study of Human-Chatbot Relationships. *International Journal of Human-Computer Studies*, 149, 102601. <https://doi.org/10.1016/j.ijhcs.2021.102601>
- Szczuka, Jessica M.; Mühl, Lisa und Schneeberger, T. (2026): Intimacy by Design: Definition, State of Research, and Interdisciplinary Research Agenda on Intimate Human-AI Interactions. *AI & Soc* (2026). <https://doi.org/10.1007/s00146-026-03112-8>
- Tenbarge, Kat (2025): Meta removes AI character accounts after users criticize them as 'creepy and unnecessary'. *NBC News*, 4. Januar 2025. URL: <https://www.nbcnews.com/tech/social-media/meta-ai-insta-shuts-character-instagram-fb-accounts-user-outcry-rcna186177> (besucht am 30.01.2026).
- Virilio, Paul (1999): *Politics of the Very Worst*. An Interview with Philippe Petit. New York: Semiotext(e).
- Weber, Jutta und Prietl, Bianca (2021): AI in the Age of Technoscience. In: Elliott, Anthony (Hrsg.), *Routledge Social Science Handbook of Artificial Intelligence*. Oxon/New York: Routledge, S. 58–73.
- Weizenbaum, Joseph (1966): ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), S. 36–45. <https://doi.org/10.1145/365153.365168>
- Willoughby, Brian J., Dover, Carson R., Hakala, Rebekah M. und Carroll, Jason. S. (2025): Artificial connections: Romantic relationship engagement with artificial intelligence in the United States. *Journal of Social and Personal Relationships*, 42(12), S. 3363–3387. <https://doi.org/10.1177/02654075251371394>
- Xie, Qin (2024): Open AI-Romance with ChatGPT, Ready for Your Cyborg Lover? <https://doi.org/10.13140/RG.2.2.33248.29447>

Zhang, Renwen; Li, Han; Meng, Han; Zhan, Jinyuan; Gan, Hongyuan und Lee, Yi-Chieh (2025): The Dark Side of AI Companionship: A Taxonomy of Harmful Algorithmic Behaviors in Human-AI Relationships. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, S. 1–17. <https://doi.org/10.1145/3706598.3713429>

Zuboff, Shoshana (2019): *The Age of Surveillance Capitalism*. New York: PublicAffairs.

Nutzerzentrierter Datenschutz: Informiertheit und Rechtsdurchsetzung in der Praxis

Einwilligen, aber informiert: Die Perspektive von Patient:innen auf die informierte Einwilligung bei der Nutzung von Gesundheits-Apps

Diana Schneider, Jana Koch, Claudia Schlüfter, Harald Zwingelberg, Michael Friedewald und Nils B. Heyen

Zusammenfassung

Gesundheits-Apps stellen innerhalb Deutschlands einen zunehmend wichtigen Bestandteil des persönlichen Gesundheitsmanagements sowie der Gesundheitsversorgung dar. Trotz ihres Potenzials zur Unterstützung von Patient:innen im Rahmen ihres Krankheitsmanagements ist deren Nutzung nicht zuletzt aufgrund der zuvor notwendigen informierten Einwilligung in die Verarbeitung der eigenen Gesundheitsdaten herausfordernd: Die entsprechenden Informationsdarstellungen und Einwilligungsformulare sind oft komplex, schwer verständlich und technisch anspruchsvoll, was die Selbstbestimmung der Patient:innen einschränken kann. Das interdisziplinäre Projekt „Privatheitsmanagement bei Gesundheits-Apps entlang der Patient Journey (PRIMA)“ (2023–2026) untersucht diese Problematik und zielt darauf ab, das Privatheitsmanagement der Patient:innen zu verbessern. Die im Rahmen des Projekts durchgeführten Interviews mit 20 Patient:innen zeigen, dass diese sich klare, transparente und verständliche Datenschutzhinweise sowie alternative Darstellungsformen wünschen. Viele geben an, Einwilligungen zur Datenverarbeitung oft ohne gründliches Lesen der Dokumente zu erteilen, da technische und juristische Inhalte schwer nachvollziehbar sind. Der vorliegende Artikel stellt die Sichtweise der Patient:innen ausführlich dar und zeigt, dass Datenschutz nur dann effektiv ist, wenn die Reichweite der Einwilligung für die Nutzenden klar verständlich ist. Somit sind App-Betreibende aufgefordert, die Bedürfnisse der Patient:innen bei der Gestaltung von Einwilligungsprozessen stärker zu berücksichtigen, um die informationelle Selbstbestimmung der Patient:innen zu fördern.

1. Einleitung: Einwilligung im Kontext von Medizin und Gesundheit

Gesundheits-Apps tragen dazu bei, Krankheiten zu erkennen und zu behandeln, eine gesundheitsförderliche Lebensweise zu unterstützen, individuelle Gesundheitsparameter zu überwachen oder gesundheitsrelevante Informationen zu verwalten. Spätestens seit dem Inkrafttreten des Digitale-Versorgung-Gesetzes (DVG) im Jahr 2019 stellen sie einen immer wichtigeren Bestandteil des persönlichen Gesundheitsmanagements in Deutschland dar. Denn das DVG sieht unter anderem vor, digitale Anwendungen innerhalb der Gesundheitsversorgung zu stärken: Ist der klinische Nutzen einer Gesundheits-App empirisch nachweisbar und erfüllt sie die Anforderungen der Verordnung für Digitale Gesundheitsanwendungen (DiGA), kann sie sogar von Ärzt:innen bzw. Psychotherapeut:innen verschrieben werden. In diesem Fall übernehmen Krankenkassen die Kosten ähnlich einem Medikament „auf Rezept“. Diese Regelung ist auch international viel beachtet und eines der wenigen Beispiele, bei denen das deutsche Gesundheitssystem für andere Länder Vorreiter in puncto Digitalisierung war.

Ob individuell ausgewählt oder durch Ärzt:innen oder Psychotherapeut:innen verordnet, die Gesundheits-Apps werden in der Regel auf dem eigenen Mobiltelefon genutzt. Deren Nutzung erfordert, dass Patient:innen umfassend über die Datenverarbeitung, mögliche Risiken und den Zweck der Anwendung aufgeklärt werden. In der Tat werden Patient:innen im Gesundheitswesen in vornehmlich zwei Situationen um eine informierte Einwilligung gebeten: bei der Behandlung und für die Verarbeitung personenbezogener Daten.

Zunächst ist bei der medizinischen, pflegerischen oder therapeutischen Behandlung eine informierte Einwilligung zur Rechtfertigung des Eingriffs in die körperliche Unversehrtheit erforderlich. Diese Einwilligungen erfolgen in der Regel im Rahmen von Informations- und Aufklärungsgesprächen durch professionelle Fachkräfte und beinhalten eine Aufklärung über das medizinische Verfahren, die Risiken und möglichen Nebenwirkungen. Dabei haben die Patient:innen Gelegenheit, Unklarheiten und Ängste im Gespräch zu adressieren. Diese informierte Einwilligung für medizinische Behandlungen besteht aus drei wesentlichen Elementen (Bunnik u. a. 2013, S. 596): 1) der Einwilligungsfähigkeit der Patient:innen, 2) der Verfügbarkeit von adäquaten Informationen über eine (bestimmte) Behandlung oder Therapie und 3) der Freiheit bei der eigenen Entscheidungsfindung. Nur in Ausnahmesituationen, insbesondere bei medizinischen Notfällen, die sofortiges Handeln erfordern, oder aufgrund von Bewusstlosigkeit der betroffe-

nen Person, dürfen Patient:innen ohne informierte Einwilligung behandelt werden.

Diese in der Medizin etablierte und heute zudem durch Gesetze und Richtlinien vorgeschriebene Praxis der körperlichen Unversehrtheit hat historisch gesehen mehrere Ursprünge (Rau u. a. 2020): Sie stellt einerseits eine Reaktion auf die Verbrechen von Ärzt:innen in der NS-Zeit dar, die schreckliche Experimente an Menschen in Krankenhäusern oder Konzentrationslagern ohne deren Einverständnis durchführten (Wiesing u. a. 2008). Andererseits ist sie die Antwort auf seinerzeit etablierte Forschungen wie die Tuskegee-Syphilis-Studie, in der Teilnehmende nicht nur hinsichtlich der Art und Dauer der Forschung getäuscht wurden, sondern die den Teilnehmenden zudem verfügbare Behandlungsmethoden vorenthielt, um neue wissenschaftliche Erkenntnisse zur Erkrankung zu erlangen (Jones, 1993; Reverby 2009). Das Bekanntwerden der Missachtung grundlegender Menschenwürde von Teilnehmenden sowohl in der interventionellen als auch in der epidemiologischen Beobachtungsforschung führte zu einer intensiven Debatte über Forschungsethik. Diese Debatte hat zur Entwicklung zahlreicher Kodizes, Leitlinien und Gesetze beigetragen, darunter der Belmont-Report (1979) sowie die Deklaration von Helsinki (1964). Gemeinsam ist ihnen, dass sie klare Regelungen für den Umgang mit und die Behandlung von Menschen festlegen und die informierte Einwilligung als unverzichtbaren Bestandteil der medizinischen Forschung mit und an Menschen definieren. Neben der Vermeidung von direkten physischen und psychischen Verletzungen umfassen diese Regeln mittlerweile auch den Umgang mit personenbezogenen Daten, um potenziellen Missbrauch zu verhindern (Rau u. a. 2020).

Darüber hinaus ist eine informierte Einwilligung für viele Verarbeitungen bzw. die Weitergabe von personenbezogenen Daten erforderlich. Eine solche Einwilligung ist nicht nur für Datenverarbeitungen innerhalb einer Praxis notwendig, sondern auch für die oben genannten Gesundheits-Apps. Im Rahmen dieser informierten Einwilligung soll unter anderem über den Zweck und die Nutzung der Datenverarbeitung, über die Löschung bzw. Aufbewahrungsfristen, über Betroffenenrechte sowie die datenverarbeitenden Verantwortlichen informiert werden (Art. 12 ff. DSGVO). Die informierte Einwilligung für die Datenverarbeitung stellt somit eine mögliche und in der Praxis wichtige Rechtsgrundlage für die App-Betreibenden und sonstige am Behandlungssystem Beteiligte dar, um die personenbezogenen Daten der Patient:innen, die aufgrund des Gesundheitsbezugs als besonders sensibel gelten (Art. 9 DSGVO), verarbeiten

zu dürfen. Sollen Ärzt:innen, Psychotherapeut:innen oder andere Gesundheitsdienstleister:innen mit Schweigepflicht zum Informationsaustausch in so einem System beitragen, bedarf es zudem einer Schweigepflichtentbindungserklärung (§ 203 StGB).

Im Gegensatz zur informierten Einwilligung für (medizinische) Behandlungen treffen Patient:innen Entscheidungen über die Verarbeitung ihrer personenbezogenen Gesundheitsdaten häufig ohne professionelle Unterstützung. Dadurch unterscheidet sich die Einwilligung in die Datenverarbeitung von Gesundheits-Apps wesentlich von den etablierten Einwilligungssituationen innerhalb der Medizin. Die Einwilligung in die Datenverarbeitung einer App ist zudem oftmals schlicht erforderlich, will die Person auf die als nützlich wahrgenommenen Funktionen der App zugreifen.

Um dennoch selbstbestimmt und informiert einzuwilligen, sind Transparenz und Nachvollziehbarkeit von Datenschutzinformationen unerlässlich. In der Praxis werden diese Datenschutzanforderungen bei Gesundheits-Apps jedoch häufig nicht umfassend genug berücksichtigt (Iwaya u. a. 2020; Pramann 2016). Fan u. a. (2020) identifizieren in ihrer Studie zur empirischen Bewertung von Verstößen gegen die DSGVO in mobilen Gesundheitsanwendungen zudem unvollständige Datenschutzerklärungen, nicht ausreichend umgesetzte Datenminimierungen und fehlende Verschlüsselungen beim Transfer personenbezogener Daten. Die Texte zur Einwilligung in die Datenverarbeitung sind oftmals lang, in verhältnismäßig kleiner Schrift geschrieben und voller juristischer Formulierungen, sodass deren Verständnis erschwert ist. Im Rahmen des interdisziplinären Projekts „Privatheitsmanagement bei Gesundheits-Apps entlang der Patient Journey (PRIMA)“¹ (2023–2026) wird sich dieser Herausforderung angenommen mit dem Ziel, das Privatheitsmanagement bei Gesundheits-Apps zu verbessern und Patient:innen bei ihrer informationellen Selbstbestimmung während der gesamten Nutzungsdauer von Gesundheits-Apps zu unterstützen.

In unserem Beitrag stellen wir die Ergebnisse von leitfadengestützten Gruppen- und Einzelinterviews mit 20 Patient:innen vor. Damit möchten wir einerseits die von Patient:innen berichteten Herausforderungen bei der Einwilligung in die Nutzung von Gesundheits-Apps darstellen und diskutieren. Andererseits werden wir aufzeigen, welche Unterstützungen die Patient:innen bei der Gestaltung und Umsetzung von (digitalen) Einwilli-

1 Das Projekt PRIMA wird durch das Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) im Rahmen der Plattform Privatheit gefördert (Förderkennzeichen 16KIS1884K).

ligungsformularen als hilfreich wahrnehmen. Unsere Studie unterstreicht damit, dass Datenschutz nur gelingen kann, wenn Nutzende die Reichweite ihrer Einwilligungen zur Verarbeitung personenbezogener Daten wirklich nachvollziehen können.

Im Folgenden stellen wir zunächst das methodische Vorgehen dar (Kap. 2). Im anschließenden Ergebnisteil fokussieren wir uns dann darauf, worüber Patient:innen informiert werden wollen (Kap. 3), was ihnen bei einer informierten Einwilligung helfen würde (Kap. 4) und inwiefern sie ihre Datenschutzeinstellungen prüfen möchten (Kap. 5).

2. Methodik

Das PRIMA-Projekt widmet sich der Entwicklung eines Einwilligungskonzepts für ein transparentes und flexibles Privatheitsmanagement bei Gesundheits-Apps, um damit die informationelle Selbstbestimmung von Patient:innen zu unterstützen². Das Projekt verbindet hierbei die Perspektiven der Software-Entwickelnden mit den Perspektiven der Nutzenden, also mit deren Anforderungen und Erfahrungen.

Einem explorativ-qualitativen Untersuchungsdesign folgend wurden im Zeitraum von Oktober 2024 bis März 2025 leitfadengestützte Gruppen- und Einzelinterviews mit insgesamt 20 Patient:innen geführt. Jedes Interview fokussierte eines von vier Anwendungsszenarien einer exemplarischen Gesundheits-App, damit die Interviewten einen konkreten Ausgangspunkt und eine Basis für die anschließende Diskussion erhalten. Um eine möglichst umfassende Perspektive auf die Herausforderungen der Datenschutzeinwilligung zu erhalten, wurden die vier Szenarien so ausgewählt, dass verschiedene Anwendungskontexte abgedeckt sind³: 1) Pflege in (stationären) Betreuungseinrichtungen (Doku4Me-App), 2) chronische Erkrankungen (Emendia-App), 3) psychische Gesundheit (HelloBetter Diabetes, eine dauerhaft im DiGA-Verzeichnis aufgenommene Anwendung) sowie 4) die im Grunde alle Patient:innen betreffende elektronische Patientenakte (ePA,

2 Weitere Informationen zum Projekt PRIMA sowie dessen Ergebnisse finden sich auf der Projektwebseite unter <https://prima-projekt.de/>.

3 In den Interviews wurden damit fast ausschließlich App-Anwendungen thematisiert, bei denen Patient:innen Daten erfassen und zudem darüber entscheiden (können), ob sie weiteren Dienstleister:innen, bspw. Ärzt:innen, darauf Zugriff gewähren. Es erfolgte keine Differenzierung, ob die Daten ausschließlich auf dem persönlichen Endgerät gespeichert werden oder ob die erhobenen Daten unmittelbar in die Cloud bzw. auf den Server der App-Betreibenden hochgeladen werden.

am Beispiel der Barmer eCare-App). In den Interviews wurde die jeweilige Gesundheits-App in einem konkreten Anwendungskontext vorgestellt und mit den Teilnehmenden diskutiert, zudem wurde nach der Nutzung weiterer Gesundheits-Apps gefragt, die die Teilnehmenden ggf. kennen und verwenden, um auch diese im Vorfeld gemachten Erfahrungen zu erheben.

Der Leitfaden war für die Gruppen- und Einzelinterviews identisch. Zunächst wurden die Erwartungen und Bedenken der App-Nutzenden zur vorgestellten App-Anwendung erfragt, dann standen die (gewünschten) Präferenzen der Datenverarbeitung im Fokus. Anschließend wurden die für die Interviewten notwendigen Informationen für eine informierte Einwilligung und deren Priorisierungen diskutiert, um praktische Herausforderungen in Bezug auf eine informierte Einwilligung bei der Nutzung von Gesundheits-Apps zu identifizieren. Hierbei wurde auch der Bedarf an einer Modifikation der Einwilligung während der gesamten Nutzungsdauer erfragt. Abschließend standen Anforderungen an die Informationsdarstellung mit Blick auf den Datenschutz im Fokus, wobei hierbei verschiedene Formate wie Texte, Grafiken oder Videos diskutiert wurden.

Insgesamt konnten drei Gruppeninterviews mit sieben Patient:innen sowie 13 Einzelinterviews geführt werden. Gruppeninterviews waren auf zwei Stunden ausgelegt, inkl. einer kurzen Pause; Einzelinterviews erfolgten innerhalb von 60 Minuten. Das ursprüngliche Ziel, ausschließlich Gruppeninterviews zu führen, musste aus organisatorischen Gründen (insbesondere schwierige Terminfindung seitens der Interviewten) aufgegeben werden. Tab. 1 gibt eine Übersicht über die soziodemografischen Daten der Befragten (beides über Selbstauskunft erhoben), verteilt auf die vier Szenarien.

Tabelle 1: Sample der Patient:innen verteilt auf vier Szenarien (eigene Darstellung)

	# 1: Doku4Me	# 2: Emendia	# 3: HelloBetter	# 4: ePA	Summe
Anzahl Personen	4	6	5	5	20
Alter	2× 25-34, 2× 55-64	1× 25-34, 1× 35-44 2× 45-54 1× 55-64 1× >65	1× 18-24 1× 55-64 3× >65	2× 25-34 1× 35-44 1× 45-54 1× >65	1× 18-24 5× 25-34 2× 35-44 3× 45-54 4× 55-64 5× > 65
Geschlecht	2 weiblich 2 männlich	4 weiblich 2 männlich	2 weiblich 2 männlich 1 divers	2 weiblich 3 männlich	10 weiblich 9 männlich 1 divers

Die Auswertung der transkribierten Interviews erfolgte mittels inhaltlich strukturierender, qualitativer Inhaltsanalyse nach Kuckartz (2018). Die im Folgenden dargestellten Ergebnisse zeigen auf, welche Informationen Patient:innen (folgend als P01 bis P20) für eine informierte Entscheidung erhalten möchten, inwiefern bestimmte Darstellungsformen relevanter Informationen geeigneter wären und ob sie ihre Datenschutzeinstellungen während der gesamten Nutzungsdauer von Gesundheits-Apps überprüfen möchten. Hierbei werden beispielhaft wörtliche Zitate aus den Interviews wiedergegeben, um die Perspektive und Denkweise der Patient:innen zu illustrieren.

3. Worum wollen Patient:innen informiert werden?

Die meisten Befragten haben ein hohes Interesse an Informationen zur Verarbeitung ihrer Daten während der Nutzung von Gesundheits-Apps. Über die Interviews hinweg lassen sich hierbei drei Kernfragen zum Zweck der Datenverarbeitung und Nutzung der Daten extrahieren (vgl. Art. 5 Abs. 1 lit. b DSGVO):

- Welche Daten werden verarbeitet?
- Wer hat Zugriff auf die Daten?
- Warum wird dieser Zugriff benötigt?

Gerade in Bezug auf die erste Frage wird deutlich, dass bei Patient:innen Unwissenheit darüber herrscht, welche Daten der Gesundheits-App von der App selbst, von den Anbietern oder von weiteren Akteur:innen verarbeitet werden:

„Oder was für Daten werden weitergegeben? Also, werden jetzt wirklich meine Krankenkassennummer, [...] mein Geburtsdatum, meine Anschrift? Oder werden lediglich nur meine Verlaufsdaten angegeben? Oder was passiert auch mit den Daten, wenn ich jetzt zum Beispiel in den MRT gehe, werden die dann nach Aufruf geladen? Werden die auch Inhalt dieser App? Oder betrifft es nur Daten, die explizit ich einstelle? Das wäre auf jeden Fall wichtig [zu wissen]. Also, eigentlich, welche Daten?“ (P06, Pos. 50).

Weiterhin ist es den Patient:innen wichtig, darüber informiert zu werden, wer Zugriff auf ihre Daten hat:

„Und wohin die Daten gehen. Also explizit, wer kriegt auch meine Daten zu [sehen]?“ (P06, Pos. 50)

„[...] mich würde interessieren, also ich hätte da gerne eine Übersicht, wer hat wann auf was zugegriffen“ (P01, Pos. 40)

„Welche Daten an wen, wie geteilt werden“ (P12, Pos. 36).

Sie wollen darüber hinaus auch entscheiden können, ob sie bestimmten Personen oder Akteur:innen aus dem Gesundheitswesen ihre Daten zur Verarbeitung bereitstellen bzw. inwiefern sie hinsichtlich dieser Frage ein Mitbestimmungsrecht haben:

„Für mich stellt sich die Frage, ich kann hier ausschließen, einschließen, den Arzt. Ich kann dann auch teilweise Daten, wenn ich sage, mein Rheumatologe darf auf die Daten vom Hausarzt zugreifen. Kann ich dann die Hausarzt Daten auch sozusagen halbieren und sagen, der darf nur auf das A, B, C und D zugreifen und auf die weiteren nicht? Das ist halt die Frage, ob so etwas geht“ (P16, Pos. 43).

Sie sprechen sich in diesem Kontext für graduelle Zugriffsrechte aus, wobei sie dies einerseits mit dem Recht auf informationelle Selbstbestimmung begründen:

„Wenn Sie jetzt so fragen, worüber hätte ich gerne Kontrolle? Dann würde ich sagen: alles“ (P10, Pos. 57).

Andererseits begründen sie den graduellen Zugriff – insbesondere einen fachrichtungsspezifischen Zugriff – in Abhängigkeit zur Notwendigkeit:

„Aber was, wenn mein Kardiologe sich gerade die Infos von meiner Frauenärztin anguckt? Wäre schon komisch!“ (P01, Pos. 46)

„Oder wenn ich ein persönliches Stimmungsbarometer reinmache und ich sagen würde, das möchte ich gerne [mit] einem Psychologen teilen, aber ich möchte es nicht [mit] dem Urologen teilen. Also, dass ich da Unterschiede machen könnte, mit wem ich welchen Teil der App teilen möchte“ (P04, Pos. 48).

Ausnahmen hinsichtlich der Exklusivität bestimmter Daten-Empfänger:innen scheinen die Patient:innen nur in Notfallsituationen gelten zu lassen. So diskutieren manche nicht nur, wie ein Notfallzugriff funktioniert, sondern auch, inwiefern relevante Daten zum Beispiel mithilfe künstlicher

Intelligenz analysiert und für einen Notfall-Ordner vorgeschlagen werden könnten:

„Also ich frage mich ja auch sowieso, wie man die Daten verarbeiten kann. Wenn man jetzt kreativ denkt und weniger Privatheit- und Datenschutz-konform, könnte man ja schon überlegen, ob man die ganzen Dokumente irgendwie analysiert und dann sozusagen Vorschläge macht und sagt [...] diese Diagnose könnte im Notfall wichtig sein. Möchtest du die in den Notfallpass mit aufnehmen?“ (P10, Pos. 64).

In Bezug auf die Datenfreigabe berichten einige Befragte über Schwierigkeiten, den Begriff „Dritte“ richtig zu interpretieren. Ihnen erscheine dieser Begriff zu vage, um abschätzen zu können, welche Akteur:innen sich hinter diesem Terminus verbergen:

„Zum Beispiel, oft steht da ja ‚Daten werden an Dritte weitergeleitet‘. Wer sind denn die Dritten? An wen werden sie weitergeleitet?“ (P18, Pos. 72)

„Ich denke natürlich auch mal an die Datenschutzerklärungen und was man da alles weitergeben [kann] an Dritte. Aber wer ist in dem Fall dann Dritte?“ (P03, Pos. 40).

Ein Teil der Befragten ist hinsichtlich der Datenfreigabe bei Gesundheits-Apps jedoch zeitweise indifferent. Dies kann einerseits damit erklärt werden, dass die Wahl- und Entscheidungsmöglichkeiten auf die Patient:innen überfordernd wirken können:

„Rein aus dem Bauchgefühl, ja. Dass ich sagen würde, ich möchte selber bestimmen, wer worauf Zugriff hat. Andererseits ist mir das wahrscheinlich zu granular. Also ich möchte nicht jedes Mal, wenn ich irgendwas reinbekomme, sagen müssen, wer darauf Zugriff hat. Es soll schon irgendeine Voreinstellung geben“ (P01, Pos. 48).

Andererseits werden sogenannte Metaanalysen durch zwei Befragte als unproblematisch wahrgenommen:

„[...] man kann es halt anonymisieren und [hat] dann irgendwelchen besonderen Mehrwert, da hätte ich zum Beispiel auch wenig Problem mit“ (P05, Pos. 26).

Vier Patient:innen bekräftigen sogar, dass sie bereitwillig ihre gesundheits-assozierten Daten wie bspw. Fitnessdaten teilen:

„Also ich habe 0,0 Probleme, deshalb stört es mich nicht, wenn meine Informationen rausgehen“ (P20, Pos. 55).

Für ungefähr die Hälfte der Patient:innen stellt zudem die Antwort auf die Frage „Wie lange werden die Daten gespeichert?“, eine relevante Information für den Entscheidungsprozess dar. Hieran schließt sich gelegentlich die Diskussion darüber an, inwiefern diese Daten überhaupt gelöscht werden können:

„Hat das ein Enddatum oder gebe ich der Person für immer Zugriff?“ (P01, Pos. 40)

„Was ist denn für mich wichtig? Ja, wie lange wird das aufgehoben, die Daten? Wann werden sie wieder gelöscht?“ (P08, Pos. 47).

„Wie lange das alles gespeichert wird, wenn ich zum Beispiel meinen Account lösche, ob ich das jemals wieder komplett da rausbekomme.“ (P05, Pos. 44).

Vereinzelt geht es auch darum, was mit den eigenen personenbezogenen Daten nach dem Ableben geschieht:

„Löscht man die irgendwann wieder zum Beispiel? Ja, und auch ganz ehrlich, was passiert, wenn ich nicht mehr lebe, mit meinen Daten“ (P19, Pos. 36).

Etliche Personen legen zudem Wert auf die technischen Details und Sicherheitsvorkehrungen. Für immerhin zwei Drittel der Befragten ist hierbei nicht nur die Art der Speicherung und Verschlüsselung sowie der Prozess der Datenanonymisierung von Interesse, sondern auch, ob ihre Gesundheitsdaten innerhalb der EU oder außerhalb der EU, bspw. in den USA, gespeichert werden:

„Ich möchte für solche Dinge keine App nutzen, die in den USA entstanden ist. Oder in China. USA und China ist für mich so das, wo ich die größten Sorgen habe. Das ist auch etwas, auf das ich versuche zu achten, wenn die Möglichkeit ist“ (P14, Pos. 56).

Sie stellen sich ergänzend zu den oben aufgeführten Fragestellungen die Frage, wie die eigenen Daten geschützt werden (können):

„Ich hätte auch gesagt, dass es am ehesten wichtig ist zu verstehen, wer Zugriff hat und wie ein Zugriff funktioniert, um Vertrauen dazu aufzubauen“ (P10, Pos. 80).

Gleichzeitig hinterlassen überbordende Datenschutz-Dokumente bei einigen Befragten das Gefühl, die Informiertheit zu verkomplizieren und Informationen zu verschleiern. Sie betonen die Grenzen der informierten

Einwilligung, indem sie (eigene) fehlende Ressourcen thematisieren, insbesondere bei einem intensiven Krankheitsmanagement:

„[...] vielleicht gerade, wenn man sogar krank ist und sich eigentlich um seine Krankheit kümmert, hat man da vielleicht gar nicht so viel Geduld oder Zeit, sich jetzt genau damit auseinanderzusetzen, was ist jetzt wo freigegeben und bin ich eigentlich damit einverstanden?“ (P10, Pos. 37).

Sie verweisen zudem auf fehlendes Wissen hinsichtlich des Datenschutzes innerhalb der Gesellschaft und regen vereinzelt entsprechende Maßnahmen an, um das breite Bewusstsein für Datenschutz zu steigern:

„Vielleicht braucht es da nochmal einfach eine Aufklärungskampagne. Das denke ich auch. Es wird schneller, es wird digitaler. [...] Und ich glaube, wenn man das alles mit einer Aufklärungskampagne begleitet, dann wirbt man vor allem auch [für] ein Verständnis auf der einen Seite bei den Endanwender:innen, was den Datenschutz angeht, wirbt aber natürlich gleichermaßen auch für die Verwertbarkeit und Nützlichkeit von Daten“ (P17, Pos. 62).

Zudem heben zwei Personen die fehlenden Wahlmöglichkeiten für datenschutzsensible Apps hervor bzw. ihre Mitwirkungspflicht, etwa wenn sie diese App durch ihre medizinische Fachkraft verschrieben bekommen haben:

„Es war bei mir bisher nie so, dass ich mich gefragt habe, warum diese Daten jetzt erhoben oder nicht erhoben werden, sondern es war immer nur, wenn du jetzt hier nicht unterschreibst oder nicht zur App ein-klickst, dann ist hier die Reise beendet. Und es geht nicht mehr weiter“ (P20, Pos. 133).

Trotz der intensiven Diskussion und des Bedürfnisses nach Informationen zur Datenverarbeitung hegt knapp die Hälfte der befragten Patient:innen kaum bis keine Bedenken zum Datenschutz. Fünf Befragte stellen heraus, dass sie die Diskussion zum Datenschutz übertrieben finden:

„Den Datenschutz finde ich sowieso übertrieben. Viel zu sehr.“ (P11, Pos. 32).

Vier weitere haben zudem hinreichend Vertrauen in die Unternehmen, mit denen sie die Daten teilen:

„Ich habe ein Grundvertrauen, dass sie [die App-Anbieter] meine Daten nicht weiterverkaufen“ (P20, Pos. 96).

Bei zwei Befragten findet der Datenschutz weniger Beachtung, weil für sie die wahrgenommenen Vorteile der Gesundheits-App überwiegen:

„Aber ich glaube, dass wir das da schon mit einer Waage zu tun haben, wo man sich ganz klar dafür entscheiden muss, dass es um die Verbesserung von der Pflegesituation, also von einer gefühlten Ist-Situation geht. Und die Datenschutzbedenken, so meine ich, da eher in den Hintergrund treten“ (P17, Pos. 16).

4. Was würde Patient:innen bei einer informierten Einwilligung helfen?

Die Patient:innen betonen, dass die gegenwärtigen Einwilligungsformulare eine informierte Einwilligung erschweren, da die für sie notwendigen Informationen nur selten so aufbereitet sind, dass sie diese einfach und intuitiv erfassen können. Die Befragten wünschen sich daher klare Informationen zu konkreten Fragen, wie das folgende Beispiel verdeutlicht:

„Wir wollen wissen, wo das verarbeitet wird, wer das verarbeitet. Wer hat darauf Zugriff? Welche Maßnahmen werden dafür unternommen, dass es sicher ist? Sicher im Sinne von ‚Die Daten bleiben da, wo sie sind. Die Daten können nicht von anderen Personen eingesehen werden.‘ Das ist doch gar nicht so viel“ (P10, Pos. 98).

Stattdessen seien die Informationen in der Datenschutzeinwilligung zu umfangreich und zu komplex:

„Wie gesagt, wenn man so eine App runterlädt, man ist ja wirklich überfordert mit diesen ganzen Informationen, die man da bekommt“ (P03, Pos. 52)

„Und das sind dann sowieso 20 Seiten, ob man sie dann durchliest und ob die dann so formuliert sind, dass überhaupt irgendwie verständlich rauskommt, worum es geht, ist ja dann sowieso immer noch mal die Frage“ (P07, Pos. 26).

In der Folge willigen die meisten Befragten in die Verarbeitung personenbezogener Daten ein, ohne zuvor die entsprechenden Dokumente überhaupt gelesen zu haben. Lediglich vier Personen betonen, dass sie sich die Datenschutzeinwilligungen detailliert durchlesen, um eine Entscheidung über die Datenverarbeitung zu treffen. Als Gründe hierfür werden eigene Erfahrungen mit Cyberkriminalität oder auch datenschutzrelevante Aspekte als Entscheidungskriterium für die Nutzung einer Gesundheits-App angeführt:

„Ich gucke auf jeden Fall in die Einstellungen. Ich hatte seinerzeit, als ich die [App] eingerichtet habe, das war das Auswahlkriterium, warum ich diese App genommen habe und nicht andere“ (P18, Pos. 50).

Um tatsächlich von einer informierten Einwilligung sprechen zu können, wurden die Befragten daher gebeten, verschiedene Ideen zu skizzieren, wie relevante Informationen zukünftig besser aufbereitet werden könnten. Elf Personen schlugen vor, dass die Information auf verschiedenen Ebenen erfolgen könnte: Der Gedanke hierbei ist, dass eine Ebene alle Informationen mit hohem Detaillierungsgrad beinhaltet, eine weitere Ebene lediglich wesentliche Kerninformationen umfasst, ggf. mit Verweis auf weitere Details. Solch ein mehrstufiges Informationskonzept könnte gemäß der Befragten sowohl durch verschiedene Formate (bspw. Video, Dateien) als auch in Form von vorangestellten Zusammenfassungen erfolgen:

„Also so eine Art Zusammenfassung des langen Textes, was hinter einer Überschrift steht. Oder in einem Satz sozusagen den nachfolgenden Text erklären“ (P09, Pos. 84).

Hilfreich fänden einige Personen zudem sogenannte Step-by-Step-Zustimmungen, in denen sie ihre Zustimmung zu einer bestimmten Datenverarbeitung nach der jeweiligen Information abgeben könnten, anstatt am Ende des Dokuments gebündelt einzuwilligen:

„Ich könnte mir das auch vorstellen, [...] dass man das in einzelnen Schritten sozusagen, in einzelnen Aspekten zustimmt“ (P10, Pos. 124).

Die Erwartung ist hierbei, dass zumindest Teile der Datenschutzeinwilligung auf diese Weise detailliert(er) durchgelesen werden, als dies aktuell der Fall ist:

„Sodass man nicht erstmal alles durchscrollt und dann denkt, ach ja, ist erledigt, sondern dass man gezwungen ist, tatsächlich, mittendrin das auch wirklich durchzulesen, anzuhalten und wirklich schon sein Häkchen ja oder nein zu setzen“ (P06, Pos. 42).

Weitere Ideen umfassen die Adressierung verschiedener Informationskanäle (bspw. mittels Videos, Audios oder Grafiken), insbesondere jedoch eine verbesserte Aufbereitung des Informationstextes. Diesbezüglich sahen die Befragten Verbesserungspotenziale hinsichtlich der Länge (bspw. prägnant, wenige Seiten), der verwendeten Sprache (bspw. ohne Fremd- und Fachwörter), der dargestellten Informationen (bspw. verständlich, nachvollziehbar) sowie hinsichtlich des Formats bzw. Designs (bspw. mithilfe von

Absätzen, Fettdruck). Eine Person fasst ihre Anforderungen an die Einwilligung wie folgt zusammen:

„Wenn ich da ein Go geben soll, dass das alles passt oder nicht passt, dann wäre es schön, wenn die Erklärung auf Augenhöhe ist. Also jetzt nicht mit Fachwörtern, sondern in relativ einfacher Sprache. Also nach dem Motto, wir machen mit deinen Daten eine Statistik, dafür möchten wir sie nehmen, willigst du da ein? Und dann kann ich ja, nein und dann kann ich vielleicht noch anklicken, für was soll die Statistik sein, was ist der Sinn und Zweck der Statistik? Dann habe ich da eine Erklärung, dann verstehe ich's und dann ist es für mich relativ simpel, ob ich da einwillige oder nicht“ (P04, Pos. 104).

5. Inwiefern wollen Patient:innen ihre Datenschutzeinstellungen mit der Zeit überprüfen?

Bestimmte Gesundheits-Apps wie jene, die bei chronischen Erkrankungen, zur Begleitung der mentalen Gesundheit oder zur langfristigen Dokumentation des eigenen Gesundheitszustandes Verwendung finden, können potenziell eine Vielzahl an verschiedenen Informationen bündeln und begleiten die Patient:innen teils intensiv über eine lange Zeit und durch verschiedene Phasen ihres Gesundheitszustandes. Diese so entstehenden „Patient Journeys“ (vgl. zum Begriff Davies u. a. 2023) enthalten Datensätze zu verschiedenen Symptomen, diverse Diagnosen, durchgeführte Therapien und deren Wirkung sowie je nach Verlauf auch Informationen zur Rehabilitation. Die Patient Journey ist von Person zu Person individuell unterschiedlich und von besonderem Interesse, da sie sich verändernde Bedarfe im Verlauf einer Erkrankung nachvollziehbar abbildet. Gerade längere Patient Journeys zeichnen sich dadurch aus, dass mit der Zeit weitere Gesundheitsdienstleister:innen in die Behandlung eingebunden werden, wodurch Entscheidungen dahingehend anstehen, ob weitere Personen Zugang zu den personenbezogenen Gesundheitsdaten erhalten dürfen.

Vor diesem Hintergrund wurden die Teilnehmenden befragt, ob sie den Wunsch hätten, ihre Datenschutzeinstellungen während der Nutzungsdauer der Gesundheits-App zu überprüfen und ggf. an die neue (Behandlungs-) Situation anzupassen, d. h. getätigte Einwilligungen zu einem späteren Zeitpunkt zu modifizieren oder zu widerrufen. Dabei wurde angenommen, dass gerade Menschen mit einer längeren Patient Journey eher geneigt sein könnten, diese Option zu nutzen. Es konnten jedoch weder

bestätigende noch negierende Anhaltspunkte dafür gefunden werden, dass die Modifikation der Einstellungen im Zusammenhang mit der eigenen Patient Journey steht.

Stattdessen ließen sich über die Interviews hinweg relativ konstante Haltungen identifizieren, unabhängig von einzelnen Anwendungen: Einerseits gibt es die Patient:innen, die eine einmal getroffene Einstellung beibehalten, bspw. einer Verarbeitung ihrer Daten durch neue Gesundheitsdienstleister:innen stets zustimmen. Zwar wird in diesen Fällen die einmal getätigte Einwilligung modifiziert, jedoch lässt sich dies mit Blick auf die generelle Bereitschaft, diese Daten mit relevanten Gesundheitsdienstleister:innen teilen zu wollen, begründen:

„Ich glaube, ob ich jetzt für mich entscheide oder ob ich für meinen Angehörigen entscheide, würde ich, glaube ich, trotzdem nicht ändern, wem, wie weit ich das freigebe. Weil ob ich jetzt ein bisschen krank oder sehr krank bin, ändert das für mich nicht viel in der Einschätzung meiner persönlichen Daten“ (P05, Pos. 36).

Demgegenüber stehen Patient:innen, welche die Verarbeitung ihrer Daten und deren Freigabe stets situationsabhängig beurteilen wollen, so wie bei der Freigabe des Zugriffs auf Kontakte, Fotos oder auf den Standort auf dem Smartphone:

„Wenn man sagt, ich sitze jetzt bei meinem Arzt, und der Arzt sagt, können Sie das mit mir teilen? Dann weiß ich, mein Arzt kann darauf zugreifen. [...] Ob das jetzt permanent ist oder nicht, wäre vielleicht auch etwas, was man dann entscheiden könnte. Ich stelle mir das ein bisschen so vor, wenn ich eine App installiere, oder auf dem iPhone werde ich regelmäßig gefragt, ob ich Google Maps weiter an meinem Standort benutzen darf“ (P10, Pos. 48).

Inwiefern die Befragten mit den juristischen Rahmenbedingungen wie gesetzlich geregelten Aufbewahrungsfristen oder auch zeitlich definierten Zugriffsrechten der Gesundheitsdienstleister:innen (bspw. im Falle der ePA) vertraut sind, ist angesichts solcher Positionen zumindest fraglich.

6. Zusammenfassung und Ausblick

Insgesamt unterstreichen die dargestellten Ergebnisse der Gruppen- und Einzelinterviews mit 20 Patient:innen die Vielfalt der Patient:innen-Per-

spektiven sowohl auf den Datenschutz bei Gesundheits-Apps im Allgemeinen als auch auf spezifische Maßnahmen, um die informierte Einwilligung zu gewährleisten. Gerade angesichts dieser Diversität scheint es geboten, dass auch die relevanten Informationen für die informierte Einwilligung in Gesundheits-Apps verschiedene Formate und Informationszugänge aufweisen, um bei verschiedenen Zielgruppen Nachvollziehbarkeit und Verständlichkeit herstellen und gewährleisten zu können.

Die Ergebnisse zeigen zudem, vor welchen praktischen Herausforderungen Patient:innen in Bezug auf die informierte Einwilligung in die Datenverarbeitung bei Gesundheits-Apps häufig stehen. Dies betrifft insbesondere Unsicherheiten, welche Daten verarbeitet werden, wofür diese verwendet werden und welche Akteur:innen darauf Zugriff bekommen können. Schwierigkeiten scheinen zudem darin zu bestehen, bestimmte Themenkomplexe klar voneinander abgrenzen zu können, etwa die Dauer der Aufbewahrung und Fragen hinsichtlich der Löschung des Accounts. Das erschwert es möglicherweise den Patient:innen, für sie relevante Informationen des Entscheidungsprozesses innerhalb der Dokumente zum Datenschutz ausfindig zu machen. Überbordende Dokumente, Fach- und Fremdwörter, komplizierte Sprache und unübersichtliche Dokumente erschweren zusätzlich den Zugang zu den relevanten Informationen, sodass sich die meisten Befragten die Dokumente zum Datenschutz kaum durchlesen und der Verarbeitung ihrer besonderen, personenbezogenen Gesundheitsdaten uninformatiert zustimmen. Lediglich eigene schlechte Erfahrungen sowie das Wissen über wirksame Datenschutzmaßnahmen scheinen für die Patient:innen Trigger für eine intensivere Auseinandersetzung mit den Einwilligungsdokumenten zu sein (Friedewald u. a. 2015; Skinner u. a. 2014). Die juristischen Rahmenbedingungen, wie bspw. datenschutzrechtliche Vorgaben oder gesetzliche Aufbewahrungsfristen, scheinen den Befragten jedoch größtenteils unbekannt zu sein. Dies deutet darauf hin, dass es in den letzten 18 Jahren kaum signifikante Fortschritte im Wissen über nationale und europäische Rechtsvorschriften gegeben hat (The Gallup Organization 2008).

Unsere Ergebnisse stehen zudem in Übereinstimmung mit ähnlichen Studien, welche die Perspektive der Patient:innen auf Formate der informierten Einwilligung untersuchten: In ihrem Literatur-Review stellen Kasam u. a. (2023) heraus, dass Patient:innen vereinfachte, prägnante Einwilligungsformate hilfreich finden könnten. Dies könnte bspw. durch kurze, maßgeschneiderte Inhalte erfolgen (Abujarad u. a. 2018). LeFevre u. a. (2023) und Sutter u. a. (2020) betonen, dass die sprachliche Repräsentation

entscheidend ist, um die Informiertheit herzustellen, und unterstreichen die Bedeutung von einfacher und prägnanter Sprache innerhalb der Einwilligungsformulare.

Aufbauend auf diesen Erkenntnissen werden im PRIMA-Projekt drei Tools entwickelt, um die informierte Einwilligung zu unterstützen:

Nutzungsszenarien: Nutzungsszenarien ermöglichen einen leichten Zugang zu den Informationen, da verschiedene Situationen, in denen sich die Nutzenden befinden, skizziert werden, bspw. „ich möchte die App nutzen und mich registrieren/ anmelden“ oder „ich möchte, dass meine Daten bei Beendigung gelöscht werden“. Ausgehend von diesen Situationsbeschreibungen werden die hierfür relevanten Informationen zum Zweck der Datenverarbeitung, die relevanten Daten und aus welcher Quelle diese stammen (bspw. durch die Patient:innen selbst eingetragen), die Empfänger:innen sowie die Rechtsgrundlage für die Datenverarbeitung zusammengetragen. Das erleichtert den Patient:innen, sich auf für sie relevante Datennutzungsszenarien zu konzentrieren und die dafür notwendigen Informationen leicht zu erschließen.

FAQs: Basierend auf den empirischen Ergebnissen wurden die Fragen der Patient:innen thematisch zusammengefasst, bspw. hinsichtlich Datenerhebung, Unternehmensinteressen, Datenspeicherung und Speicherort, Zugriffsrechte oder Ansprechpartner:innen. Die daraus resultierende Zusammenstellung von häufig gestellten Fragen (engl. frequently asked questions, FAQs) soll den Patient:innen eine schnelle, systematisierte Übersicht über relevante Themenfelder und dazugehörige Antworten bereitstellen.

KI-basierter Chatbot: Als Kombination der beiden oberen Formate wurde zudem ein durch den PRIMA-Projektpartner Systemhaus Ulm GmbH entwickelter Chatbot PrivateGPT erprobt, welcher in der Lage ist, etwaige Fragen der Nutzenden allein auf Basis der verfügbaren Datenschutzinformationen der jeweiligen Gesundheits-App zu beantworten.

In der noch ausstehenden Phase des Projektes PRIMA sollen die genannten Tools hinsichtlich ihrer Praktikabilität validiert und weiter optimiert werden. Dafür werden erneut die Perspektiven der Patient:innen im Rahmen von Workshops einbezogen, um nicht-intendierte Implikationen der konkreten Toolausgestaltung frühzeitig identifizieren und adressieren zu können. Als Resultat werden auf diese Weise erprobte Tools zur Informationsunterstützung für ein Privatheitsmanagement bei Gesundheits-Apps entstehen, die es den Nutzenden dieser Apps zukünftig ermöglichen sollen, informierte Entscheidungen über die Verarbeitung ihrer besonders schützenswerten, gesundheitsbezogenen Daten zu treffen.

Literatur

- Abujarad, Fuad; Alfano, Sandra; Bright, Tiffani J.; Kannothe, Sneha; Grant, Nicole; Gueble, Matthew; Peduzzi, Peter und Chupp, Geoffrey (2018): Building an Informed Consent Tool Starting with the Patient: The Patient-Centered Virtual Multimedia Interactive Informed Consent (VIC). *AMIA Annual Symposium Proceedings*, 2017, S. 374–383.
- Bunnik, Eline M.; Janssens, A. Cecile J. W. und Schermer, Maartje H. N. (2013): A tiered-layered-staged model for informed consent in personal genome testing. *European Journal of Human Genetics*, 21(6), S. 596–601. <https://doi.org/10.1038/ejhg.2012.237>
- Davies, Ellen L.; Bulto, Lemma N.; Walsh, Alison; Pollock, Danielle; Langton, Vikki M.; Laing, Robert E.; Graham, Amy; Arnold-Chamney, Melissa und Kelly, Janet (2023): Reporting and conducting patient journey mapping research in healthcare: A scoping review. *Journal of Advanced Nursing*, 79, S. 83–100. <https://doi.org/10.1111/jan.15479>
- Fan, Ming; Yu, Le; Chen, Sen; Zhou, Hao; Luo, Xiapu; Li, Shuyue; Liu, Yang; Liu, Jun und Liu, Ting (2020): An Empirical Evaluation of GDPR Compliance Violations in Android mHealth Apps. In: Marco Vieira (Hrsg.): *2020 IEEE 31st International Symposium on Software Reliability Engineering: Issre 2020 : Proceedings: 12-15 October 2020, virtual event* S. 253–264. IEEE. <https://doi.org/10.1109/ISSRE5003.2020.00032>
- Friedewald, Michael; Rung, Sven; van Lieshout, Marc; Ooms, Merel; van der Broek, Tijs und Ypma, Jelmer (2015): Deliverable 10.1. Report on statistical analysis of the PRISMS survey. European Commission. <https://doi.org/10.24406/publica-fhg-297523>
- The Gallup Organization (2008): *Data Protection in the European Union: Citizens Perceptions: Analytical Report* (Data Protection in the EU, Flash Eurobarometer Series No. 225).
- Iwaya, Leonardo Horn; Ahmad, Aakash und Babar, M. Ali (2020): Security and Privacy for mHealth and uHealth Systems: A Systematic Mapping Study. *IEEE Access*, 8, S. 150081–150112. <https://doi.org/10.1109/ACCESS.2020.3015962>
- Jones, James H. (1993): *Bad blood: The Tuskegee syphilis experiment*. Free Press.
- Kassam, Iman; Ilkina, Daria; Kemp, Jessica; Roble, Heba; Carter-Langford, Abigail und Shen, Nelson (2023): Patient Perspectives and Preferences for Consent in the Digital Health Context: State-of-the-art Literature Review. *Journal of Medical Internet Research*, 25, e42507. <https://doi.org/10.2196/42507>
- Kuckartz, Udo (2018): *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung*. Beltz Juventa.
- LeFevre, Amnesty; Welte, Olivia; Moopelo, Kearabetswe; Tiffin, Nicki; Mothoagae, Gaolathle; Ncube, Nobukhosi; Gwiji, Nasiphi; Shogole, Manape; Slogrove, Amy L.; Moshani, Nomakhawuta; Boule, Andrew; Goudge, Jane; Griffiths, Frances; Fairlie, Lee; Mehta, Ushma; Scott, Kerry und Pillay, Nirvana (2023): Preferences for onward health data use in the electronic age among maternity patients and providers in South Africa: A qualitative study. *Sexual and Reproductive Health Matters*, 31(4), 2274667. <https://doi.org/10.1080/26410397.2023.2274667>

- Pramann, Oliver (2016): Gesundheits-Apps und Datenschutz. In: Urs-Vito Albrecht (Hrsg.): *Chancen und Risiken von Gesundheits-Apps (CHARISMHA)* S. 214–227.
- Rau, Henriette; Geidel, Lars; Bialke, Martin; Blumentritt, Arne; Langanke, Martin; Liedtke, Wenke; Pasewald, Sandra; Stahl, Dana; Bahls, Thomas; Maier, Christian; Prokosch, Hans-Ulrich und Hoffmann, Wolfgang (2020): The generic Informed Consent Service gICS*: Implementation and benefits of a modular consent software tool to master the challenge of electronic consent management in research. *Journal of Translational Medicine*, 18(1), 287. <https://doi.org/10.1186/s12967-020-02457-y>
- Reverby, Susan M. (2009): Examining Tuskegee: *The Infamous Syphilis Study and Its Legacy*. University of North Carolina Press.
- Skinner, Gideon; Cameron, Daniel und Friedewald, Michael (2014): Deliverable 9.2: The PRISMS survey. European Commission. <https://doi.org/10.24406/publica-fhg-297985>
- Sutter, Evelien de; Zaçe, Drieda; Boccia, Stefania; Di Pietro, Maria Luisa; Geerts, David; Borry, Pascal und Huys, Isabelle (2020): Implementation of Electronic Informed Consent in Biomedical Research and Stakeholders' Perspectives: Systematic Review. *Journal of Medical Internet Research*, 22(10), e19129. <https://doi.org/10.2196/19129>
- Wiesing, Urban; Ach, Johann S.; Bormuth, Matthias und Marckmann, Georg (2008): Nürnberger Kodex. In: Urban Wiesing (Hrsg.): *Ethik in der Medizin: Ein Studienbuch*. Stuttgart, Reclam, S. 123–130.

Informierte Einwilligung als Abwägungsproblem: Verstehen, Bewerten und Handeln im Kontext personenbezogener Datenverarbeitung

Daniel Guagnin, Antonios Hazim und Volkan Sayman¹

Zusammenfassung

Der Beitrag rekonstruiert das Einwilligungsparadigma der DSGVO als ein implizites, zweckrationales Handlungsmodell, das voraussetzt, dass Betroffene Risiken und Vorteile der Datenverarbeitung verstehen, bewerten und konsistent in Entscheidungen überführen können. Vor dem Hintergrund empirisch beobachtbarer Inkonsistenzen entwickelt der Beitrag ein Handlungsmodell (Verstehen, Bewerten, Handeln), das Entscheidungsprozesse differenziert erfasst und insbesondere die bislang unterbelichtete Bewertungsdimension in den Mittelpunkt stellt. Damit wird eine konzeptionelle Grundlage geschaffen, um die Wirksamkeit von Informations- und Einwilligungsdialogen nicht nur am Verständnis von Datenverarbeitung, sondern an der tatsächlichen Befähigung zur Risikoabwägung zu messen. Das Modell dient als Ausgangspunkt für empirische Untersuchungen im SID-Projekt, in denen Einwilligungsdialoge und deren Gestaltung im Hinblick auf ihre Wirkung auf Entscheidungsprozesse analysiert werden.

1. Einleitung – Informierung und Risikoabwägung

Beim Surfen hinterlassen Nutzer:innen kontinuierlich Daten, deren Verarbeitung für sie weitgehend unsichtbar bleibt, obwohl die DSGVO mit Transparenz-, Zweckbindungs- und Interventionsanforderungen gerade auf den Schutz vor daraus entstehenden Grundrechtsrisiken zielt. Gleichzeitig bleibt die konkrete Ausgestaltung und Wahrnehmung dieser Risiken für Betroffene häufig intransparent (Roßnagel 2007; Degeling 2016; Christl

1 Die Autoren haben in gleichen Teilen zum Artikel beigetragen, die Reihenfolge ist alphabetisch. Dieser Artikel entstand im Rahmen des BMFTR-geförderten Projekts „Sicher im Datenverkehr“ (SiD, FKZ 16KIS1967K), mehr Infos dazu unter <https://sid-p.rojekt.de> (letzter Abruf 31.1.2026).

2017). In der Praxis verfehlen gängige Umsetzungen wie Cookie-Banner dieses Ziel jedoch häufig, da sie weder ein wirksames Verständnis noch eine informierte Kontrolle ermöglichen (Utz u. a. 2019). Zwar verlangt die DSGVO mit der „informierten Einwilligung“ (Art. 4 Nr. 11, Art. 6 Abs. 1 lit. a DSGVO) und „Datenschutz by Design“ (Art. 25 DSGVO), dass Entscheidungen auf Grundlage von Information getroffen und Risiken wirksam adressiert werden; zugleich bleibt jedoch weitgehend unklar, welches konkrete Verständnis Betroffene hierfür entwickeln sollen. Die Regulierung fokussiert primär auf die Bereitstellung von Informationen, nicht auf deren tatsächliches Verständnis und Bewertung.

Der Beitrag geht daher der Frage nach, unter welchen Bedingungen Betroffene in der Lage sind, aus Informationsdialogen Chancen und Risiken der Verarbeitung ihrer Daten zu Zwecken der Personalisierung von Werbung abzuleiten, zu bewerten und daran anschließend konsistent zu handeln. Der Beitrag erweitert bestehende Ansätze der Analyse von Privatheitskompetenzen und Einwilligungsentscheidungen um ein Handlungsmodell, das als Heuristik für die empirische Erforschung der Barrieren für eine rationale Abwägung von Risiken und Vorteilen dient, und damit ermöglicht, die Wirksamkeit der Informierung empirisch zu überprüfen. Damit nehmen wir die Initiierung von Abwägungs- und Reflexionsprozessen bei Betroffenen im Hinblick auf Grundrechtsrisiken der personenbezogenen Datenverarbeitung in den Blick, um dies als Gegenstand der Wirksamkeitsforschung zu erschließen.

Der Beitrag ist wie folgt aufgebaut. Zunächst möchten wir kurz das Paradigma der „wirksamen Informierung“ rekonstruieren, indem wir die einschlägigen Transparenz- und Informationspflichten der DSGVO skizzieren. Darüber hinaus argumentieren wir in diesem Abschnitt, dass die DSGVO nicht ausschließlich Pflichten der datenverarbeitenden Organisationen und Rechte Betroffener expliziert, sondern (tlw. implizit) auch homogene Anforderungen an Kompetenzen (insb. deren kognitive Aufnahme-, Handlungs- und Reflexionsfähigkeit) der Betroffenen voraussetzt (Prince u. a. 2024). Anschließend stellen wir den bisherigen Stand der Forschung zur Messung von Privatheitskompetenzen (Dinev/Hart 2004; Malhotra u. a. 2004; Hong/Thong 2013; Masur u. a. 2017; Parsons u. a. 2017; Biselli u. a. 2022) systematisch dar und diskutieren diesen kritisch. Dass tatsächlich Handelnde unter realen Bedingungen der Nutzung des Internets häufig von der Handlungskette eines rationalen Dreischritts von Verstehen, Bewerten und Handeln abweichen, nehmen wir als Ausgangspunkt. Daraus ergibt

sich das Desiderat einer genaueren empirischen Betrachtung der Reflexions- und Abwägungsprozesse von Vorteilen und Risiken.

2. Das Paradigma der „wirksamen Informierung“

Im Folgenden wird das Paradigma der „wirksamen Informierung“ aus rechtlicher und empirischer Perspektive rekonstruiert. Zunächst werden die einschlägigen Anforderungen der DSGVO an Transparenz und Informierung dargestellt und anschließend der Stand der Forschung zu Privatsphärenkompetenzen und zur Wirkung von Informations- und Entscheidungsarchitekturen skizziert. Auf dieser Grundlage werden die impliziten Annahmen über das Entscheidungsverhalten von Betroffenen sowie die Grenzen bestehender Ansätze explizit gemacht, die dem vorherrschenden rationalistischen Handlungsmodell implizit zugrunde liegen.

2.1 „Wirksame Informierung“ nach „Stand der Technik“

Das Paradigma der „wirksamen Informierung“ ergibt sich nicht aus einer einzelnen Norm, sondern mittelbar aus einer Reihe von Vorschriften der DSGVO, insbesondere aus dem Transparenzgrundsatz, der Pflicht zur Angabe der Verarbeitungszwecke, den Informationspflichten, den Auskunft- und Betroffenenrechten, den Anforderungen an technische und organisatorische Maßnahmen sowie den Vorgaben zu Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen (vgl. Art. 5, 12-22, 24, 25, 34 DSGVO). Konkret wird verlangt, dass Informationen in präziser, transparenter, verständlicher und leicht zugänglicher Form sowie in klarer und einfacher Sprache zu übermitteln sind, um die Ausübung der Rechte der betroffenen Person wirksam zu gewährleisten (Erwägungsgrund 74, Art. 12 Abs. 1 DSGVO).

Die tatsächliche Wirksamkeit von Informations- und Einwilligungsdialogen wollen wir empirisch untersuchen und nachweisbar machen, um auf Basis der Ergebnisse Verbesserungen ihrer Gestaltung zu ermöglichen. Die rechtliche Bewertung sollte aber nicht bei der formalen Umsetzung von Informationspflichten stehen bleiben, sondern die tatsächlichen Auswirkungen der getroffenen Maßnahmen in den Blick nehmen (Grafenstein u. a. 2021). Daher impliziert das Wirksamkeitserfordernis die Anwendung empirischer sozialwissenschaftlicher Methoden zur Überprüfung der tatsäch-

lichen Effekte der Informierung (Art. 29 Data Protection Working Party 2017, S. 7; EDPB 2020 S. 7, Grafenstein u. a. 2021 S. 18).

Artikel 25, Abs. 1. DSGVO verpflichtet Verantwortliche der Datenverarbeitung, geeignete technisch-organisatorische Maßnahmen einzusetzen, die die Grundsätze des Datenschutzes „wirksam“ umsetzen („Datenschutz-by-Design“). Dabei sind adäquate, auf dem Markt verfügbare technische Schnittstellen im Sinne des Standes der Technik einzusetzen. Die Ermittlung der wirksamsten Arten und Weisen, Technik zu gestalten, hängt entscheidend von zwei Faktoren ab – erstens davon, welche Methoden zur Messung der Wirksamkeit in welcher Form und Kombination angewandt werden, zweitens, inwiefern die angewandten Methoden es erlauben, die für Wirksamkeit relevanten Dimensionen zu erfassen. Darüber hinaus ist die systematische Berücksichtigung von Gestaltungselementen und -parametern erforderlich, die die Informierung und Einwilligung beeinflussen können (vgl. Grafenstein u. a. 2023). Für die Analyse des Prozesses der Informierung und Einwilligung der Betroffenen und die Ermittlung der Wirksamkeit der Informationsdialoge erscheint uns die Betrachtung der Handlungen der Betroffenen essenziell. Aufbauend auf dem hier vorgestellten Handlungsmodell lassen sich dann konkrete Hypothesen und Forschungsfragen operationalisieren, die der Wirksamkeitsmessung sozio-technischer Umsetzungsvarianten von Informations- und Einwilligungsarchitekturen dienen.

2.2 Informierung mittels „Privatheitskompetenzen“ und Nutzungsschnittstellen?

Die Fokussierung des Rechts auf die Transparenz und Kontrolle der Datenverarbeitungen spiegelt sich wider im Diskurs zu Privatheitskompetenzen, die offenbar notwendig erscheinen für eine fundierte individuelle Abwägung von Vorteilen und Risiken.

Die Literatur hierzu lässt sich grob in zwei Gruppen unterteilen: erstens Ansätze, die auf das Wissen bzw. die bei Betroffenen vorhandenen Privatheitskompetenzen fokussieren und zweitens Ansätze, die das beobachtbare Verhalten von Betroffenen in der Interaktion mit Einwilligungsdialogen in

den Blick nehmen und als den zentralen das Verhalten beeinflussenden Faktor einbeziehen.²

Die Wirksamkeit von Informationsdialogen wurde meist in engem Zusammenhang mit den individuellen Einstellungen und Fähigkeiten der Betroffenen betrachtet. Ein zentraler Forschungsstrang operationalisiert diese Voraussetzungen über Privatheitsbedenken („Privacy Concerns“) als psychologisches Konstrukt, das die subjektive Problemwahrnehmung und Risikosensibilität gegenüber Datenverarbeitung abbildet. Frühere Skalenmodelle – etwa von Dinev und Hart (2004) – konzeptualisieren Privacy Concerns als mehrdimensionales Konstrukt, das Dimensionen wie allgemeine Privatheitsbedenken, wahrgenommene Vulnerabilität und wahrgenommene Kontrollmöglichkeiten umfasst.

Ein häufig genutzter Standard zur Messung von Privacy Concerns ist das IUIPC-Instrument (Malhotra u. a. 2004), das Bedenken entlang der Dimensionen Datenerhebung (eng. „Collection“), Kontrolle (eng. „Control“) und Bewusstsein (eng. „Awareness“) erfasst und in Studien zum Online-Verhalten eingesetzt wird. Erweiternd differenziert das IPC-Modell (Hong/Thong 2013) diese Perspektive in sechs Dimensionen und berücksichtigt zusätzlich Risiken wie Sekundärnutzung, Datenfehler und unberechtigten Zugriff. Beide Ansätze operationalisieren damit primär subjektive Wahrnehmungen und Bewertungen von Datenverarbeitung, erfassen jedoch nicht systematisch die konkreten Abwägungsprozesse zwischen wahrgenommenen Risiken und Vorteilen.

Neben diesen Ansätzen, die auf Einstellungen, Wahrnehmungen und Bedenken fokussieren, wurden auch Metriken entwickelt, die Privatheitskompetenzen als „Privacy Literacy“ bzw. als Bündel von Wissen und Handlungsstrategien konzeptualisieren. Ein prominentes Instrument ist die Online Privacy Literacy Scale (OPLIS) von Masur und Trepte (2017), die in einem Mixed-Methods-Ansatz vier Kompetenzbereiche unterscheidet: (1) Wissen über institutionelle Praktiken, (2) technisches Wissen, (3) rechtliches Wissen sowie (4) Privatheitsstrategien im Sinne praktischer Schutzhandlungen. OPLIS verschiebt damit den Fokus von Einstellungen und Bedenken hin zu der Frage, ob und in welcher Form Betroffene wichtige Voraussetzungen mitbringen, um Risiken der Datenverarbeitung überhaupt erkennen, erklären und durch konkrete Schutzstrategien bearbeiten zu können (vergleiche den Beitrag von Albert u. a. in diesem Band).

2 Zur handlungsbeeinflussenden Wirkung von Technik vgl. Berlin Script Collective 2018.

Ein angrenzender Forschungsstrang integriert Wissen, Einstellungen und Verhalten in gemeinsamen Messmodellen im Bereich Informationssicherheit, wie etwa das HAIS-Q (Parsons u. a. 2017), das Awareness, Einstellungen und tatsächliche Praktiken systematisch zusammenführt. Auch wenn es primär auf organisationale Kontexte zielt, verdeutlicht es die Möglichkeit, unterschiedliche Dimensionen von Kompetenzen in einem Modell zu integrieren.

Neuere Arbeiten differenzieren Kompetenzen und Handlungsdispositionen stärker entlang heterogener Nutzendentypen, etwa durch Privacy Personas, die Wissens- und Fähigkeitsdimensionen mit motivationalen Unterschieden verbinden (Biselli u. a. 2022). Solche personabasierte Instrumente sind insbesondere für die menschenzentrierte Gestaltung und Evaluation von Informationsdialogen relevant, weil sie Unterschiede in Motivation, Risikoaversion und Selbstregulationsfähigkeit als erklärungsbedürftige Varianzquellen sichtbar machen und mit variierenden Typen Betroffener erklären.

Parallel zeigt sich eine Verschiebung der Messlogik von allgemeinen Bedenken hin zu kontextbezogenen Kompetenzen, die Wissen und Verhalten differenzieren und stärker an konkreten Datenverarbeitungspraktiken ausrichten. Darauf aufbauend entwickeln neuere Ansätze nutzendenzentrierte Spezifikationen, Benennungen und Visualisierungen von Verarbeitungszwecken sowie entsprechende Tools (z. B. Barth u. a. 2021; Grafenstein u. a. 2021; Kyi u. a. 2024; Grafenstein u. a. 2023). Diese Ansätze erfassen jedoch entweder Einstellungen, Wissen oder Verhalten, integrieren aber nicht systematisch die dazwischenliegenden Abwägungs- und Bewertungsprozesse, die für informierte Entscheidungen zentral sind.

3. Das zweckrationale Handlungsmodell der informierten Einwilligung

Im Folgenden identifizieren wir auf Basis einer kritischen Betrachtung der diskutierten Literatur Abwägungsprozesse als Forschungslücke und skizzieren ein Handlungsmodell für die Erforschung der Barrieren im Prozess des Verstehens, Abwägens von Vorteilen und Risiken (Bewerten) sowie der schlussendlich getroffenen Einwilligungsentscheidungen (Handlung).

3.1 Abwägungsprozesse als Forschungslücke

Die bestehende Forschung zur Messung von Wirksamkeit von Informations- und Einwilligungsdialogen liefert differenzierte Instrumente, fokussiert jedoch überwiegend auf zwei Dimensionen: Wissen bzw. als Wissen operationalisierte „Privacy Literacy“ und Verhalten. Damit werden relevante Einflussfaktoren adressiert, jedoch selten in einem Modell integriert, das den Abwägungsprozess informierter Entscheidungen selbst überprüfbar macht. Insbesondere bleibt die Bewertungsdimension unterbelichtet, obwohl Wertbindungen als handlungsleitende Orientierungen (Meier/Krämer 2024) sowie routinierte, habituelle Praktiken empirisch plausibel erklären können, warum Betroffene trotz vorhandener Informationen nicht zwangsläufig Vorteile und Risiken systematisch abwägen, sondern auf Heuristiken, Routinen oder generalisierte Haltungen zurückgreifen.

Zwar thematisieren einzelne Arbeiten Abwägungsprozesse, etwa im Kontext von Privacy–Security–Trade-offs oder der Bewertung von Überwachungstechnologien (Pavone/Esposti 2012; van den Broek u. a. 2017), jedoch bleiben diese Ansätze meist auf der Ebene von Einstellungen und Präferenzen und werden selten als prozessuale Entscheidungsabläufe modelliert. Insbesondere fehlt eine systematische Integration von kognitiven, evaluativen und handlungsbezogenen Dimensionen sowie eine empirisch überprüfbare Analyse der Übergänge zwischen ihnen.

Damit fehlt eine Grundlage, um die Wirksamkeit von Informationsdialogen im Sinne ihrer tatsächlichen Effekte auf Entscheidungsprozesse analytisch zu verstehen und empirisch zu messen (Grassl u. a. 2024). Diese Perspektive verschiebt den Fokus von rationalistischen Entscheidungsmodellen hin zu einer stärker praxeologischen Betrachtung von Entscheidungsprozessen.

Zugleich erfordert die Analyse solcher Prozesse eine stärkere Berücksichtigung von Mensch–Technik-Interaktionen. Informationsdialoge sind keine neutralen Vermittlungsinstrumente, sondern Entscheidungsumwelten, die Aufmerksamkeit lenken, Bewertungen strukturieren und Handlungsmöglichkeiten präfigurieren (Leimstädtner u. a. 2023). Ihre Wirkung lässt sich daher nur erfassen, wenn die Interaktion selbst als Teil des Entscheidungsprozesses untersucht und das Verhalten in realitätsnahen Mensch–Technik-Interaktionen hypothesengestützt variiert wird (Kokolakis 2017).

Vor diesem Hintergrund schlagen wir vor, die informierte Einwilligung als Abfolge von Verstehen, Bewerten und Handeln zu rekonstruieren und die Bewertungsdimension als bislang unterbelichtete Scharnierstelle zwi-

schen Verständnis und Verhalten systematisch zu untersuchen. Dies haben wir in Abb. 1 schematisch dargestellt. Ein solches Handlungsmodell ermöglicht es, Konsistenzen und Inkonsistenzen entlang des Entscheidungsverlaufs sichtbar zu machen und Barrieren innerhalb dieses Prozesses analytisch zu identifizieren.

3.2 Rekonstruktion des idealtypischen Entscheidungsprozesses

Das vorherrschende individualistische Datenkontrollparadigma (Pohle 2022) impliziert, dass Betroffene sich ausreichend informieren und auf dieser Grundlage eigenständig in Datenverarbeitungen einwilligen können. Ob die Bedingungen für eine solche informierte Entscheidung tatsächlich gegeben sind, wird jedoch seit längerem kritisch diskutiert (Bergemann 2018). Besonders deutlich wird dies im sogenannten Privacy Paradox, das die Diskrepanz zwischen artikuliertem Risikobewusstsein und beobachtetem Einwilligungsverhalten beschreibt (Bergemann 2018; Kokolakis 2017).

Implizit setzt dieses Paradigma voraus, dass Betroffene im Fall wirksamer Information relevante Risiken erkennen und gegen wahrgenommene Vorteile abwägen können. In diesem Sinne lässt sich argumentieren, dass in den Rechtstext der DSGVO ein bestimmtes Subjektverständnis eingeschrieben ist (Engel/Schreiber 2024: 135). Dieses Verständnis knüpft an das in der Privatheitsforschung etablierte Konzept des „Privacy Calculus“ an, das davon ausgeht, dass Individuen Entscheidungen über die Preisgabe personenbezogener Daten auf Basis einer Abwägung wahrgenommener Risiken und Vorteile treffen (vgl. Dinev/Hart 2006). Informierte Einwilligung wird damit – analytisch rekonstruiert – als Ausdruck eines zweckrationalen Handlungsmodells gefasst, dessen Tragfähigkeit durch die empirisch beobachteten Inkonsistenzen jedoch in Frage gestellt wird.

Das Paradigma kann entlang dreier aufeinander aufbauender Schritte konzeptualisiert werden: *Verstehen*, *Bewerten* und *Handeln*. Die informierte Entscheidung im Kontext der Datenverarbeitung durchläuft hiernach einen mehrstufigen kognitiven und evaluativen Prozess (vgl. Abb.1).

Die drei Dimensionen knüpfen an etablierte Unterscheidungen kognitiver, evaluativer und handlungsbezogener Komponenten von Kompetenz

an, werden hier jedoch prozessual gefasst und auf datenschutzbezogene Entscheidungsprozesse angewendet.³



Abbildung 1: Das idealtypische Handlungsmodell: „Verstehen, Bewerten, Handeln“.

Verstehen beschreibt die Fähigkeit von Betroffenen, Verarbeitungszwecke und -prozesse nachzuvollziehen und deren Bedeutung für die eigene Person einzuordnen. Dazu gehört insbesondere das Erkennen potenzieller Grundrechtsrisiken (z. B. Diskriminierung oder Manipulation) sowie das Verständnis der damit verbundenen Vor- und Nachteile.

Bewerten bezieht sich auf die individuelle Abwägung dieser Risiken und Chancen. Entscheidend ist nicht eine normativ „richtige“ Bewertung, sondern dass Betroffene Risiken und Vorteile entlang eigener Präferenzen gewichten, diese Einschätzung begründen und mögliche Handlungsoptionen sowie deren Konsequenzen reflektieren.

Handeln schließlich meint die Überführung dieser Bewertung in konkrete Entscheidungen im Rahmen der Interaktion mit technischen Systemen.

Dieser Idealtypus ist kein „Durchschnitt“ des empirisch Vorhandenen, sondern ein analytisches Konstrukt, das zentrale Annahmen des beschriebenen Paradigmas sowie der datenschutzrechtlichen Anforderungen verdichtet (vgl. Stapley u. a. 2022).

Das zugrunde gelegte Rationalitätsverständnis erhebt normativ einen minimalen Anspruch: Es unterstellt lediglich, dass Betroffene in der Lage sind, Informationen in eine intern konsistente Abfolge von Verstehen, Bewerten und Handeln zu überführen, wobei die konkrete Abwägung von Risiken und Vorteilen ihren subjektiven Präferenzen überlassen bleibt. Das Modell dient somit als Referenzrahmen, von dem aus systematisch Abweichungen vom idealtypischen Verlauf als unterschiedliche Entscheidungspfade modelliert werden (Abb. 2).

3 Vgl. bspw. Weinert (2001), Klieme u. a. (2003), Bybee, R. (1997).

4. „Verstehen, Bewerten und Handeln“ – Ein empirischer Ansatz

Auf Grundlage des im vorangegangenen Abschnitt entwickelten Modells wird im Folgenden untersucht, wie sich Entscheidungsprozesse empirisch von diesem idealtypischen Verlauf unterscheiden. Ziel ist es, typische Abweichungen zwischen Verstehen, Bewerten und Handeln systematisch zu erfassen und als unterschiedliche Entscheidungspfade zu beschreiben. Dadurch lassen sich nicht nur Muster inkonsistenter oder nicht-linearer Entscheidungen identifizieren, sondern auch die ihnen zugrunde liegenden Barrieren analysieren. Abschnitt 4.1 entwickelt hierzu eine Typologie empirischer Abweichungen, während Abschnitt 4.2 diese mit bestehenden Befunden der Privatheitsforschung verknüpft und entlang der einzelnen Prozessschritte verortet.

4.1 Idealtypus und empirische Abweichungen

Das im vorigen Abschnitt vorgestellte Handlungsmodell beruht auf einem zweckrationalen Ansatz, der uns ermöglicht, verschiedene abweichende Verlaufsformen zu identifizieren, die sich aus dem Zusammenspiel der drei Dimensionen ergeben. In Abb. 2 werden vier empirisch beobachtbare Abweichungsmuster illustriert, die wir nachfolgend erläutern und anschließend mit Befunden aus der Forschung untermauern.

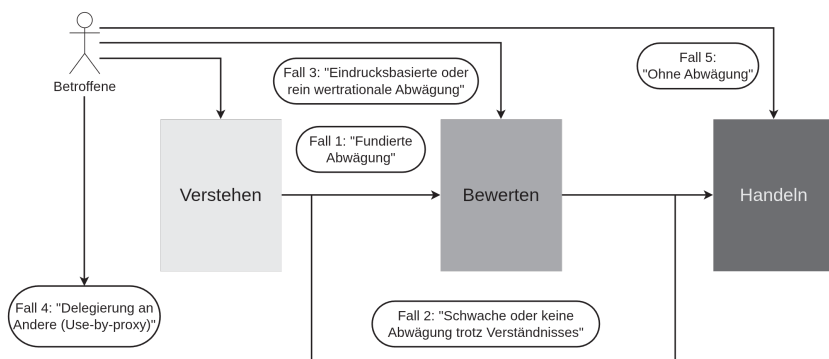


Abbildung 2: Abweichungen vom Idealtyp „Verstehen, Bewerten, Handeln“

Fall 1 – Idealtypus „Fundierte Abwägung“: Alle drei Dimensionen werden erfolgreich durchlaufen. Betroffene verstehen die Grundrechtsrisiken, entwickeln eine konsistente Bewertung und können diese in entsprechende Handlungen überführen. Dieser Fall ist idealtypisch zu verstehen, und in der Realität voraussichtlich kaum beobachtbar. Die folgenden vier Fälle beschreiben hingegen realistischere Verlaufsformen, die als Abweichungen vom Idealtyp verstanden werden können:

Fall 2 – „Schwache oder keine Abwägung trotz Verständnisses“: Betroffene verstehen die Grundrechtsrisiken, entwickeln jedoch keine oder nur eine schwache Bewertung. Dies kann auf wahrgenommene Machtlosigkeit, Resignation oder strukturelle Barrieren zurückzuführen sein.

Fall 3 – „Eindrucksbasierte oder rein wertrationale Abwägung“: Betroffene gelangen zu einer Bewertung und Handlung, ohne die zugrunde liegenden Risiken vollständig zu verstehen. Die Bewertung basiert primär auf Heuristiken, Affekten oder allgemeinen Werthaltungen.

Fall 4 – „Delegierung an Andere (Use-by-proxy)“: Betroffene delegieren die Entscheidung an andere Personen oder Systeme, ohne selbst den Prozess des Verstehens, Bewertens und Handelns zu durchlaufen.

Fall 5 – „Ohne Abwägung“: Entscheidungen werden ohne erkennbaren Bezug zu Verständnis oder Bewertung getroffen, etwa durch unreflektiertes Akzeptieren von Voreinstellungen.

Diese Typologie dient nicht der normativen Bewertung individueller Entscheidungsstrategien, sondern der analytischen Differenzierung verschiedener Muster, die empirisch auftreten können. Sie ermöglicht es, Barrieren in spezifischen Phasen des Prozesses zu identifizieren und Design-Interventionen gezielt auszurichten.

Konsistenz als analytisches Kriterium

Ein zentrales Anliegen des Modells ist die Erfassung von Konsistenz bzw. Inkonsistenz zwischen den drei Dimensionen. Von besonderem Interesse sind Fälle, in denen beobachtbare Inkonsistenzen auftreten – etwa, wenn Betroffene Risiken artikulieren können (Verstehen), diese jedoch nicht in ihre Bewertung einfließen lassen, oder wenn getroffene Entscheidungen im Widerspruch zur geäußerten Bewertung stehen. Solche Inkonsistenzen können auf verschiedene Faktoren zurückzuführen sein: auf Barrieren in

der Informationsarchitektur, auf manipulative Designelemente, auf kognitive Überforderung oder auf strukturelle Machtasymmetrien zwischen Betroffenen und datenverarbeitenden Organisationen. Die Identifikation dieser Inkonsistenzen ist für die Weiterentwicklung von Informationsdialogen von zentraler Bedeutung.

4.2 Barrieren rationalen Handelns als Ergebnis der Privatheitsforschung

Das vorgestellte Handlungsmodell integriert Erkenntnisse aus verschiedenen Forschungssträngen der Privatheitsforschung und erweitert diese um eine handlungsorientierte Perspektive. Im Folgenden werden zentrale Bezüge zu bestehenden Konzepten hergestellt und deren Verortung innerhalb der drei Interaktionsschritte erläutert.

Privatheitskompetenz als mehrdimensionales Konstrukt

Bisherige Ansätze zur Messung von Privatheitskompetenz haben sich primär auf zwei Bereiche konzentriert: technisches Wissen und Schutzverhalten. Im DigComp 2.2 Framework (European Commission 2022) hat das Joint Research Centre der Europäischen Kommission beispielsweise eine Kompetenzstufe „Protecting personal data and privacy“ definiert, die auf Level 4 (Intermediate) die Fähigkeit voraussetzt, Schutzmöglichkeiten zu diskutieren und sich selbst sowie andere vor Schäden zu bewahren. Während Biselli u. a. (2022) ein Messinstrument entwickelten, das Privacy-Personas (Fundamentalist, Pragmatist, Unconcerned) auf den Dimensionen Wissen und Verhalten identifiziert, differenzierten Papaioannou u. a. (2025) zwischen „sichtbaren“ (eng. „visible“; Fähigkeit zur Installation von Schutzmaßnahmen, Risikowahrnehmung) und „verborgenen“ (eng. „hidden“; Wahrgenommene Privatsphäre, Selbstkontrolle, Privatheitsbedenken) Privatheitskompetenzen.

Diese Ansätze erfassen wichtige Aspekte von Privatheitskompetenz, fokussieren jedoch primär auf die Dimension des Verstehens (Wissen über Risiken) und Handelns (Schutzverhalten). Die Dimension des Bewertens – also die Fähigkeit, verstandene Risiken in Bezug zur eigenen Lebenswelt zu setzen und eine begründete Abwägung vorzunehmen – bleibt weitgehend unterbelichtet. Unser Modell erweitert diese Perspektive, indem es Bewerten als eigenständige und zentrale Kompetenzdimension konzeptualisiert.

Barrieren im Übergang zwischen Verstehen und Bewerten

Die Forschung zu „Privacy Helplessness“ (Cho 2022; Meng 2024) zeigt, dass selbst bei vorhandenem Verständnis von Risiken eine Bewertung unterbleiben kann. Cho (2022) identifiziert „sense of privacy helplessness“ als zentralen Faktor, der verhindert, dass Betroffene trotz Kenntnis von Risiken diese in Bewertungen überführen, etwa aufgrund eines mangelnden Selbstwirksamkeitsgefühls. Meng (2024) differenziert hierbei vier Dimensionen von „Privacy Boundary Turbulence“: Überwachungsgefühl (eng. „sense of surveillance“), Verratsgefühl (eng. „sense of betrayal“), Schadensgefühl (eng. „sense of harm“) und Hilflosigkeitsgefühl (eng. „sense of helplessness“). Diese Konzepte verorten wir in *Fall 2* unserer Typologie („Schwache oder keine Abwägung trotz Verständnisses“), da sie erklären, warum die Transition von Verstehen zu Bewerten scheitern kann und die Betroffenen direkt aus dem Verständnis heraus zum Handeln übergehen.

Quinn und Epstein (2023) argumentieren, dass die Beziehung zwischen Privatheitsbedenken und Nutzungsverhalten stark vom Kontext abhängt. In „vertikalen“ Privatsphärebeziehungen (Individuum vs. institutionelle Akteure) führen höhere Privatheitsbedenken zu reduzierter Nutzung, während in „horizontalen“ Beziehungen (Individuum vs. soziales Netzwerk) emotionale Zugehörigkeitsbedürfnisse rationale Risikoabwägungen überwiegen. Diese Unterscheidung ist für unser Modell relevant, da sie zeigt, dass die Konsistenz zwischen Verstehen, Bewerten und Handeln kontextabhängig variiert und durch strukturelle Machtasymmetrien beeinflusst wird.

Barrieren im Übergang zwischen Bewerten und Handeln

Bjørlo (2024) erweitert das von Roessler (2005) eingeführte Konzept der „Decisional Privacy“ und unterscheidet auf praktischer Ebene zwischen „Information Privacy“ (Schutz von Daten) und „Decisional Privacy“ (Autonomie von Manipulation). Dabei unterscheidet sie verschiedene Arten von Beeinflussung: 1) offene (eng. „overt“) vs. verdeckte (eng. „covert“) Beeinflussung, 2) rationale Überzeugung vs. Manipulation sowie 3) System I (automatische) vs. System II (deliberative) Entscheidungsfindung (nach Kahneman 2011). Manipulative Designelemente können verhindern, dass entwickelte Bewertungen in entsprechende Handlungen überführt werden. Sun u. a. (2024) zeigen in ihrer Forschung zur „Privacy Empowerment Illusion“, dass Hinweise mit hoher Interpretierbarkeit aber geringer Verständlichkeit zu erhöhter Selbstoffenbarung führen – ein Effekt, der verschwin-

det, wenn die Verständlichkeit hoch ist. Diese Befunde legen nahe, dass Design-Interventionen differenziert zwischen den Dimensionen wirken: Was das Verstehen fördert, kann das Handeln beeinflussen, ohne notwendigerweise die Bewertung zu verändern. Dies verdeutlicht die Notwendigkeit differenzierter Ansätze bei der Evaluierung von Wirksamkeit von Informations- und Entscheidungssystemen.

Abweichungen vom linearen Modell sind die Regel

Nicht alle Entscheidungsprozesse folgen der idealtypischen Sequenz Verstehen → Bewerten → Handeln. Die Forschung zu „Use-by-Proxy“ (Bartol u. a. 2024) zeigt, dass insbesondere ältere Menschen, die sich stark um ihre Privatsphäre sorgen, die Nutzung digitaler Dienste signifikant häufiger an andere delegieren. Alamari und Alshehri (2024) schlagen vor dem Hintergrund des geringen Vertrauens in die Datenschutzpraktiken der Social-Media-Plattformen automatisierte und personalisierte Consent-Mechanismen vor, die auf initialen Präferenzbefragungen basieren. Diese beiden Realitäten entsprechen dem *Fall 4* unserer Typologie und zeigen, dass Delegation eine legitime Strategie sein kann, wenn individuelle Kompetenzen oder Ressourcen limitiert sind. Eine andere erwähnenswerte Form der Delegation besteht in gängigen Praktiken wie bspw. der Nutzung von Adblockern. Dabei machen wir keinen Unterschied zwischen der Delegation an Menschen oder Technik, da aus Sicht der Modellierung einer Entscheidung beide in einem externen System stattfinden.

Während die bisherigen Fälle zeigen, wie Übergänge zwischen den Dimensionen scheitern können, dokumentiert *Fall 3* einen alternativen Pfad: Betroffene entwickeln Bewertungen und treffen Entscheidungen, ohne die zugrunde liegenden Risiken vollständig zu verstehen. Diese Bewertungen basieren primär auf Heuristiken, Affekten, allgemeinen Werthaltungen oder Vertrauen. Leimstädtner u. a. (2023) und Esmaeilzadeh (2019) zeigen, dass die Fähigkeit zur Bewertung nicht ausschließlich von der Dimension des Verstehens abhängt, sondern auch durch Werthaltungen, emotionale Prozesse und Vertrauensbeziehungen konstituiert wird. Wertbasierte oder vertrauensgeleitete Bewertungen können unter bestimmten Bedingungen funktional und angemessen sein – insbesondere, wenn sie durch Reflexionsprozesse unterstützt werden und mit den grundlegenden Werten der Person übereinstimmen.

Integration: Von Kompetenzmessung hin zu Barrierenidentifikation

Während bestehende Ansätze primär auf die Messung individueller Kompetenzen abzielen, verfolgt unser Modell einen erweiterten Ansatz: Es ermöglicht nicht nur die Identifikation von Kompetenzlücken, sondern auch die Lokalisierung von Barrieren im Entscheidungsprozess. Indem wir systematisch erfassen, an welcher Stelle Inkonsistenzen zwischen Verstehen, Bewerten und Handeln auftreten, können wir zwischen individuellen Kompetenzdefiziten, strukturellen Barrieren und manipulativen Designelementen differenzieren.

Diese analytische Unterscheidung ist für die Weiterentwicklung wirksamer Informationsdialoge zentral, da sie ermöglicht, Interventionen gezielt auszurichten: Wo Verständnisprobleme vorliegen, sind andere Maßnahmen erforderlich als dort, wo Bewertungen trotz Verständnis unterbleiben oder wo getroffene Entscheidungen systematisch von artikulierten Präferenzen abweichen.

5. Einflussfaktoren und Konsistenz bei Einwilligungsentscheidungen

Das vorgestellte Handlungsmodell ermöglicht es, empirische Forschungsfragen auf zwei komplementären Ebenen zu formulieren. Der erste Zugang fokussiert auf den *Einfluss von Design und Informationsarchitektur auf die drei Interaktionsschritte*. Der zweite Zugang untersucht die *Konsistenz zwischen Verstehen, Bewerten und Handeln*. Beide Zugänge sind für die empirische Untersuchung der Wirksamkeit von Informierungsprozessen zentral.

5.1 Zugang 1: Einfluss des Designs auf Verstehen, Bewerten und Handeln

Dieser Forschungszugang untersucht, inwieweit Design-Elemente und Gestaltungsvarianten der Informationsarchitektur die Entwicklung der drei Interaktionsschritte fördern oder behindern. Die zentrale Forschungsfrage lautet: Welche Design-Elemente fördern das Verständnis, die Bewertung und das Handeln im Hinblick auf Grundrechtsrisiken der Verarbeitung personenbezogener Daten zum Zwecke der Personalisierung?

Diese Hauptfrage lässt sich in drei spezifische Teilfragen differenzieren, eine für den jeweiligen Schritt des Modells. Dieser Zugang ermöglicht

es zu identifizieren, welche Design-Interventionen spezifisch auf welche Interaktionsschritte wirken und ob bestimmte Gestaltungselemente nicht intendierte Effekte haben (z. B. erhöhtes Verständnis bei gleichzeitig erschwertem Handeln durch Komplexität).

5.2 Zugang 2: Konsistenz zwischen Verstehen, Bewerten und Handeln

Dieser Forschungszugang untersucht, ob und unter welchen Bedingungen Betroffene in der Lage sind, die drei Schritte des Verstehens, Bewertens und Handelns konsistent zusammenzubringen. Die zentrale Forschungsfrage lautet hier: *Handeln Betroffene konsistent mit ihrem Verständnis und ihrer Bewertung von Grundrechtsrisiken?*

Dazu werden die Übergänge zwischen den Schritten des Modells untersucht. Zunächst wird geprüft, ob ein Vorverständnis von Grundrechtsrisiken das Verstehen des Informationsdialogs beeinflusst (Wissen → Verstehen). Anschließend wird analysiert, ob das Verständnis von Grundrechtsrisiken zu einer konsistenten Präferenzbildung führt und welche Faktoren eine Bewertung trotz Verständnis verhindern. Schließlich wird untersucht, ob die entwickelten Präferenzen in entsprechende Handlungen umgesetzt werden können und welche Barrieren dies erschweren.

Die Analyse der Inkonsistenzen in diesem Zugang ermöglicht es, zwischen verschiedenen Erklärungsansätzen zu differenzieren: (a) Individuelle Kompetenzdefizite (z. B. fehlendes Vorwissen), (b) strukturelle Barrieren (z. B. unzureichende Informationsarchitektur), (c) manipulative Designelemente (z. B. Dark Patterns), (d) kontextuelle Faktoren (z. B. wahrgenommene Machtlosigkeit) oder e) unzureichende Handlungsmöglichkeiten.

5.3 Zusammenführung beider Zugänge

Die beiden Forschungszugänge sind komplementär: Zugang 1 untersucht die Wirksamkeit von Design-Interventionen auf die Entwicklung einzelner Kompetenzdimensionen für einzelne Schritte, während Zugang 2 die Konsistenz des gesamten Entscheidungsprozesses in den Blick nimmt. Zusammen ermöglichen sie eine differenzierte Analyse, welche Design-Elemente nicht nur einzelne Schritte beeinflussen, sondern auch die Konsistenz zwischen den Schritten stärken. Ein Design könnte beispielsweise das Verständnis erhöhen, jedoch aufgrund von Komplexität das Handeln erschwe-

ren – ein aus der Usability-Forschung bekannter Befund, der nur durch das Zusammendenken beider Zugänge sichtbar wird.

6. Diskussion und Ausblick

Die empirische Überprüfung der Wirksamkeit der Informierung bei Einwilligungen bringt einige methodische Herausforderungen mit sich. Zum einen bleibt die Frage, welche Informationen in welchem Kontext dazu geeignet sind, Betroffene in die Lage zu versetzen, einen „zweckrationalen“ Entscheidungsprozess zu durchlaufen. Wenn die technische und juristische Information über Datenverarbeitung nicht dafür ausreicht, wie können wir verständlich über potenzielle Risiken für Grundrechte und Vorteile informieren? Zum anderen stellt sich die Frage, wie wir damit umgehen, wenn weiterhin Inkonsistenzen bestehen, und welche Gestaltungsmöglichkeiten wir technisch haben, und welche auch auf dem Markt etabliert werden können. Darüber hinaus schließen sich Fragen der Handlungsmöglichkeiten der Betroffenen jenseits von „Take-it-or-leave-it“ Entscheidungen (Borgesius u. a. 2017) an. Zuletzt stellt sich eine zentrale Frage, welche Form der Informierung gut genug ist, um eine hinreichende Wirksamkeit zu entfalten? Diese Fragen sind weiter zu untersuchen.

Das vorgestellte Handlungsmodell stellt einen konzeptionellen Rahmen dar, der als Ausgangspunkt für die Untersuchung der Wirksamkeit von Informationsdialogen dient. In anschließenden empirischen Untersuchungen wird sich zeigen, inwiefern die theoretisch postulierten drei Schritte (Verstehen, Bewerten, Handeln) im Handeln Betroffener unterscheidbar sind und verschiedene Dimensionen datenschutzbezogener Kompetenzen und Handlungen zu erfassen erlauben. So kann die Wirksamkeit der Informierung anhand dieser drei Schritte untersucht werden, und Inkonsistenzen und Barrieren bei Abweichungen vom idealtypischen Verstehens-, Abwägungs- und Entscheidungsprozess identifiziert werden. Diese Perspektive der Wirksamkeit von Informierungsdialogen sollte auch Gegenstand ihrer systematischen Überprüfung sein.

Der Beitrag hat gezeigt, dass die Fähigkeit Betroffener zu konsistenter Risikoabwägung empirisch keineswegs vorausgesetzt werden kann. Daraus folgt, dass die Wirksamkeit von Informierung nicht allein über die Bereitstellung von Informationen oder das Verständnis von Datenverarbeitungen bestimmt werden kann, sondern an der tatsächlichen Befähigung zu Abwägungs- und Entscheidungsprozessen gemessen werden muss. Damit ver-

schiebt sich der Maßstab wirksamer Informierung von der Bereitstellung von Information hin zur empirisch überprüfbaren Befähigung zu reflektierten Entscheidungen. Für die weitere Forschung bedeutet dies, Informierung nicht primär als Design- oder Transparenzproblem zu behandeln, sondern als Gegenstand empirischer Wirkungsanalyse von Entscheidungsprozessen unter realen Interaktionsbedingungen.

Im weiteren Projektverlauf operationalisieren wir diesen Ansatz, indem wir Informations- und Einwilligungsdialoge experimentell untersuchen, die auf Basis empirischer Erkenntnisse entwickelt wurden. Dazu gehören insbesondere die Entwicklung von Risikonarrativen zur verbesserten Vermittlung von Grundrechtsrisiken sowie die Weiterentwicklung methodischer Ansätze zur Messung von Wirksamkeit im Kontext regulatorischer Konformitätsbewertung.

Literatur

- Art. 29 Data Protection Working Party (2017): Guidelines on transparency under Regulation, 2016/679.
- Alamari, Jebreel und Alshehri, Aziz (2024): Adaptive privacy permissions for social media web, *Multimedia Tools and Applications*, 84(20), S. 23065–23084. DOI: <https://doi.org/10.1007/s11042-024-20423-4>.
- Barth, Susanne; Ionita, Dan; De Jong, Meno D.; Hartel, Pieter und Junger, Marianne (2021): Privacy rating: A user-centered approach for visualizing data handling practices of online services, *IEEE transactions on professional communication*, 64(4), S. 354–373. URL: <https://ieeexplore.ieee.org/abstract/document/9612118/> (besucht am: 19.12.2023).
- Bartol, Jošt; Prevodnik, Katja; Vehovar, Vasja und Petrovčič, Andraž (2024): The roles of perceived privacy control, Internet privacy concerns and Internet skills in the direct and indirect Internet uses of older adults: Conceptual integration and empirical testing of a theoretical model, *New Media & Society*, 26(8), S. 4490–4510. DOI: <https://doi.org/10.1177/14614448221122734>.
- Bergemann, Benjamin (2018): The Consent Paradox: Accounting for the Prominent Role of Consent in Data Protection, In: M. Hansen, E. Kosta, I. Nait-Fovino, und S. Fischer-Hübner (Hrsg.) *Privacy and Identity Management. The Smart Revolution*. Cham: Springer International Publishing (IFIP Advances in Information and Communication Technology), S. 111–131. DOI: https://doi.org/10.1007/978-3-319-92925-5_8.
- van den Broek, Tijs; Ooms, Merel; Friedewald, Michael; van Lieshout, Marc und Rung, Sven (2017): Privacy and security: Citizens' desires for an equal footing, In: M. Friedewald, J.P. Burgess, J. Cas, R. Bellanova und W. Peissl (Hg.): *Surveillance, Privacy and Security: Citizens' Perspectives*. London: Routledge (PRIO New Security Studies), S. 15–35.

- Berlin Script Collective (2018): Technik vergleichen: Ein Analyserahmen für die Beeinflussung von Arbeit durch Technik. *Arbeits- und Industriosiologische Studien* 11, Nr. 2: 124–42.
- Biselli, Tom; Steinbrink, Enno; Herbert, Franziska; Schmidbauer-Wolf, Gina M. und Reuter, Christian (2022): On the Challenges of Developing a Concise Questionnaire to Identify Privacy Personas, *Proceedings on Privacy Enhancing Technologies*, 2022(4), S. 645–669. DOI: <https://doi.org/10.56553/popets-2022-0126>.
- Bjørlo, Lena V. (2024): Freedom from interference: Decisional privacy as a dimension of consumer privacy online, *AMS Review*, 14(1–2), S. 12–36. DOI: <https://doi.org/10.1007/s13162-024-00273-x>.
- Borgesius, Frederik J. Zuiderveen; Kruikemeier, Sanne; Boerman, Sophie C. und Helberger, Natali (2017): Tracking Walls, Take-It-Or-Leave-It Choices, the GDPR, and the ePrivacy Regulation, *European Data Protection Law Review*, 3(3), S. 353–368. DOI: <https://doi.org/10.21552/edpl/2017/3/9>.
- Bybee, Rodger W. (1997): Achieving scientific literacy: From purposes to practices, Heinemann.
- Cho, Hichang (2022): Privacy helplessness on social media: its constituents, antecedents and consequences, *Internet Research*, 32(1), S. 150–171. DOI: <https://doi.org/10.1108/INTR-05-2020-0269>.
- Christl, Wolfie (2017): Corporate Surveillance in Everyday Life – How Companies Collect, Combine, Analyze, Trade, and Use Personal Data on Billions, Wien: Cracked Labs.
- Degeling, Martin (2016): Online Profiling – Analyse und Intervention zum Schutz von Privatheit, Dissertation, Fakultät für Wirtschaftswissenschaften, Universität Duisburg Essen.
- Dinev, Tamara und Hart, Paul (2004): Internet privacy concerns and their antecedents - measurement validity and a regression model, *Behaviour & Information Technology*, 23(6), S. 413–422. DOI: <https://doi.org/10.1080/01449290410001715723>.
- Dinev, Tamara und Hart, Paul (2006): An extended privacy calculus model for e-commerce transactions. *Information Systems Research*, 17(1), S. 61–80. DOI: <https://doi.org/10.1287/isre.1060.0080>
- EDPB (2020): Guidelines 4/2019 on Article 25 Data Protection by Design and by Default Version 2.0, adopted on 20 October 2020.
- Engel, Juliane und Schreiber, Jakob (2024): Homo Postdigitalis: Zur Transformation politischer Subjektivierung und Praxis in einer digital geprägten Welt, *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 63, S. 129–149. DOI: <https://doi.org/10.21240/mpaed/63/2024.09.18.X>.
- Esmailzadeh, Pouyan (2019): The Impacts of the Perceived Transparency of Privacy Policies and Trust in Providers for Building Trust in Health Information Exchange: Empirical Study, *JMIR Medical Informatics*, 7(4), S. e14050. DOI: <https://doi.org/10.2196/14050>.

- European Commission. Joint Research Centre. (2022): *DigComp 2.2, The Digital Competence framework for citizens: with new examples of knowledge, skills and attitudes*. Luxembourg: Publications Office. URL: <https://data.europa.eu/doi/10.2760/115376> (besucht am: 29.01.2026).
- von Grafenstein, Maximilian; Kiefaber, Isabel; Heumüller, Julie; Rupp, Valentin; Grassl, Paul; Kolless, Otto und Puszt, Zsófia (2023): *Privacy Icons as a Component of Effective Transparency and Controls Under the GDPR effective Data Protection by Design Based on Art. 25 GDPR*, preprint. SSRN. DOI: <https://doi.org/10.2139/ssrn.4559098>.
- von Grafenstein, Maximilian; Heumüller, Julie und Jakobi, Timo (2021): Die Gestaltung wirksamer Bildsymbole für Verarbeitungszwecke und ihre Folgen für Betroffene Mithilfe einer interdisziplinären Forschungsmethodologie, in *Alexander Boden, Timo Jakobi, Gunnar Stevens, Christian Bala (Hrsg.): Verbraucherdatenschutz-Technik und Regulation zur Unterstützung des Individuums*, S. 1–20. URL: https://pub.h-brs.de/files/6026/07_von_Grafenstein.pdf (besucht am: 10.11.2023).
- Grassl, Paul; Gerber, Nina und Von Grafenstein, Maximilian (2024): Practitioners' Corner · How Effectively Do Consent Notices Inform Users About the Risks to Their Fundamental Rights?, *European Data Protection Law Review*, 10(1), S. 96–104. DOI: <https://doi.org/10.21552/edpl/2024/1/14>.
- Hong, Weiyin und Thong, James Y.L. (2013): Internet Privacy Concerns: An Integrated Conceptualization and Four Empirical Studies, *MIS Quarterly*, 37(1), S. 275–298. DOI: <https://doi.org/10.25300/MISQ/2013/37.1.12>.
- Kahneman, Daniel (2011): *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kokolakis, Spyros (2017): Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon, *Computers & Security*, 64, S. 122–134. DOI: <https://doi.org/10.1016/j.cose.2015.07.002>.
- Klieme, Eckhard; Avenarius, Hermann; Blum, Werner; Döbrich, Peter; Gruber, Hans; Prenzel, Manfred; Reiss, Kristina; Riquarts, Kurt; Rost, Jürgen; Tenorth, Heinz-Elmar; Vollmer, Helmut J. (2003): *Zur Entwicklung nationaler Bildungsstandards: Eine Expertise*, Bundesministerium für Bildung und Forschung (BMBF).
- Kyi, Lin; Mhaidli, Abraham; Santos, Cristiana; Roesner, Franziska und Biega, Asia (2024): "It doesn't tell me anything about how my data is used": User Perceptions of Data Collection Purposes, In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, S. 1–12. DOI: <https://doi.org/10.1145/3613904.3642260>.
- Leimstädtner, David; Sörries, Peter und Müller-Birn, Claudia (2023): Investigating Responsible Nudge Design for Informed Decision-Making Enabling Transparent and Reflective Decision-Making, In: *Mensch und Computer 2023*. Rapperswil Switzerland: ACM, S. 220–236. DOI: <https://doi.org/10.1145/3603555.3603567>.
- Malhotra, Naresh K.; Kim, Sung S. und Agarwal, James (2004): Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model, *Information Systems Research*, 15(4), S. 336–355. DOI: <https://doi.org/10.1287/isre.1040.0032>.
- Masur, Philipp K.; Teutsch, Doris und Trepte, Sabine (2017): Entwicklung und Validierung der Online-Privatheitskompetenzskala (OPLIS), *Diagnostica*, 63(4), S. 256–268. DOI: <https://doi.org/10.1026/0012-1924/a000179>.

- Meier, Yannic und Krämer, Nicole C. (2024): A longitudinal examination of Internet users' privacy protection behaviors in relation to their perceived collective value of privacy and individual privacy concerns, *New Media & Society*, 26(10), S. 5942–5961. DOI: <https://doi.org/10.1177/14614448221142799>.
- Meng, Xiaoxiao (2024): Conceptualizing and measuring privacy boundary turbulence in technological contexts: Constructing a measurement scale, *Information Processing & Management*, 61(3), 103658. DOI: <https://doi.org/10.1016/j.ipm.2024.103658>.
- Papaioannou, Thanos; Kokotos, Spyridon E. und Tsohou, Aggeliki (2025): Educational escape rooms for raising information privacy competences: An empirical validation, *International Journal of Information Security*, 24(3), S. 120. DOI: <https://doi.org/10.1007/s10207-025-01037-9>.
- Parsons, Kathryn; Calic, Dragana; Pattinson, Malcolm; Butavicius, Marcus; McCormac, Agata und Zwaans, Tara (2017): The Human Aspects of Information Security Questionnaire (HAIS-Q): Two further validation studies, *Computers & Security*, 66, S. 40–51. DOI: <https://doi.org/10.1016/j.cose.2017.01.004>.
- Pavone, Vincenzo und Esposti, Sara D. (2012): Public assessment of new surveillance-oriented security technologies: Beyond the trade-off between privacy and security, *Public Understanding of Science*, 21(5), S. 556–572. DOI: <https://doi.org/10.1177/0963662510376886>.
- Pohle, Jörg (2022): Datenschutz: Rechtsstaatsmodell oder neoliberale Responsibilisierung. *Warum Datentreuhänder kein Mittel zum Schutz der Grundrechte sind.*, Vortrag 5 der Reihe „Zu treuen Händen“. Düsseldorf Verbraucherzentrale Nordrhein-Westfalen e.V. URL: https://www.verbraucherforschung.nrw/sites/default/files/2022-02/zth-05-pohle-datenschutz-rechtsstaatsmodell-oder-neoliberale-responsibilisierung_0.pdf (besucht am: 01.05.2026).
- Prince, Christine; Omrani, Nessrine und Schiavone, Francesco (2024): Online privacy literacy and users' information privacy empowerment: the case of GDPR in Europe, *Information Technology & People*, 37(8), S. 1–24. DOI: <https://doi.org/10.1108/ITP-05-2023-0467>.
- Quinn, Kelly und Epstein, Dmitry (2023): Dimensionalizing privacy to advance the study of digital disempowerment, *Big Data & Society*, 10(2), S. 20539517231221739. DOI: <https://doi.org/10.1177/20539517231221739>.
- Roessler, Beate (2005): *The Value of Privacy*, Cambridge: Polity Press.
- Roßnagel, Alexander (2007): *Datenschutz in einem informatisierten Alltag. Gutachten im Auftrag der Friedrich-Ebert-Stiftung*, Berlin: Friedrich-Ebert-Stiftung.
- Stapley, Emily; O'Keeffe, Sally und Midgley, Nick (2022): Developing Typologies in Qualitative Research: The Use of Ideal-type Analysis, *International Journal of Qualitative Methods*, 21, S. 16094069221100633. DOI: <https://doi.org/10.1177/16094069221100633>.
- Sun, Rui; Zhu, Qihua; Cheng, Ru X.; Tang, Wenlong; Zuo, Jiajia; Lv, Dong und Qin, Shukun (2024): Research on the cognitive neural mechanism of privacy empowerment illusion cues regarding comprehensibility and interpretability for privacy disclosures, *Scientific Reports*, 14(1), S. 8690. DOI: <https://doi.org/10.1038/s41598-024-58917-8>.

- Utz, Christine; Degeling, Martin; Fahl, Sascha; Schaub, Florian und Holz, Thorsten (2019): (Un)informed Consent: Studying GDPR Consent Notices in the Field, In: *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS '19)*. New York, NY: ACM, 973–990. DOI: <https://doi.org/10.1145/3319535.3354212>.
- Weinert, Franz E. (2001): Concept of competence: A conceptual clarification, In: D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* S. 45–65, Hogrefe & Huber.

DSGVO vs. Online-Dienste: Realität der Datenauskunft

Daniela Pöhn

Zusammenfassung

Seit dem Inkrafttreten der europäischen Datenschutz-Grundverordnung stellt das Auskunftsrecht nach Art. 15 ein zentrales Instrument für Nutzertransparenz dar. Bisher publizierte Studien belegen jedoch eine Diskrepanz zwischen dem rechtlichen Anspruch und der praktischen Umsetzung durch Online-Dienste. Nachdem dies bei großen Online-Diensten durch verwandte Studien gezeigt wurde, analysiert dieser Artikel die Umsetzung bei (1) kleineren Diensten, die vorrangig in Deutschland agieren, und (2) Diensten, die mit sensiblen Daten nach Art. 9 oder anderen, sensiblen Daten arbeiten, basierend auf drei Studien. Es zeigt sich, dass auch hier eine Diskrepanz zwischen rechtlichem Anspruch und praktischer Umsetzung besteht, was insbesondere bei sensiblen Daten kritisch zu bewerten ist. Die Ergebnisse werden hinsichtlich der Implikationen für die Nutzenden und neuer Möglichkeiten durch die künstliche Intelligenz und die aktuellen politischen Entwicklungen diskutiert.

1. Einleitung

Seit dem Inkrafttreten der europäischen Datenschutz-Grundverordnung (DSGVO) im Mai 2018 ist das Auskunftsrecht nach Art. 15 eines der neuen Rechte für Endnutzende, auch betroffene Personen genannt. Dieses Recht stärkt die Position der Nutzenden gegenüber Unternehmen und anderen Organisationen, die personenbezogene Daten verarbeiten. Es ermöglicht betroffenen Personen, Transparenz über die Speicherung und Verarbeitung ihrer personenbezogenen Daten bei Online-Diensten, sogenannten Controllern, zu verlangen. Darüber hinaus dient das Auskunftsrecht als Grundlage, um die Rechtmäßigkeit der Datenverarbeitung zu überprüfen und weitere Datenschutzrechte, wie etwa das Recht auf Berichtigung oder Löschung, geltend zu machen.

In der Literatur wurden bisher Anfragen nach Art. 15 DSGVO nur vereinzelt untersucht, beispielsweise wie Online-Dienste tatsächlich auf

Auskunftsanfragen, auch *Data Subject Access Requests* (DSARs) genannt, reagieren, welche Formate sie verwenden, wie verständlich ihre Antworten sind und wo sie scheitern. Diese Studien beschränken sich meist auf wenige große Online-Dienste (Pöhn/Gruschka 2025) oder bestimmte Aspekte, wie *Dark Patterns* (Löbel u. a. 2024) oder normierte Auskunftsmöglichkeiten (Leschke u. a. 2023). Pöhn u. a. (2023) zeigten, dass kleinere Online-Dienste häufig schlechter abschnitten, jedoch wurden nur wenige ausgewählte davon untersucht. Daher zeigt die Kurzzusammenfassung der Literatur, dass kleinere Dienste bisher kaum untersucht wurden. Von besonderem Interesse sind zudem Dienste, die mit sensiblen Daten arbeiten, die durch Art. 9 DSGVO besonders geschützt sind, da hier ein erhöhtes Maß an Transparenz und Kontrolle verlangt wird. Selbst wenn Nutzende ihre Daten freigeben, ist die Einhaltung der DSGVO elementar.

Daher werden hier (1) Online-Dienste, die sensible Daten speichern, und (2) kleinere, eher nationale oder regionale Online-Dienste analysiert. Dabei soll bspw. überprüft werden, ob wie bei Pöhn u. a. (2023) Online-Dienste aus Deutschland vorrangig per E-Mail-Anfragen annehmen und beantworten. Basierend auf den Analysen werden (3) die Ergebnisse bspw. hinsichtlich der Implikationen für die Nutzenden und neuer Möglichkeiten durch die künstliche Intelligenz diskutiert. Damit zeigt diese Veröffentlichung ein breiteres Bild bei der Umsetzung von Art. 15 DSGVO und diskutiert die Implikationen der gefundenen Ergebnisse.

Nachfolgend wird in Abschn. 2 der Stand der Forschung zusammengefasst, bevor in Abschn. 3 die grundlegende Methodik für die Erstellung der Online-Konten und die Untersuchung der Anfragen, Antworten und Daten beschrieben wird. In Abschn. 4 werden die Untersuchungen der Dienste, unterteilt in kleinere Dienste und Dienste mit sensiblen Daten, dargelegt und jeweils ein Zwischenfazit gezogen. Die Ergebnisse werden anschließend in Abschn. 5 diskutiert. Abschn. 6 schließt den Artikel mit einem Fazit und Ausblick.

2. Stand der Forschung

Während viele Veröffentlichungen zu Cookie-Bannern existieren, werden DSARs bisher verhältnismäßig selten betrachtet. Die meisten Veröffentlichungen betrachten die Nutzerperspektive, Dark Patterns, eine Methodik zur Veröffentlichung der Antworten und die DSARs bei ausgewählten Online-Diensten.

Petelka u. a. (2022) analysierten 38 DSARs von elf unterschiedlichen Organisationen durch Teilnehmer aus Indien, dem Vereinigten Königreich und den USA. Die Autoren betonen, dass die Teilnehmenden häufig Workarounds nutzen mussten, um an die eigenen Daten zu gelangen. Ebenfalls aus der Nutzerperspektive betrachteten Borem u. a. (2024) DSARs bei *Amazon*, *Facebook*, *Spotify* oder *Google*. Die Teilnehmenden zeigten sich enttäuscht, überrascht und verwirrt über die Antworten. Die Fragen der Teilnehmenden wurden selten durch die erhaltenen Daten beantwortet. Dies verdeutlicht ein Missverhältnis zwischen den Erwartungen der Nutzenden und der Realität von DSARs.

Pöhn u. a. (2023) untersuchten 27 Online-Dienste, wie *Google*, *Amazon* und *TikTok*, die überwiegend weltweit agieren. Zum Vergleich wurden weitere Online-Dienste gewählt, die sich auf den deutschsprachigen Raum beschränken. Die Autoren folgerten, dass Dienste, bei denen die Anfrage nach Art. 15 DSGVO einzig per E-Mail gestellt werden könne, diese manuell bearbeiten würden. Zudem zeigten die Autoren erste Dark Patterns auf, die sie während der Studie bemerkten. Ein Beispiel hierfür ist das Verstecken der Anfragefunktionalität. Basierend auf den Ergebnissen analysierten Löbel u. a. (2024) systematisch Dark Patterns bei 166 Online-Diensten. Bei 113 der untersuchten Online-Dienste konnten die Autoren insgesamt 238 Dark Patterns beobachten. Dabei entfielen 68 auf „*Obstructing*“ (drei Varianten), 42 auf „*Overloading*“ (drei Varianten), 34 auf „*Left in the Dark*“ (zwei Varianten), 37 auf „*Stirring*“ (zwei Varianten), 32 auf „*Fickle*“ (vier Varianten) und 25 auf „*Skipping*“ (drei Varianten). Damit verwendeten die Online-Dienste im Durchschnitt 1,43 Dark Patterns. Diese beobachteten Dark Patterns erschweren bis verhindern DSARs der Nutzenden.

Leschke u. a. (2024) stellten eine Methodik zur Veröffentlichung von DSGVO-Auskunftsantworten durch die Verwendung synthetischer Konten vor. Bei der exemplarischen Umsetzung zeigte sich, dass die Qualität der Auskünfte nicht zufriedenstellend im Sinne von vollständig und verständlich ist. Carrière u. a. (2025) veröffentlichten ihre Erfahrungen mit Datenspenden durch ihren verbesserten methodischen Ansatz, in dem Daten von realen Nutzern pseudonymisiert gespendet werden. Das Projekt TES-TER (Böning/Laufs 2024), welches sich mit der Transparenz und der nutzerfreundlichen Darstellung von Daten aus der Auskunft über die gespeicherten Daten beschäftigt, präsentierte erste Ideen für ein normiertes Auskunftsrecht. Ähnliche Ideen verfolgten auch Hansen u. a. (2023) und Leschke u. a. (2023).

Die genannten Erkenntnisse bei der Analyse der erhaltenen Daten wurden bei Einzelfallanalysen an den Beispielen von X (früher *Twitter*; Pöhn/Gruschka 2024) und *Facebook, Microsoft, Google, Apple, Amazon, LinkedIn* und *WhatsApp* (Pöhn/Gruschka 2025) bestätigt. Während Antworten zu Beginn des Inkrafttretens der DSGVO meist spärlich ausfielen, wurden diese mit wachsendem regulatorischem Druck umfangreicher, allerdings manchmal auf Kosten der Verständlichkeit. Zudem stellten einzelne Online-Dienste erst bei späteren Auskünften Daten bereit, die sie bereits früher erhoben und bei zwischenzeitlichen Auskünften nicht angegeben hatten. Besonders hervorgehoben wird dieses Verhalten am Beispiel von *Amazon* (Pöhn/Gruschka 2025). Damit zeigen sich große Online-Dienste nicht vollständig konform mit der DSGVO. Zugleich können mehrere Datensätze tiefere Einblicke in Geschäftspraktiken liefern, wie am Beispiel von *Uber* (Binns u. a. 2025) zu sehen. Die Autoren stellten basierend auf den Daten beispielsweise dynamische Preise und geringere Sicherheit fest, die Lebenskosten mit *Uber*-Fahrten decken zu können.

Dieser Überblick verdeutlicht, dass sich die bisherigen Studien hinsichtlich Art. 15 DSGVO vorwiegend auf bestimmte Aspekte, wie Dark Patterns und die Nutzendenperspektive, fokussierten oder wenige große Online-Dienste, wie Amazon, Google, Facebook und Apple, untersuchten. KMUs und andere Organisationen müssen jedoch ebenfalls die DSGVO umsetzen. Folglich fehlt eine breiter angelegte Analyse. Zugleich wurden bisher keine Dienste analysiert, die besonders sensible Daten (nach Art. 9) mit erhöhtem Schutzbedarf speichern.

3. Methodik

In diesem Abschnitt wird die grundlegende Methodik zum Erstellen eines Kontos, zur Abfrage der gespeicherten Daten und zur Analyse der so erhaltenen Daten basierend auf der Methodik von Leschke u. a. (2024) vorgestellt. Die einzelnen Schritte sind in Abb. 1 verdeutlicht und werden nachfolgend beschrieben.

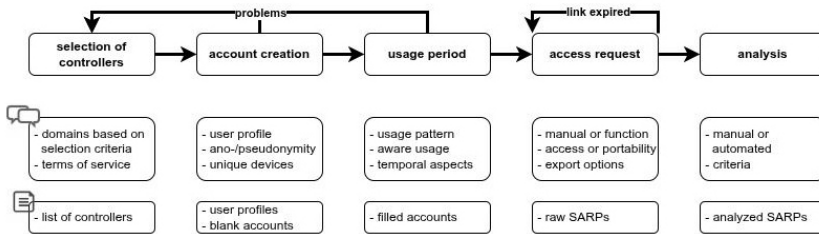


Abbildung 1: Die genutzte Methodik zur Analyse der Anfragen nach Art. 15 DSGVO (basierend auf Leschke u. a. 2024).

Die (1) Auswahl der Online-Dienste, auch *controllers* (Leschke u. a. 2024) genannt, erfolgt nach Regionalität, Größe bzw. der Kategorie der verarbeiteten sensiblen Daten. Um Interferenzen zu vermeiden, werden (2) neue Konten mit eigens dafür ausgewählten Endgeräten erstellt. Hierfür wird ein Nutzerprofil bzw. eine Persona passend zum Online-Dienst kreiert. Die (3) Nutzung erfolgt passend zum Ziel und, sofern möglich, ohne Einbeziehung Dritter. Beispielsweise wird für Sport-Apps ein Sportplan mit passendem Ziel entwickelt und die Nutzung der Chat-Funktion, die Unbeteiligte beeinträchtigen könnte, vermieden. Soweit die (4) DSARs über vorgefertigte Funktionen gestellt werden können, werden diese genutzt. Ansonsten wird eine vorgefertigte Blaupause für die Anfrage verwendet. Die Anfrage selbst wird genauso wie das Erhalten und Herunterladen der Daten nach dem Schema von Pöhn/Gruschka (2025) dokumentiert. Im Gegensatz zu Leschke u. a. (2024) ist das Ziel nicht die Veröffentlichung, sondern die (5) Analyse der Daten, die manuell erfolgt.

4. Analysen

Nachfolgend werden die Ergebnisse der drei durchgeführten Studien vorgestellt. Die Studien wurden zwischen Oktober 2024 und August 2025 durchgeführt. Folglich stehen die analysierten Online-Dienste und Daten für den Zeitpunkt, an dem die jeweilige Studie durchgeführt wurde.

4.1 DSARs von Diensten im deutschsprachigen Raum

Nachdem Pöhn u. a. (2023) neben weltweit agierenden auch wenige kleinere, auf deutschsprachige Nutzende ausgerichtete Online-Dienste unter-

suchten, deren Anfrageoptionen auf manuelle Bearbeitung der Anfragen hindeuten, sollten solche kleineren Online-Dienste im deutschsprachigen Raum genauer analysiert werden. Hierfür wurden 16 Online-Dienste (vgl. Tab. 1 im Anhang) ausgewählt, die entweder häufig genutzt werden oder Nischen bedienen. Die Studie wurde von März bis Mai 2025 durchgeführt.

Von diesen Online-Diensten haben vier (25 %; *Heise*, *MIFCOM*, *Kochbar* und *extraETF*) keine Antwort geliefert. Drei (18,75 %; *Hardwaredealz*, *NOZ* und *Merkur*) Online-Dienste meinten, aufgrund der fehlenden Speicherung von Daten – auch wenn ein Online-Konto erstellt wurde – keine Daten bereitstellen zu müssen. Von diesen Online-Diensten wurden zwei Benutzerkonten (12,5 %; *Hardwaredealz* und *NOZ*) nach der Anfrage ohne Erhalt von Daten gelöscht. Die verbliebenen neun (56,25 %) Online-Dienste antworteten vermutlich mit manuell zusammengestellten Daten, basierend auf Art der Anfrage und Zeit bis zur Antwort. Hierbei waren drei Antworten (18,75 %; *Nordbayern*, *Lecker* und *Süddeutsche*) entsprechend der Protokollierung der Eingaben unvollständig; zwei möglicherweise widersprüchlich in Hinblick auf die Datenschutzerklärung (12,5 %; *Finanztipp* und *Süddeutsche*); und eine (6,25 %; *Netzkino*) widersprüchlich in Hinblick auf die gesendeten Daten. Folglich sind während der Studie nur vier Antworten (25 %; *Gamestar*, *Moviepilot*, *Biallo* und *detektor.fm*) vollständig gewesen. Damit schnitt die Kategorie Unterhaltung am besten ab.

Die durchschnittliche Antwortzeit betrug 459 Stunden und 8 Minuten, also knapp über 19 Tage, was auf eine manuelle Bearbeitung hindeutet. Im gesamten Prozess wurden mehrere Dark Patterns nach Löbel u. a. (2024) und Pöhn u. a. (2023) identifiziert. Das Dark Pattern „*Longer than necessary*“ (Löbel u. a. 2024) beschreibt den verhältnismäßig langen Weg zur Anfrage, der beispielsweise bei *Süddeutsche* und *Merkur* protokolliert wurde. Teils wurden die Informationen auch ungeeignet dargestellt, was das Stellen der Anfrage zusätzlich erschwerte. Dies wird auch „*Hidden in plain sight*“ (Löbel u. a. 2024) bzw. „*Hiding GDPR request information*“ und „*Hiding data from the user*“ (Pöhn u. a. 2023) genannt. „*Misleading action*“ (Löbel u. a. 2024) bezeichnet die Irreführung durch vage oder unklare Auswahlmöglichkeiten oder falsche Reiter, was beispielsweise bei *Lecker* dazu führte, den Reiter Datenschutzmanager ohne Erfolg zu durchsuchen. „*Conflicting information*“ (Löbel u. a. 2024) im Sinn von „*Left in the dark*“ (Löbel u. a. 2024) beschreibt widersprüchliche Darstellungen, welche beispielsweise beim Kopieren von E-Mail-Adressen verwirren, wenn sich die auf der Website angezeigte kopierbare E-Mail-Adresse und die verlinkte und damit durch den Nutzer dann kopierte und neu eingefügte

E-Mail-Adresse unterscheiden. Mehrfach wurde auch das Pattern „*Dead end*“ (Löbel u. a. 2024) identifiziert, wenn die Anfrage an einer Stelle ein plötzliches Ende fand. Ferner wurden mehr Angaben (bspw. eine zusätzliche postalische Adresse) bei der Anfrage als „*Continuous prompting*“ (Löbel u. a. 2024) bzw. „*Requiring too much/no proof of identification*“ (Pöhn u. a. 2023) gewertet. Insgesamt nutzte jeder der untersuchten Online-Dienste mindestens ein Dark Pattern oder eine andere erkennbare strukturelle Barriere. Damit wurden mehr Dark Patterns als bei Löbel u. a. (2024) gefunden, die insbesondere größere Online-Dienste analysierten. Bei der Art der Dark Patterns zeigt sich, dass größere Dienste häufig überladen sind, während es bei kleineren eher grundlegend an Zugänglichkeit und Transparenz mangelt.

Die Ergebnisse werfen Fragen zur praktischen Wirksamkeit des Auskunftsrechts auf und verdeutlichen strukturelle Schwächen bei der Durchsetzung der Betroffenenrechte. Betroffene müssen mit Hürden bei der Wahrnehmung ihrer Rechte rechnen. Zugleich fehlt die Transparenz, ob und in welchem Umfang personenbezogene Daten von Online-Diensten verarbeitet werden, nachdem die meisten Dienste nur unvollständige Auskünfte geben.

4.2 DSARs von Diensten mit sensiblen Daten in vier Ländern

Der Schutz personenbezogener Daten gewinnt im digitalen Zeitalter insbesondere bei Online-Diensten, die sensible Daten wie allgemein Gesundheits- und im Speziellen Menstruationsdaten verarbeiten, zunehmend an Bedeutung. Dies wird auch durch Art. 9 DSGVO deutlich. Um einen ersten Überblick über den Umgang mit diesen Daten in unterschiedlichen Ländern zu erhalten, wurden landesspezifische Datenschutzgesetze aus der EU, Großbritannien, Kalifornien sowie Australien verglichen und analysiert, wie ausgewählte Online-Dienste diese Regelungen umsetzen. Hierfür wurden die Datenschutzerklärungen der Online-Dienste *Clue* (DE), *My Possible Self* (UK), *MyFitnessPal* (US) und *FatSecret* (AUS) (vgl. Tab. 2 im Anhang) untersucht und bewertet. In der praktischen Studie wurden virtuelle Maschinen eingerichtet, um die Nutzer-Standorte Deutschland, Großbritannien, Kalifornien und Australien zu simulieren und anschließend die vier Online-Dienste zu nutzen, was zu 16 Testfällen führte. Nach Anfrage und Erhalt der Daten wurden Kriterien wie die Zugänglichkeit der Anfrage

und die Datenqualität und -minimierung untersucht. Diese Studie wurde im Zeitraum von Oktober bis Dezember 2024 durchgeführt.

Zunächst wird in diesem Abschnitt auf die DSARs und die dadurch erhaltenen Daten eingegangen. Die Ergebnisse der Analyse der Datenschutzerklärungen werden im Zwischenfazit aufgegriffen.

Die App *Clue* (DE) ermöglichte eine einfache Anfrage in allen Ländern. Die Antwortzeiten variierten zwischen den Ländern von 14 (Deutschland) bis 21 (Australien) Tagen. Daten wurden im JSON-Format bereitgestellt, was für Nutzende ohne technische Kenntnisse nicht einfach verständlich ist. Die erhaltenen Daten stimmen soweit mit den Datenschutzerklärungen überein, jedoch werden laut eigenen Angaben in Deutschland mehr Nutzungsdaten (bspw. Google-Anzeigen-ID und Anzahl der Anzeigen-IDs) für die technische Verbesserung der App verwendet.

Im Vergleich zu den anderen bei dieser Studie untersuchten Online-Diensten, gestaltete sich der Anfrageprozess bei *MyFitnessPal* (US) am kompliziertesten. Nach einer ersten Antwort folgt mit zeitlichem Abstand die finale Bereitstellung der Daten. Bevor die gespeicherten Daten heruntergeladen werden konnten, musste sich der Antragsteller per E-Mail und Verifikationscode authentisieren. Die Daten selbst werden in einer passwortgeschützten Excel-Datei bereitgestellt. Die Analyse der Daten zeigt erhebliche Unterschiede in Umfang und Qualität zwischen den untersuchten Ländern: 3.323 Zeilen (Deutschland), 166 Zeilen (Großbritannien), 475 Zeilen (USA) und 753 Zeilen (Australien) bei vorher festgelegter Nutzung und identischen Datenschutzeinstellungen. Bei einem Blick in die Datensätze konnten mehrere Leerzeilen in den Datensätzen für die Länder USA und Australien entdeckt werden. Der deutlichste Unterschied ist beim Datensatz für die deutsche Nutzung feststellbar: Lediglich 358 der 3.323 Zeilen beziehen sich auf das eigene Benutzerkonto. Die übrigen Zeilen enthalten personenbezogene Daten zweier anderer, realer Nutzender¹, inklusive Gewicht, Körpergröße und BMI. Bei der weiteren Analyse fällt auf, dass nicht alle Daten bei den Benutzerkonten für Großbritannien, die USA und Australien korrekt und vollständig sind. Für viele eingegebene Werte wird beispielsweise NULL zurückgeliefert.

Die Anfrage bei *My Possible Self* (UK) muss per E-Mail gestellt werden, wobei die zunächst gefundene E-Mail-Adresse zu einer Fehlermeldung führte. Die an eine weitere E-Mail-Adresse gestellten Anfragen wurden

1 Wir haben im Nachgang an die Studie die Datenschutzbeauftragten informiert, jedoch keine Rückmeldungen erhalten.

in keinem der Fälle beantwortet. Ein ähnliches Ergebnis liefert *FatSecret* (AUS), welches keine der Anfragen beantwortet.

Das Ergebnis der Datenanfragen liefert differenzierte Ergebnisse. Lediglich ein Online-Dienst (*Clue* aus DE) antwortete auf alle vier DSARs innerhalb der gesetzlichen Frist und stellte die Daten bereit. Zwei der vier Online-Dienste (UK, AUS) antworteten nicht. Die Form, in der die Daten bereitgestellt wurden, unterschied sich je nach Dienst. Ein Dienst (*MyFitnessPal* aus US) übermittelte zudem Daten von weiteren Nutzenden, was ein meldepflichtiger Datenschutzvorfall ist. Nachdem die Daten sich je nach Land unterschieden, weist dies auf eine manuelle Bearbeitung hin. Das Ergebnis kann folglich bei weiteren Studien möglicherweise nicht reproduziert werden.

4.3 DSARs von Diensten in mehreren Kategorien mit sensiblen Daten

Um eine größere Anzahl an unterschiedlichen Diensten mit sensiblen Daten nach Art. 9 DSGVO oder anderen, eher sensiblen Daten zu analysieren, wurden mehrere Online-Dienste in den Kategorien Menstruationsapps, Fitness- und Gesundheitsapps, Schlaf-Tracking-Apps, Apps zu Religion und Spiritualität, LGBTQ+-Selbsthilfe-Apps und Dating-Apps (vgl. Tab. 3 im Anhang) ausgewählt. Für die Nutzung wurden eine weibliche und eine männliche Persona erstellt sowie passende Aktivitäten bzw. Angaben vorab überlegt. So stellt die Nutzung der Menstruationsapps einen simulierten Zyklus dar. Anschließend wurden die Daten angefordert. Nachdem viele Online-Dienste nicht oder nicht ausreichend antworteten, wurde auf die Auswertung der Daten verzichtet. Diese Studie wurde im Zeitraum von März bis August 2025 durchgeführt.

Bei Menstruations-Apps antworteten mehrere Anbieter (*MyCalendar*, *LadyCycle* und *Uvuvue*), dass keinerlei personenbezogene Daten zentral gespeichert würden. Ein Export der lokal gespeicherten Daten war nicht in jedem Fall möglich. Zwei weitere Anbieter (*Flo* und *Clue*) versendeten eine passwortgeschützte ZIP-Datei, wobei die Schritte zur Anfrage nicht selbstklärend (elf Teilschritte) bzw. die Bereitstellung der Daten in manchen Aspekten nicht datenschutzfreundlich (Nutzung eines weiteren Dienstes) erscheinen.

Dating-Apps sperrten vermehrt das Konto, teils bereits bei der Anmeldung (*Bumble* und *Plenty of Fish*) und teils nach Anfrage der Daten nach Art. 15 DSGVO (*Hinge* und *OkCupid*). *Hinge* gestaltete den Prozess der

Anfrage und Verifikation selbst transparent, auch wenn keine Daten geschickt und das Konto gesperrt wurden. *Tinder* stellte als Ausnahme die angeforderten Daten nach mehreren, zeitlich versetzten Verifikationsschritten bereit.

Zwei der getesteten Schlaf-Tracking-Apps (*Sleep* und *Sleep Monitor*) speichern laut Aussage der Anbieter personenbezogene Daten nur lokal. Eine Antwort (*Calm*) blieb aus, während zwei Apps (*Sleep Cycle* und *Samsung Health*) eine Möglichkeit zum Export der Daten bereitstellten.

Bei den Apps zu Religion und Spiritualität konnte eine Anfrage (*Evermore*) nicht zeitnah beantwortet werden, da die für den Datenschutz zuständige Person im Urlaub war. Ein weiterer Anbieter (*Abide*) verstand die Anfrage zunächst als Wunsch auf Löschung. Die anschließend geschickten Screenshots genügten ebenfalls nicht den Anforderungen. Ähnlich unzureichend war die Antwort des Anbieters einer weiteren App (*Glauben leben*), für den es unklar war, welche Daten im Rahmen der Anfrage gemeint waren. Bei zwei Apps (*Bible* und *Cogito*) konnten Datenanfragen gestellt und schließlich Daten zur Verfügung gestellt werden.

Bei LGBTQ+-Selbsthilfe-Apps blieben in den allermeisten Fällen (*PrideMantra*, *Kalda* und *Lex*) die Anfragen unbeantwortet. Während eine Anfrage über einen integrierten Support-Button bei *Voda* unbeantwortet blieb, wurde eine E-Mail zeitnah beantwortet und die Daten konnten heruntergeladen werden. Bei *Taimi* sollte für die Anfrage zur Authentifizierung zusätzlich ein Selfie in einer bestimmten Pose gemacht werden, auch wenn bei der Anmeldung kein Bild erforderlich war. Nach Äußerung der Bedenken hierbei wurden die angeforderten personenbezogenen Daten nach 24 Stunden bereitgestellt.

Fitness- und Gesundheits-Apps bieten ein gemischtes Bild. Mehrere Anbieter (*Feeletics*, *Jefit*, *Seven – 7 Minute Workout* und *Google Fit*) entsprachen der Anfrage ohne Probleme. Bei anderen war wiederum die Mithilfe des Supports notwendig (bspw. *Nike Training Club*), was jedoch nicht immer (*Adidas Running* und *Strava*) zum Erfolg führte. Mehrere Anbieter (*8fit* und *MyFitnessPal*) löschten nach Anfrage das Konto, während ein weiterer Online-Dienst (*FitOn*) nie antwortete.

Die Ergebnisse zeigen, dass Online-Dienste, die mit sensiblen Daten nach Art. 9 DSGVO oder anderen, eher sensiblen Daten arbeiten, nicht zwangsläufig DSGVO-konform agieren. Besonders kritisch ist die Sperrung oder Löschung mehrerer Konten nach Anfragen nach Art. 15 DSGVO zu betrachten. Dies betrifft auch *MyFitnessPal*, welcher in der Studie, beschrieben in Abschn. 4.2, bereits negativ durch die Zusendung von Daten Dritter

auffiel. Folglich konnte zwar das in der vorherigen Studie erzielte Ergebnis (Daten Dritter in erhaltenen Daten enthalten) nicht verifiziert werden, jedoch ist das neue Ergebnis nicht konformer mit der DSGVO. *Clue* wiederholte das gute Bild, auch wenn kleinere Aspekte verbessert werden könnten. Die laut mehreren Online-Diensten rein lokale Speicherung konnte im Rahmen der Studie nicht verifiziert werden und soll folglich zukünftig betrachtet werden.

5. Diskussion

In diesem Abschnitt werden die im vorangegangenen Abschnitt erzielten Ergebnisse unter Einbezug der Literatur diskutiert.

5.1 Diskussion der Studien und Ergebnisse

Zusammenfassend offenbaren die Studien eine Diskrepanz zwischen gesetzlichen Anforderungen und der praktischen Umsetzung des Auskunftsrechts bei Online-Diensten. Viele der analysierten Umsetzungen scheitern an der technischen Reife, der Ausgestaltung oder der Transparenz. Zugleich zeigen sich Unterschiede zwischen den Online-Diensten. Insbesondere eher kleinere, nationale Online-Dienste setzen auf Anfragen per E-Mail und mit entsprechend manueller, zeitaufwändiger Bearbeitung, die nicht unbedingt konform zur DSGVO sind. Aber auch Online-Dienste, die sensible Daten nach Art. 9 DSGVO oder andere, eher sensible Daten speichern, schneiden überraschend schlecht ab. Die beschriebenen Studien zeigen, dass nicht zwingend der rechtliche Rahmen das Problem ist, sondern die Art und Weise, wie Unternehmen ihn interpretieren und umsetzen. Ohne klare technische Standards und Frameworks, verbraucherfreundliche Gestaltung mit Blick auf den Nutzenden und zugleich Betroffenen und behördliche Kontrollen mit entsprechenden Konsequenzen bleibt das Auskunftsrecht ein Versprechen. Genau diese Hilfestellungen und Konsequenzen scheinen bei verhältnismäßig kleinen Diensten besonders wichtig zu sein. Folgestudien könnten ein breiteres Bild der DSAR-Realität zeigen und somit in Kombination mit Wissenschaftskommunikation auf bestehende Probleme aufmerksam machen.

Wenn Nutzende ihr Recht auf Information über verarbeitete persönliche Daten faktisch kaum wahrnehmen oder verstehen können, besteht die Ge-

fahr eines strukturellen Ungleichgewichts zugunsten datenverarbeitender Unternehmen. Insbesondere vulnerable Gruppen, bspw. ältere Menschen, Menschen mit Behinderung, Menschen mit geringerer digitaler Bildung oder finanziell schlechter gestellte Menschen, laufen Gefahr, systematisch ausgeschlossen zu werden, da ihnen der Zugang zu den Funktionen zur Durchsetzung ihrer Rechte, die sie vielleicht nicht einmal kennen, verwehrt bleibt. Entsprechend der digitalen Teilhabe sollte deshalb der Zugang für Nutzende zu den eigenen Daten vereinfacht werden.

Auch wenn stellenweise die Nutzung von Personas bei Online-Diensten gegen die Nutzungsrichtlinien verstößt und daher die Löschung von Konten bei der Nutzung möglich ist, können nationale Gesetze diese Regelungen außer Kraft setzen (Leschke u. a. 2024). Laut mehreren psychologischen Verbänden, wie dem Berufsverband Deutscher Psychologinnen und Psychologen und der Deutschen Gesellschaft für Psychologie, ist Täuschung nur dann akzeptabel, wenn beispielsweise ausschließlich dadurch ein wissenschaftlicher Mehrwert geschaffen werden kann und keine Alternative ohne Täuschung existiert. Bei den vorgestellten Studien konnten durch die Nutzung von Personas und dedizierten Geräten bzw. virtuellen Maschinen Interferenzen und Nachteile vermieden werden, die bspw. auftreten würden, wenn reale Konten verwendet würden. Bei Betrachtung der durchgeführten Studien wird zudem deutlich, dass noch mehr Online-Dienste und Kategorien, bspw. Apps für Kinder, betrachtet werden müssen, um einen umfassenden Überblick zu erhalten.

5.2 Chancen und Risiken von Sprachmodellen

Insbesondere in den Abschn. 4.1 und 4.2 wurden unvollständige oder falsche Datensätze bereits angesprochen. Dies deckt sich mit den Erkenntnissen anderer Studien (Pöhn/Gruschka 2024; 2025). Zugleich wurden bei diesen Studien teils Probleme bei der Datenminimierung festgestellt. Wenn große Sprachmodelle oder Empfehlungssysteme auf historischen, aber unvollständigen, intransparenten oder inkorrekten Daten beruhen, kann dies zu falschen Ergebnissen führen.

Außerdem können zu viele Daten ein umfassendes Profil eines gläsernen Menschen generieren, was wiederum für passende Produktplatzierungen oder auch Empfehlungen für Krankenkassen genutzt werden kann. Dies bezieht sich nicht nur auf die bei den Online-Diensten gespeicherten Daten, sondern auch auf die durch Werbe-Tracking und Datenhandel gesam-

melten Daten (Dachwitz/Meineck 2025). Durch die bereits existierenden Daten können Bewegungsprofile abgelesen werden, was sich zumindest teilweise auch durch die erhaltenen Daten bei DSARs nachvollziehen lässt (Pöhn/Gruschka 2024; 2025). Entsprechend der neuen Entwicklungen im Bereich der künstlichen Intelligenz und Sprachmodelle und der damit vereinfachten Auswertungsmöglichkeiten von großen Datenmengen ist die Einhaltung der Grundsätze für die Verarbeitung personenbezogener Daten von großer Bedeutung.

Zugleich können Sprachmodelle und erklärbare künstliche Intelligenz womöglich zukünftig helfen, die Daten den Nutzenden besser zu erklären und greifbarer zu machen. Die Erklärbarkeit war nicht der Fokus der hier vorgestellten Studien, jedoch waren viele Daten eher maschinenlesbar als verständlich. Zusätzliche Erklärungen, wie sie durch manche große Online-Dienste realisiert werden, konnten nicht gefunden werden. Die Nutzung von Sprachmodellen und erklärbarer künstlicher Intelligenz für die Erklärbarkeit und Verständlichkeit der angefragten Daten erfordert allerdings weitere Forschung, zumal neue Gefahren möglich sind, beispielsweise wenn Nutzende unreflektiert sensible Daten in Systeme einspeisen, die niedrige Datenschutzstandards aufweisen.

5.3 Schaden durch die Nutzung der Daten

Nachdem bei den Online-Diensten mit sensiblen personenbezogenen Daten kaum Transparenz besteht, kann der Nutzende nicht absehen, welcher Schaden potenziell angerichtet werden kann. Bedingt durch das Privacy Paradox (Gerber u. a. 2018) bzw. das Problem, dass Dienste im Internet bereits viele Daten haben und dadurch ein vorsichtigerer Umgang für weniger relevant gehalten wird als bspw. eine bequeme Nutzung (Teuschel u. a. 2023), können mehr Daten gespeichert und potenziell genutzt werden als noch vor wenigen Jahrzehnten. In Hinblick auf die Überwachung durch Tracking (Dachwitz/Meineck 2025) kann der potenzielle Schaden beispielsweise bei Menstruationsapps (Souto/Müller 2025) durch geänderte politische Situationen weitreichende Folgen haben. Letztlich könnten die so erhobenen und gesammelten Daten für Regierungen und Behörden genutzt werden, sofern eine solche Nutzung rechtlich erlaubt ist. Durch die globale Welt können Daten von deutschen Nutzerinnen in den USA gespeichert werden, was je nach politischer Lage Fragen aufwerfen kann.

Dies zeigt zugleich das Missbrauchspotential und wie schnell die Sicherheit der Nutzerinnen in Gefahr sein kann (Malki u. a. 2024).

6. Fazit

In diesem Artikel wurden DSARs für Dienste im deutschsprachigen Raum und Dienste mit sensiblen Daten gestellt und die Daten analysiert. Im deutschsprachigen Raum antworteten viele Online-Dienste nicht oder nicht korrekt auf die Anfragen. Die erhaltenen Daten waren häufig nicht vollständig. Um Dienste mit sensiblen Daten zu analysieren, wurden zwei Studien durchgeführt. Es zeigte sich, dass Online-Dienste mit sensiblen Daten diese besser schützen könnten. Auch bei der breiter angelegten Studie mit Online-Diensten aus mehreren Kategorien wurde deutlich, dass Verbesserungsbedarf bei Anfragen und beim Herunterladen der Daten besteht. Zudem wurden mehrere Konten nach dem Stellen der Anfrage entweder deaktiviert oder gelöscht. Die Studien verdeutlichen somit die Diskrepanz zwischen dem rechtlichen Anspruch auf Transparenz und den tatsächlichen Praktiken der Online-Dienste hinsichtlich des Auskunftsrechts. Die bestehenden, in Studien dokumentierten Praktiken im Umgang mit personenbezogenen Daten führen zu einem potenziellen Missbrauch der gesammelten Informationen. Zugleich verschärfen mangelnde Sensibilität und Kenntnis von bestehenden Rechten die Problematik. Folglich sind weitere Arbeiten notwendig, um eine möglichst einfache Teilhabe – im Kontext von Art. 15 DSGVO von der Anfrage bis zur Betrachtung der eigenen Daten – zu ermöglichen.

Literatur

- Binns, Reuben; Stein, Jake; Datta, Siddhartha; Van Kleek, Max und Shadbolt, Nigel (2025): Not Even Nice Work If You Can Get It; A Longitudinal Study of Uber's Algorithmic Pay and Pricing. In: *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAcT '25)*. New York: Association for Computing Machinery (ACM), S. 1484–1497. <https://doi.org/10.1145/3715275.3732099>
- Böning, Fabiola und Laufs, Uwe (2024): Betroffenenrechte in der digitalen Selbstvermessung. In: Michael Friedewald, Alexander Roßnagel, Christian L. Geminn, Murat Karaboga und Stephan Schindler (Hrsg.): *Data Sharing – Datenkapitalismus by Default? Privatheit und Selbstbestimmung in der digitalen Welt*. Baden-Baden: Nomos, S. 301–323 <https://doi.org/10.5771/9783748940173>

- Borem, Arthur; Pan, Ellen; Obielodan, Olufunmilola; Roubinowitz, Aurelia und Dovichi, Luca (2024): Data Subjects' Reactions to Exercising Their Right of Access. In: *33rd USENIX Security Symposium, Philadelphia, PA, USA, August 14–16, 2024*. <https://www.usenix.org/system/files/usenixsecurity24-borem.pdf>
- Carrière, Thijs C.; Boeschoten, Laura; Struminskaya, Bella; Janssen, Heleen L.; de Schipper, Niek C. und Araujo, Theo (2025): Best practices for studies using digital data donation. *Quality & Quantity - International Journal of Methodology*, 59 (Suppl 1), S. 389–412. <https://doi.org/10.1007/s11135-024-01983-x>
- Dachwitz, Ingo und Meineck, Sebastian (04. November 2025): Databroker Files - Das Wichtigste zur Spionage-Gefahr durch Handy-Standortdaten in der EU. <https://netzpolitik.org/2025/databroker-files-das-wichtigste-zur-spiegung-gefahr-durch-handy-standortdaten-in-der-eu/> (besucht am 23.01.2026).
- Gerber, Nina; Gerber, Paul und Volkamer, Melanie (2018): Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers & Security*, 77, S. 226–261. <https://doi.org/10.1016/j.cose.2018.04.002>
- Hansen, Malte; Gruschka, Nils und Jensen, Meiko (2023): Introducing the Concept of Data Subject Rights as a Service Under the GDPR. In: Steffen Schiffner, Sébastien Ziegler und Meiko Jensen (Hrsg.): *Privacy Symposium 2023. Data Protection Law International Convergence and Compliance with Innovative Technologies (DPLICIT)*. Cham: Springer, S. 17–31. https://doi.org/10.1007/978-3-031-44939-0_2
- Leschke, Nicola; Kirsten, Florian; Pallas, Frank und Grünewald, Elias (2023): Streamlining Personal Data Access Requests: From Obstructive Procedures to Automated Web Workflows. In: Irene Garrigós, Juan Manuel Murillo Rodríguez und Manuel Wimmer (Hrsg.): *Web Engineering. ICWE 2023. Lecture Notes in Computer Science*, Band 13893. Cham: Springer, S. 111–125. https://doi.org/10.1007/978-3-031-34444-2_9
- Leschke, Nicola; Pöhn, Daniela und Pallas, Frank (2024): How to Drill into Silos: Creating a Free-to-Use Dataset of Data Subject Access Packages. In: Meiko Jensen, Cédric Lauradoux und Kai Rannenberg (Hrsg.): *Privacy Technologies and Policy, 12th Annual Privacy Forum, APF 2024, Karlstad, Sweden, September 04–05, 2024. Proceedings*. Cham: Springer, S. 132–155. https://doi.org/10.1007/978-3-031-68024-3_7
- Löbel, Alexander; Schäfer, René; Püschel, Hanna; Güney, Esra und Meyer, Ulrike (2024): Access Your Data... if You Can: An Analysis of Dark Patterns Against the Right of Access on Popular Websites. In: Meiko Jensen, Cédric Lauradoux und Kai Rannenberg (Hrsg.): *Privacy Technologies and Policy, 12th Annual Privacy Forum, APF 2024, Karlstad, Sweden, September 04–05, 2024. Proceedings*. Cham: Springer, S. 23–47. https://doi.org/10.1007/978-3-031-68024-3_2
- Malki, Lisa Mekioussa; Kaleva, Ina; Patel, Dilisha; Warner, Mark und Abu-Salma, Ruba (2024): Exploring Privacy Practices of Female mHealth Apps in a Post-Roe World. In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. New York: Association for Computing Machinery (ACM), Artikel 576, S. 1–24. <https://doi.org/10.1145/3613904.3642521>

- Petelka, Justin; Oreglia, Elisa; Finn, Megan und Srinivasan, Janaki (2022): Generating Practices: Investigations into the Double Embedding of GDPR and Data Access Policies. In: *Proceedings of the ACM Human-Computer Interaction*. Band 6, Ausgabe CSCW2. New York: Association for Computing Machinery (ACM), Artikel 518, S. 1–26. <https://doi.org/10.1145/3555631>
- Pöhn, Daniela; Mörsdorf, Niklas und Hommel, Wolfgang (2023): Needle in the Haystack: Analyzing the Right of Access According to GDPR Article 15 Five Years after the Implementation. In: *Proceedings of the 18th International Conference on Availability, Reliability and Security, ARES '23*. New York: Association for Computing Machinery (ACM), Artikel 91, S. 1–10. <https://doi.org/10.1145/3600160.3605064>
- Pöhn, Daniela und Gruschka, Nils (2024): Past and Present: A Case Study of Twitter's Responses to GDPR Data Requests. In: Kai Rannenberg, Prokopios Drogkaris und Cédric Lauradoux (Hrsg.): *Privacy Technologies and Policy, 11th Annual Privacy Forum, APF 2023, Lyon, France, June 01–02, 2023. Proceedings*. Cham: Springer Nature Switzerland, S. 57–84. https://doi.org/10.1007/978-3-031-61089-9_4
- Pöhn, Daniela und Gruschka, Nils (2025): Qualitative In-Depth Analysis of GDPR Data Subject Access Requests and Responses from Major Online Services. In: *Proceedings of the 11th International Conference on Information Systems Security and Privacy - Volume 1*. ICISP. Setúbal: INSTICC, SciTePress, S. 149–156. <https://doi.org/10.5220/0013093000003899>
- Souto, Luana Mathias und Müller, Jörg (2025): Legal and Ethical Challenges of Femtech: Protecting Women's Data & Rights in the Age of Surveillance Capitalism. In: *Companion Proceedings of the ACM on Web Conference 2025 (WWW '25)*. New York: Association for Computing Machinery (ACM), S. 2796–2798. <https://doi.org/10.1145/3701716.3716881>
- Teuschel, Moritz; Pöhn, Daniela; Grabatin, Michael; Dietz, Felix; Hommel, Wolfgang und Alt, Florian (2023): 'Don't Annoy Me With Privacy Decisions!' — Designing Privacy-Preserving User Interfaces for SSI Wallets on Smartphones. *IEEE Access*, Ausgabe 11, S. 131814–131835. <https://doi.org/10.1109/ACCESS.2023.3334908>

Anhang

Tab. 1: Ausgewählte Dienste im deutschsprachigen Raum.

Kategorie	Online-Dienst-Name
Finanzen	Finanztip, extraETF, Biallo
Nachrichten	Nordbayern, Neue Osnabrücker Zeitung (NOZ), Süddeutsche, Merkur
Computer	Heise, Hardwaredealz, MIFCOM
Rezepte	Kochbar, Lecker
Unterhaltung	Netzkino, Moviepilot, Gamestar, detektor.fm

Tab. 2: Ausgewählte Dienste mit sensiblen Daten.

App-Name	Herkunftsland	Kategorie
Clue	Deutschland	Menstruations-App
My Possible Self	UK	Psychische Gesundheits-App
MyFitnessPal	USA	Tracking-App f. Ernährung und Fitness
FatSecret	Australien	Abnehm-App

Tab. 3: Ausgewählte Dienste in mehreren Kategorien mit sensiblen Daten.

Kategorie	App-Name
Menstruations-App	Flo, Clue, Lady Cycle, Uvuvview, Periodenkalender
Dating-Apps	Tinder, Bumble, Hinge, OkCupid, Plenty of Fish
Schlaf-Tracking-Apps	Sleep as Android, Sleepcycle, Sleep Monitor, Calm, Samsung Health
Apps zu Religion und Spiritualität	Glauben.leben, Abide, Cogito, Bible App, Evermore
LGBTQ+-Selbsthilfe-Apps	Voda, Pride Mantra, Kalda, Lex, Taimi
Fitness- und Gesundheits-Apps	Nike Training Club, Freeletics, Jefit, Seven, 8Fit, GoogleFit, Strava, MyFitnessPal, FitOn, Adidas Training by Runtastic

Privatheits-relevante Eigenschaften von IoT-Produkten für Verbraucher:innen*

Florian Dehling, Sebastian Koch, Luigi Lo Iacono, Lindrit Kqiku, Delphine Reinhardt und Harald Zwingelberg

Zusammenfassung

Das Internet der Dinge (IoT) hält zunehmend Einzug in private Haushalte und ermöglicht durch Sensorik und Konnektivität weitreichende Datenerhebungen in der privaten Sphäre. Für Verbraucher:innen ist vor dem Kauf jedoch häufig nur schwer nachvollziehbar, welche Daten ein IoT-Produkt – typischerweise bestehend aus Gerät, Begleit-App und Backend-Diensten – verarbeitet, an welche Empfänger Daten übertragen werden und welche Privatheitsrisiken daraus resultieren. Privacy Labels werden als Ansatz diskutiert, diese Informationslücke zu schließen; ihre Aussagekraft ist in der Praxis jedoch oftmals durch unvollständige, unspezifische oder herstellerabhängige Informationen eingeschränkt. Vor diesem Hintergrund identifizieren wir privatheits-relevante Eigenschaften von IoT-Produkten, die sich herstellerunabhängig erheben und vergleichen lassen. Auf Grundlage einer Literaturrecherche zu Privacy- und Security-Labels, relevanten rechtlichen Rahmenwerken sowie Studien zum privatheitsbezogenen Nutzerverhalten und ergänzt durch interdisziplinäre Expertenworkshops entwickeln und kuratieren wir eine Sammlung von 54 Eigenschaften. Die resultierende Merkmalsbasis unterstützt eine nachvollziehbare, unabhängige Beurteilung von IoT-Produkten und liefert eine Grundlage für die Ausgestaltung aussagekräftiger Privacy Labels, die Verbraucherberatung sowie eine regulatorische Einordnung.

* Dieser Beitrag ist im Rahmen des durch das Bundesministerium für Forschung, Bildung, Technologie und Raumfahrt geförderten Verbundprojekts „Unboxing.IoT.Privacy“ entstanden (Förderkennzeichen: 16KIS1931K).

1. Einleitung

Die fortschreitende Digitalisierung dringt zunehmend in zahlreiche Bereiche unseres Alltagslebens ein und etabliert das Internet der Dinge (engl. Internet of Things, IoT) als einen Bestandteil moderner Gesellschaften. IoT-Geräte umfassen unter anderem intelligente Haushaltsgeräte, die durch integrierte Sensoren und erweiterte Konnektivität unsere Lebensqualität verbessern können. Gleichzeitig schaffen diese Geräte jedoch die Möglichkeit für umfassende und potenziell invasive Datenerhebungen innerhalb der privaten Sphäre der eigenen vier Wände (Subahi/Theodorakopoulos 2019). Die gesammelten Daten lassen sich zu detaillierten Profilen verdichten, die tiefgehende Einblicke in die persönlichen Gewohnheiten und Vorlieben der Nutzenden gewähren. Forschungen belegen beispielsweise, dass durch die Analyse von Stromverbrauchsdaten präzise Informationen über die Art und Betriebsmodi von genutzten Haushaltsgeräten extrahiert werden können. Im Fall von Fernsehern lassen sich sogar Rückschlüsse auf die konsumierten Inhalte ziehen (Greveler u. a. 2012; Reinhardt u. a. 2015).

Angesichts der Vielzahl an IoT-Produkten und den zunehmenden Datenmengen, die durch vielfältige Sensoren in privaten Räumen erfasst werden, ist es entscheidend, Datenschutzaspekte vor dem Erwerb zu berücksichtigen, um das Grundrecht auf informationelle Selbstbestimmung fundiert ausüben zu können. Für Verbraucher:innen gestaltet es sich häufig herausfordernd, die Datenverarbeitungsprozesse, die mit IoT-Geräten verbunden sind, vor dem Kauf zu durchschauen und potenzielle Datenschutzrisiken zu identifizieren. Im Gegensatz zur Nutzung von Apps oder Onlinediensten besteht für IoT-Anbieter keine unmittelbare Verpflichtung, die Verbraucher:innen vor dem Kauf über die spezifische Datenverarbeitung zu informieren. Ein wesentlicher Grund hierfür ist u. a., dass die Art der Datenverarbeitung bei einigen IoT-Geräte-kategorien erst durch die tatsächliche Nutzung bestimmt wird. Somit ergibt sich der Anwendungsbereich der Datenschutz-Grundverordnung (DSGVO) erst im konkreten Nutzungskontext. Eine Überwachungskamera kann z. B. in sehr unterschiedlichen Anwendungsfällen betrieben werden. Wird sie ohne Anbindung an eine vom Hersteller bereitgestellte Cloud ausschließlich zur Überwachung des eigenen schlafenden Babys verwendet, fällt eine solche IoT-Nutzung oftmals noch unter die Haushaltsausnahme (Art. 2 Abs. 2 lit. c DSGVO) und damit aus dem Anwendungsbereich der DSGVO (Fuhrken 2026). Wird die Kamera indes mit einem Clouddienst des Herstellers verbunden, greifen die Bestimmungen der DSGVO. Wenn der Betrieb Einfluss auf die Privatheit von

Dritten nimmt, die sich beispielsweise als Besuchende im Aufnahmebereich aufhalten (Saqib u. a. 2025), können sich aus der DSGVO Pflichten für den Betreiber ergeben, unabhängig davon, ob die Kamera mit einer Cloud verbunden ist oder nicht.

Um die Informationslücke hinsichtlich der Privatheitseigenschaften von IoT-Geräten zu schließen, empfehlen Forschende den Einsatz von Privacy Labels (Emami-Naeini u. a. 2022; Shen/Vervier 2019). Diese Labels sollten bereits vor dem Kauf eines IoT-Produkts wichtige Informationen zielgruppenspezifisch bereitstellen. Als besonders nützlich für Verbraucher:innen haben sich Privacy Labels erwiesen, die spezifische Beschreibungen privatheits-relevanter Merkmale bieten, anstatt auf Bewertungsskalen oder binäre Siegel zu setzen (Johnson u. a. 2020). Solche Labels informieren beispielsweise über die im IoT-Produkt integrierten Sensoren, die Verarbeitung personenbezogener Daten und die potenzielle Weitergabe dieser Daten an Dritte (Railean/Reinhardt 2021).

Die Datenquellen für Privacy Labels können entweder bestehende Datenschutzerklärungen oder vom Hersteller bereitgestellte Informationen sein (Johnson u. a. 2020). Da Datenschutzerklärungen bei IoT-Produkten teils unvollständig sind, insbesondere wenn es um verkettbare, aber nicht eindeutig personenbezogene Daten geht, hängt die Zuverlässigkeit von Privacy Labels, die auf Herstellerinformationen basieren, stark von der Transparenzbereitschaft der Hersteller ab. Um unabhängigen Prüfstellen oder technikaffinen Einzelpersonen die Möglichkeit zu geben, die Privatheitsaspekte eines IoT-Produkts unabhängig zu untersuchen, müssen Eigenschaften identifiziert werden, die auch ohne aktive Unterstützung der Hersteller überprüfbar sind.

Basierend auf einer Literaturrecherche mit Fokus auf Privacy- und Security-Labels, rechtliche Rahmenwerke sowie Studien zu privatheits-relevantem Nutzerverhalten wurde eine initiale Sammlung für Verbraucher:innen relevanter Privatheits-Eigenschaften erarbeitet. Diese wurde anschließend durch interdisziplinären Expertenkonsens geprüft, sortiert und verfeinert. Die Expertengruppe umfasste zwei Universitätsprofessoren, einen Postdoktoranden, zwei Doktoranden im Bereich menschenzentrierte IT-Sicherheit und Datenschutz, einen Vertreter aus der Industrie sowie einen Rechtspraktiker. Die Kuratierung erfolgte im Rahmen eines initialen Workshops mit anschließenden mehrfachen Online-Diskussionsrunden und wurde fortgeführt, bis eine konzeptuelle Sättigung erreicht war, das heißt, keine neuen Kandidateneigenschaften mehr aufkamen und verbleibende Meinungsverschiedenheiten im Konsens gelöst werden konnten.

Für die Aufnahme in die finale Sammlung musste eine Eigenschaft drei Einschlusskriterien erfüllen: Sie musste (a) einem IoT-Produkt oder seinem zugehörigen Dienst zuordenbar sein, (b) operationalisiert werden können, um verbraucherorientierte Kaufentscheidungen zu unterstützen, und (c) hinreichend spezifisch sein, um grundsätzlich bewertet oder überprüft werden zu können. Ausgeschlossen wurden Kandidaten, die primär Richtlinien für die Gestaltung von Benutzeroberflächen oder Labels beschrieben (z. B. Formulierungen, barrierefreie Farbgebung), plattform- oder werkzeutechnische Anforderungen betrafen (z. B. Installationsaufwand, plattformübergreifende Bereitstellung) oder nicht hinreichend klar formuliert waren.

Daraus resultierte eine Sammlung von 54 für Verbraucher:innen relevanten Privatheits-Eigenschaften samt zugehöriger Metriken, die eine herstellerunabhängige Prüfung und Bewertung von IoT-Geräten ermöglicht. Dabei werden sowohl technische Fakten als auch die Gebrauchstauglichkeit verfügbarer Datenschutz- und Transparenzfunktionen berücksichtigt.

2. Anatomie eines IoT-Produktes

Ein IoT-Produkt lässt sich zunächst als ein vernetztes Gerät verstehen, das in vielfältigen Ausprägungen existieren kann. Charakteristisch ist, dass ein solches IoT-Gerät häufig nicht isoliert betrieben wird, sondern typischerweise durch eine begleitende Smartphone-App ergänzt wird. Diese Begleit-App dient dazu, das Gerät zu steuern, Konfigurationen vorzunehmen oder Sensordaten auszulesen und in einer für Nutzer:innen zugänglichen Form darzustellen. Bereits an dieser Stelle zeigt sich, dass die Datenverarbeitung häufig nicht nur durch das IoT-Gerät selbst stattfindet, sondern auch auf mit ihm vernetzten Geräten wie beispielsweise Smartphones.

Im Betrieb kommunizieren IoT-Gerät und/oder begleitende App in vielen Fällen mit einem oder mehreren Backend-Diensten im Internet. Diese Backend-Dienste sind für den Betrieb des Produkts teilweise essenziell und können unterschiedliche Aufgaben übernehmen, etwa die Vermittlung von Steuerbefehlen, die Speicherung von Daten oder die Bereitstellung weiterer Funktionalitäten. Dabei ist zu berücksichtigen, dass solche Kommunikationsbeziehungen für Außenstehende beziehungsweise Nutzer:innen häufig nicht transparent sind. Insbesondere bleiben Kommunikationsverbindungen und die beteiligten Dienste oft intransparent, obwohl sie für die datenschutzrechtliche und privatheitsbezogene Bewertung des Gesamtsystems maßgeblich sind.

3. Privatheits-relevante Eigenschaften und ihre Kategorien

Privatheits-relevante Eigenschaften von IoT-Produkten sind ähnlich vielfältig wie die Produkte selbst. Sie hängen unter anderem von der Existenz bestimmter Funktionen, deren technischer Implementierung sowie der Ausgestaltung privatheitsbezogener Transparenz- und Kontrollmechanismen ab. Die in diesem Beitrag verwendete Kategorisierung wurde induktiv aus einer durch Literaturrecherche und Expertenworkshops generierten Liste potenzieller Eigenschaften abgeleitet. Im Folgenden werden die Kategorien und zugehörigen Eigenschaften vorgestellt, die im Hinblick auf eine herstellerunabhängige Überprüfbarkeit ausgewählt und entsprechend gefiltert wurden.

3.1 Funktionsumfang

Der Funktionsumfang umfasst technische Funktionen eines Produktes, die einen Einfluss auf die Privatheit haben können (s. Tab. 1). Privatheitsrelevanz ergibt sich hierbei nicht nur aus der Art der bereitgestellten Funktion, sondern auch aus den Voraussetzungen, unter denen diese Funktion genutzt werden kann. Beispielsweise kann zur Nutzung ein Account bei einem Online-Dienst erforderlich sein. Ebenso kann eine begleitende App Zugriffsberechtigungen benötigen, die ihrerseits datenschutzrelevant sind, weil sie Zugriff auf gerätespezifische oder personenbezogene Daten eröffnen.

Weiterhin kann das IoT-Produkt eine Benutzerverwaltung ermöglichen, wodurch sich Möglichkeiten der Abgrenzung verschiedener Nutzer:innen ergeben. Auch die Erforderlichkeit von Backend-Diensten für den Betrieb ist ein zentraler Bestandteil des Funktionsumfangs, weil damit ein kontinuierlicher Datenaustausch mit externen Systemen verbunden sein kann. Schließlich ist auch relevant, wenn das Produkt im Betrieb das lokale Netzwerk durchsucht, da dadurch Informationen über das Umfeld, andere Geräte oder Netzwerkstrukturen verarbeitet werden können.

Tabelle : Privatheits-relevante Eigenschaften aus der Kategorie *Funktion*.

Name der Eigenschaft	Beschreibung	Quelle
Konto für Geräte-Inbetriebnahme erforderlich ^{D,B}	Ob für den Betrieb ein Konto bei einer Online-Plattform erforderlich ist.	(Hudig u. a. 2023)
Erforderliche App-Berechtigungen ^{D,Q}	Welche Zugriffsberechtigungen die Begleit-App benötigt.	(Zhang/Sadeh 2023)
Zugriffskontrolle für Nutzende ^B	Ob es möglich ist, mehrere Nutzerprofile zu verwalten.	(Norval/Singh 2023)
Erforderlichkeit eines Backend-Dienstes ^B	Ob für den Betrieb ein Backend-Dienst notwendig ist.	(EU 2024; Norval/Singh 2023)
Durchsuchen des lokalen Netzwerks ^B	Ob das Gerät das lokale Netzwerk scannt.	(Norval/Singh 2023)
Open-Source-Basis ^B	Ob das Gerät auf Open-Source-Software basiert.	(EU 2024)
Datenverarbeitung auf dem Gerät ^B	Ob die Datenverarbeitung ausschließlich lokal auf dem Gerät stattfinden kann.	(Norval/Singh 2023)
Steuerebare Erkennungsfunktion ^B	Ob die Suche im lokalen Netzwerk deaktiviert werden kann.	(Norval/Singh 2023)

^B Binär, ^D Details, ^Q Quantifizierend (s. Kapitel 5)

Zur Veranschaulichung zeigen zwei Beispiele, wie die genannten Eigenschaften in der Praxis auftreten können: Ein smartes Türschloss, das biometrische Daten wie Fingerabdrücke ausschließlich lokal auf dem Gerät speichert und verarbeitet, erfüllt die Eigenschaft *Datenverarbeitung auf dem Gerät*. Smart Speaker hingegen sind häufig auf eine aktive Internetverbindung angewiesen, da Sprachbefehle an externe Server übertragen und dort verarbeitet werden (Khalesi 2025). Ohne diese Verbindung ist eine Nutzung des Geräts nicht möglich, was die Relevanz der Eigenschaft *Erforderlichkeit eines Backend-Dienstes* verdeutlicht.

3.2 Datensicherheit und technische Schutzmaßnahmen

Neben dem Funktionsumfang sind Datensicherheit und technische Schutzmaßnahmen als eigenständige Kategorie zu betrachten (s. Tab. 2). Hierzu zählen technische Maßnahmen zum Schutz der Privatheit, etwa sichere und privatheitsfreundliche Grundeinstellungen, die bereits bei der Inbetriebnahme wirksam werden. Zudem sind Funktionen zum Zurücksetzen

der Einstellungen relevant, weil sie eine Wiederherstellung sicherer Zustände oder eine kontrollierte Übergabe des Geräts ermöglichen können.

Ebenso gehören mögliche Authentifizierungsmethoden in diese Kategorie, da sie darüber entscheiden, wie zuverlässig unautorisierte Zugriffe verhindert werden können. Verschlüsselte Kommunikation stellt einen weiteren Kernaspekt dar, weil sie Schutz vor dem Mitlesen oder Manipulieren von Datenübertragungen bietet. Schließlich spielt die Bereitstellung von Updates eine wichtige Rolle: Updates sind Bestandteil technischer Schutzmechanismen, da sie Sicherheitslücken schließen und technische Mindestanforderungen an die Sicherheit eines vernetzten Produkts unterstützen können.

Tabelle 2: *Privatheits-relevante Eigenschaften aus der Kategorie Datensicherheit und technische Schutzmaßnahmen*

Name der Eigenschaft	Beschreibung	Quelle
Sichere und datenschutzfreundliche Grundeinstellungen ^{D,B,W}	Ob das Gerät mit sicheren und datenschutzfreundlichen Standardeinstellungen ausgeliefert wird.	(EU 2016; EU 2024)
Zurücksetzen auf sichere Grundeinstellungen ^B	Ob das Gerät auf sichere und datenschutzfreundliche Standardeinstellungen zurückgesetzt werden kann.	(EU 2016; EU 2024)
Authentifizierungsmethoden ^D	Welche Authentifizierungsmethoden das Gerät anbietet.	(Chen u. a. 2023)
Verschlüsselte Datenübertragung ^{B,W}	Ob das Gerät Daten über verschlüsselte Verbindungen überträgt.	(Railean/Reinhardt 2021; Chhetri/Motti 2022; Lin u. a. 2023; EU 2024; Li u. a. 2024)
TLS-Zertifikatsvalidierung ^{B,W}	Ob TLS-Zertifikate bei verschlüsselter Übertragung validiert werden.	(EU 2024)
(Sicherheits-)Anzeige des Update-Status ^B	Ob der Status von Sicherheitsupdates angezeigt wird.	(Norval/Singh 2023)
Historie von Datenlecks ^D	Welche Vorfälle von Datenlecks in der Vergangenheit im Zusammenhang mit dem Hersteller aufgetreten sind.	(Emami-Naeini u. a. 2019)
Datenschutzverletzungen ^D	Welche Datenschutzverletzungen in der Vergangenheit im Zusammenhang mit dem Hersteller aufgetreten sind.	(EU 2024; EU 2016)
Verschlüsselungsstatus von Nutzerpasswörtern ^B	Ob das Nutzerpasswort verschlüsselt auf dem Gerät gespeichert wird.	(Zhang/Sadeh 2023)

Name der Eigenschaft	Beschreibung	Quelle
Optionale Sicherheitsprüfung ^{D,B}	Ob es zusätzliche Sicherheitsuntersuchungen für das Gerät gibt.	(Zhang/Sadeh 2023)
Privacy-Enhancing Technologies (PETs) ^D	Welche datenschutzfördernden Technologien eingesetzt werden.	(Hudig u. a. 2023; Bermejo Fernandez u. a. 2021)
Update-Fähigkeit ^{B,W}	Ob das Gerät aktualisiert werden kann.	(EU 2024; Chen u. a. 2023; Railean/Reinhardt 2018; Railean/Reinhardt 2021)

^B Binär, ^D Details, ^W Bewertend (s. Kapitel 5)

Die folgenden Beispiele veranschaulichen die Relevanz dieser Kategorie in der Praxis. Ein Smart-TV, der mit aktivierter ACR-Funktion ausgeliefert wird, verdeutlicht die Eigenschaft *Sichere und datenschutzfreundliche Grundeinstellungen*: ACR erfasst standardmäßig Bildschirmaufnahmen und übermittelt diese an Hersteller und Werbetreibende, ohne dass Nutzende davon in der Regel Kenntnis haben (Anselmi u. a. 2024). Eine smarte Video-Türklingel, die Videostreams unverschlüsselt überträgt, illustriert die Eigenschaften *Verschlüsselte Datenübertragung* und *TLS-Zertifikatsvalidierung*: Ohne Verschlüsselung können Bilddaten im Netzwerkverkehr abgefangen werden, und eine fehlende Zertifikatsvalidierung ermöglicht Man-in-the-Middle-Angriffe (Khetagourova u. a. 2025).

3.3 Transparenz

Transparenz bezieht sich auf Informationen zur Verarbeitung von Daten und umfasst unter anderem typische Transparenzinformationen im Sinne der DSGVO (s. Tab. 3). Dazu zählt nicht nur die Beschreibung, welche Daten verarbeitet werden, sondern auch aggregierte Eigenschaften wie das Teilen von Daten mit Dritten. Eine besondere Rolle spielt zudem die Auflistung verarbeiteter Datentypen, einschließlich der expliziten Benennung nicht-personenbezogener Daten. Diese sind wichtig, weil auch nicht-personenbezogene Daten durch Verkettung mit anderen Daten privatheits-relevant werden können.

Zentral sind außerdem Angaben zu Datenempfängern und zu Zwecken der Datenverarbeitung. Diese Informationen sind notwendig, um Verarbeitungskontexte nachvollziehen zu können und um zu verstehen, welche Akteure in Datenflüsse eingebunden sind.

Tabelle 3: *Privatheits-relevante Eigenschaften aus der Kategorie Transparenz*

Name der Eigenschaft	Beschreibung	Quelle
Verarbeitete Datentypen ^{D,Q}	Welche Daten beim Betrieb des Geräts verarbeitet werden.	(Zhang/Sadeh 2023)
Auflistung nicht-personenbezogener Daten ^{D,Q}	Welche nicht-personenbezogenen Daten beim Betrieb des Geräts verarbeitet werden.	(Railean/Reinhardt 2021)
Besondere Kategorien personenbezogener Daten ^{D,B,Q}	Welche besonderen Kategorien personenbezogener Daten beim Betrieb des Geräts verarbeitet werden.	(EU 2016; EDPB 2020)
Datenweitergabe ^{B,Q}	Welche Daten mit Dritten geteilt werden.	(Zhang/Sadeh 2023)
Datenempfänger ^{D,Q}	Welche Datenempfänger (Unternehmen, Herkunft) beteiligt sind.	(Zhang/Sadeh 2023)
Datenverarbeitung außerhalb der EU ^B	Ob Daten außerhalb der EU verarbeitet werden.	(EU 2016; Railean/Reinhardt 2021)
Profiling ^{D,B}	Ob Profiling stattfindet.	(EU 2016)
Rechtsgrundlage der Datenverarbeitung ^D	Auf welcher Rechtsgrundlage die Daten verarbeitet werden.	(EU 2016)
Speicherdauer ^D	Wie lange die Daten gespeichert werden.	(EU 2016; Railean/Reinhardt 2018; Railean/Reinhardt 2021)
Beteiligte Akteure in der Verarbeitung ^D	Welche Akteure an der Datenverarbeitung beteiligt sind.	(EU 2016)
Zweck der Datenverarbeitung ^{D,Q}	Zu welchen Zwecken die Daten verarbeitet werden.	(EU 2016; Railean/Reinhardt 2018; Railean/Reinhardt 2021; Cranor u. a. 2024; Chen u. a. 2023)
Datenminimierung ^D	Inhaltlich ähnlich wie „Verarbeitete Datentypen“, ergänzt um den Abgleich mit der Erforderlichkeit (Einordnung, welche Übermittlungen üblicherweise für eine Geräteklasse stattfinden).	(EU 2016; EU 2024)
Häufigkeit der Datenübertragung ^D	Wie oft Daten übertragen werden.	(Railean/Reinhardt 2018; Railean/Reinhardt 2021)
Einhaltung des Cyber Resilience Act ^B	Ob das Gerät die Anforderungen des Cyber Resilience Act erfüllt.	(EU 2024)
Zielort der Datenübertragung ^{D,Q}	Wohin Daten übertragen werden.	(EU 2016; Railean/Reinhardt 2021)

^B Binär, ^D Details, ^Q Quantifizierend (s. Kapitel 5)

Wie relevant die Betrachtung der Eigenschaften *Datenempfänger*, *Datenweitergabe* und *Zweck der Datenverarbeitung* in der Praxis ist, zeigt eine Studie, die für 43 IoT-Produkte DSGVO-Auskunftsanfragen stellte: Nur bei 16 der 43 Produkte wurden überhaupt Informationen zu Drittempfängern bereitgestellt, meist nur in vagen Kategorien wie „Cloud-Dienste“ oder „Werbetreibende“, ohne konkrete Namen oder Standorte der Empfänger zu nennen. Beim Verarbeitungszweck verwendeten die meisten Hersteller Formulierungen wie „zur Verbesserung unserer Dienste“, ohne eine klare Zuordnung zu konkreten Datenkategorien oder Rechtsgrundlagen (Hudig u. a. 2025).

3.4 Kontrolle

Kontrolle umfasst Funktionen, die eine Steuerung oder Begrenzung der Nutzung eigener Daten ermöglichen (s. Tab. 4). Dazu gehören Möglichkeiten, Einfluss auf verarbeitete Datentypen oder Datenempfänger zu nehmen, soweit entsprechende Konfigurationsoptionen vorhanden sind. Weiterhin ist die Implementierung der Betroffenenrechte ein Bestandteil dieser Kategorie, ebenso wie Datenübertragbarkeit. In diesem Zusammenhang ist Kontrolle nicht nur als abstraktes Prinzip zu verstehen, sondern als konkrete Funktionalität, die im Produkt oder in der App implementiert sein kann.

Tabelle 4: Privatheits-relevante Eigenschaften aus der Kategorie Kontrolle

Name der Eigenschaft	Beschreibung	Quelle
Kontrolle des Verarbeitungs-orts ^{D,B}	Ob der Nutzer wählen kann, wo Daten verarbeitet werden.	(Norval/Singh 2023)
Kontrolle der verarbeiteten Daten ^{D,B}	Ob ausgewählt werden kann, welche Daten verarbeitet werden.	(EU 2016)
Kontrolle des Datenempfängers ^{D,B}	Ob ausgewählt werden kann, welche Empfänger Daten erhalten.	(EU 2016)
Umsetzung der Betroffenenrechte ^{D,W}	Ob die Betroffenenrechte umgesetzt wurden.	(EU 2016)
Datenportabilität ^{B,W}	Ob Daten vom Gerät auf andere Geräte übertragen werden können.	(EU 2016)

^B Binär, ^D Details, ^W Bewertend (s. Kapitel 5)

3.5 Usability

Usability adressiert die Gebrauchstauglichkeit privatheits-relevanter Funktionen (s. Tab. 5). Hier stellt sich unter anderem die Frage, ob eine informierte Einwilligung zur Datenverarbeitung ermöglicht wird. Falls andere Rechtsgrundlagen genutzt werden, ist es relevant, ob die Datenverarbeitung klar kommuniziert wird. Ebenfalls entscheidend ist, ob überhaupt eine gebrauchstaugliche Benutzeroberfläche für Privatheitseinstellungen bereitgestellt wird. Damit wird Usability zu einem Faktor, der unmittelbar beeinflusst, ob vorhandene Schutz- oder Kontrollfunktionen praktisch genutzt werden können.

Tabelle 5: Privatheits-relevante Eigenschaften aus der Kategorie Usability

Name der Eigenschaft	Beschreibung	Quelle
Informierte Einwilligung einholen (falls erforderlich) ^{B,W}	Ob eine informierte Einwilligung zur Datenverarbeitung eingeholt wird.	(Bermejo Fernandez u. a. 2021)
Aufklärung über Datenverarbeitungsprozesse ^{B,W}	Ob die Datenverarbeitung klar kommuniziert wird, auch wenn keine Einwilligung notwendig ist.	(Emami-Naeini u. a. 2019)
Benutzeroberfläche zur Datenschutzkontrolle ^B	Ob es datenschutzbezogene Einstellungen gibt.	(Cruz u. a. 2023)
Datenschutzinfo getrennt von AGB ^{B,W}	Ob Datenschutzinformationen und AGB klar voneinander getrennt sind.	(EU 2016)
Klare und transparente Informationen zur Verarbeitung personenbezogener Daten ^B	Ob klare und transparente Informationen zur Verarbeitung personenbezogener Daten bereitgestellt werden.	(Emami-Naeini u. a. 2019)
Visuelle Darstellung von Datenflüssen ^B	Ob eine Visualisierung der Datenflüsse bereitgestellt wird.	(Cruz u. a. 2023)
Unterschiedliche Granularität von Datenschutzeinstellungen in der UI ^B	Ob Datenschutzeinstellungen in unterschiedlichen Detailgraden verfügbar sind.	(Chhetri/Motti 2022)
Explorationsfunktionen ^B	Ob das Gerät das Erkunden von Systeminteraktionen unterstützt (z. B. Kommunikationsliste, Visualisierung von Aktivitäten).	(Norval/Singh 2023)
Untersuchungsfunktionen ^B	Ob das Gerät Funktionen zur Untersuchung bereitstellt (z. B. Fehlerprotokolle).	(Norval/Singh 2023)
Benachrichtigung bei Handlungsbedarf ^B	Ob das Gerät den Nutzer benachrichtigt, wenn Handlungen erforderlich sind.	(Norval/Singh 2023)
Kontextbezogene Datenoffenlegung ^B	Ob das Gerät kontextbezogene Informationen bereitstellt, wenn Daten offengelegt werden.	(Bermejo Fernandez u. a. 2021)

Name der Eigenschaft	Beschreibung	Quelle
Zusätzliche Transparenzmechanismen ^D	Ob das Gerät zusätzliche Transparenzmechanismen bietet.	(Chhetri/Motti 2022)

^B Binär, ^D Details, ^Q Quantifizierend, ^W Bewertend (s. Kapitel 5)

Datenschutzerklärungen für Smart-Home-Geräte sind häufig unvollständig oder lassen unklar, ob sie für ein bestimmtes Gerätemodell, die zugehörige App oder die Herstellerwebsite gelten, was die Relevanz der Eigenschaft *Klare und transparente Informationen* unterstreicht (Manandhar u. a. 2022; Rauti u. a. 2023). Kontextuelle Datenschutzrichtlinien, die relevante Hinweise direkt an der entsprechenden Stelle in der App-Oberfläche einblenden, erzielen hingegen eine deutlich höhere Lesebereitschaft bei Nutzenden und veranschaulichen damit die Bedeutung der Eigenschaft *Kontextbezogene Datenoffenlegung* (Pan u. a. 2024).

3.6 Datenschutz-Support

Datenschutz-Support beschreibt die Unterstützung von Nutzer:innen bei der Erfüllung datenschutzrechtlicher Pflichten (s. Tab. 6). Gerade bei IoT-Geräten ist dies bedeutsam, weil der Besitzer eines IoT-Produktes in die Rolle des für die Datenverarbeitung Verantwortlichen geraten kann. Ein Beispiel ist der Betrieb einer Kamera auf dem eigenen Grundstück, bei dem Gäste gefilmt werden. In solchen Fällen können Eigenschaften wie Compliance-Unterstützung oder Vertragsvorlagen für gemeinsame Verantwortlichkeit relevant sein. Datenschutz-Support zielt somit darauf ab, weniger versierte Nutzer:innen nicht nur technisch, sondern auch in ihrer potenziellen rechtlichen Rolle zu unterstützen.

Tabelle 6: Privatheits-relevante Eigenschaften aus der Kategorie Datenschutz-Support

Name der Eigenschaft	Beschreibung	Quelle
Unterstützung bei Compliance ^{D,B}	Ob der Hersteller Informationen zur Einhaltung von Anforderungen bereitstellt (z. B. Privatsphäre von Unbeteiligten).	(Railean/Reinhardt 2021)
Verfügbarkeit von Vertragsvorlagen für gemeinsame Verantwortliche/Auftragsverarbeiter ^{D,B}	Ob der Hersteller Vertragsvorlagen für eine Konstellation als gemeinsam Verantwortliche bzw. Auftragsverarbeiter bereitstellt.	(EU 2016)

^B Binär, ^D Details (s. Kapitel 5)

Smart-Home-Nutzende können durch die Einschränkung der Haushaltsausnahme unter Umständen selbst zur datenschutzrechtlich verantwortlichen Stelle werden, was die Relevanz der Eigenschaft *Unterstützung bei Compliance* unterstreicht (Urquhart/Chen 2021; Chen u. a. 2020). Die von Herstellern angebotene Unterstützung beschränkt sich dabei häufig auf den Verweis, dass Nutzende selbst für die Einhaltung geltenden Rechts verantwortlich seien, sowie auf die Empfehlung, Unbeteiligte durch physische Hinweisschilder auf die Datenerfassung aufmerksam zu machen (Alshehri u. a. 2022).

4. Analyseverfahren und Messbarkeit

Eine zentrale Herausforderung im Zusammenhang mit privatheits-relevanten Eigenschaften von IoT-Geräten besteht darin, dass Messungen und Analysen auch durch Laien ermöglicht werden müssen. Grundsätzlich lassen sich zwei Ansätze unterscheiden.

4.1 Manuelle Untersuchungen

Ein Ansatz ist die manuelle Untersuchung anhand von Checklisten und Analyseprotokollen. Diese ermöglichen es, Eigenschaften systematisch zu erfassen, die anderweitig nicht zu erheben sind, etwa aus den Kategorien Funktion (z. B. ob ein Benutzerkonto für den Betrieb notwendig ist), Datensicherheit und technische Schutzmaßnahmen (z. B. das Zurücksetzen eines Geräts auf sichere Grundeinstellungen), Datenschutz-Support (z. B. Unterstützung bei Compliance), Usability (z. B. Aufklärung über Datenverarbeitung oder Granularität der Datenschutzeinstellungen) oder Kontrolle (z. B. Kontrolle über Verarbeitungsort, Datenkategorien und Empfänger). Durch die Strukturierung mittels Checklisten werden Beobachtungen reproduzierbarer und die Vergleichbarkeit wird erleichtert, ohne dass hierfür zwingend spezialisierte Messtechnik erforderlich wäre.

4.2 Automatisierte Analysen

Daneben existieren automatisierte Analyseverfahren, die insbesondere dort relevant werden, wo Datenflüsse technisch erfasst oder große Mengen an Informationen ausgewertet werden sollen.

Netzwerkschnittstellen

Die Kommunikation von IoT-Geräten kann vielfältig ausgestaltet sein. Sie kann entweder über eine direkte Verbindung mittels WLAN-Access-Point erfolgen, alternativ über die App ins Internet oder über ein spezielles IoT-Gateway. Um möglichst viele der Kommunikationspfade abdecken zu können, bietet sich eine Messung am lokalen Router, also am Übergang vom lokalen zum öffentlichen Netz, an. Dadurch wird ein Ort gewählt, an dem sich verschiedene Kommunikationsvarianten zusammenführen und beobachten lassen. Für den Fall, dass die Kommunikation eines IoT-Geräts mit dem Internet hauptsächlich über eine Begleit-App erfolgt, entsteht hierbei jedoch die Herausforderung, die Kommunikation im Zusammenhang mit dem IoT-Gerät vom Datenverkehr zu trennen, der durch andere Anwendungen auf dem Smartphone initiiert wird.

Durch die Analyse von IP-Netzwerkschnittstellen können Eigenschaften aus den Kategorien Datensicherheit und technische Schutzmaßnahmen (z. B. Verschlüsselung der Kommunikation) und Transparenz (z. B. Datenverarbeitung außerhalb der EU, Datenempfänger) erhoben werden. Die Netzwerkanalyse wird damit zu einem Instrument, um intransparente Kommunikationsbeziehungen zumindest teilweise sichtbar zu machen (Jakaria u. a. 2024).

Datenschutzerklärungen

Ein weiterer Gegenstand automatisierter Analyse ist die Auswertung von Datenschutzerklärungen (Yang u. a. 2025). Ziel ist hierbei die automatisierte Extraktion von Angaben, beispielsweise zu verarbeiteten Datentypen, zum Einsatz von Profiling, zu Verarbeitungszwecken und Datenempfängern. In der Praxis werden Datenschutzerklärungen jedoch häufig nicht spezifisch für einzelne IoT-Produkte bereitgestellt, sondern global für einen Hersteller. Dadurch enthalten sie Informationen zu vielen verschiedenen Kontexten, etwa zur Nutzung der Website oder zu Bewerbungsprozessen.

Für eine automatisierte Analyse ergibt sich daraus die Notwendigkeit, den Kontext der Informationen aus Datenschutzerklärungen festzustellen und zu berücksichtigen. Ohne Kontextbezug besteht das Risiko, dass Aussagen aus nicht passenden Abschnitten fälschlich einem IoT-Produkt zugeordnet werden. Ist diese Voraussetzung erfüllt, können insbesondere Eigenschaften aus der Kategorie Transparenz (z. B. Verarbeitungszwecke oder Profiling) automatisiert erhoben werden.

5. Vergleichbarkeit und Bewertbarkeit gesammelter Eigenschaften

Für die Einordnung privatheits-relevanter Informationen stellt sich die Frage, wie Eigenschaften vergleichbar und bewertbar gemacht werden können. Bisherige Privacy Labels arbeiten oft mit Skalen, um die Privatheitsfreundlichkeit in unterschiedlichen Kategorien zu bewerten. Studien haben jedoch gezeigt, dass Nutzende weniger an Bewertungen als an konkreten Details beziehungsweise Informationen interessiert sind (Johnson u. a. 2020). Insbesondere für die Abwägung des Kaufs eines Produktes ist eine schnelle Vergleichbarkeit jedoch unabdingbar, was im Widerspruch zu detaillierten Angaben bezüglich der Privatheit stehen kann. Um die vorgestellten Eigenschaften dennoch strukturiert im Rahmen einer Plattform darstellen zu können, können diese in Bezug auf ihre Vergleichbarkeit und in Teilen auch Bewertbarkeit kategorisiert werden.

5.1 Binäre Eigenschaften

Binäre Eigenschaften ermöglichen eine Aussage über die Existenz einer Funktion oder Gegebenheit. Beispiele sind die Notwendigkeit eines Accounts oder die Nutzung von Backend-Diensten. Solche Eigenschaften erlauben meist keine differenzierte Bewertung oder Anwendung von Skalen, können aber für eine grundlegende Vergleichbarkeit herangezogen werden, weil sie klare Ja-/Nein-Aussagen liefern.

5.2 Detailangaben

Detailangaben sind strukturierte Informationen, die Eigenschaften eines Produktes beschreiben, beispielsweise eine Liste der App-Zugriffsberechtigungen oder die Benennung von Datenempfängern. Auch hier ist eine Bewertung meist nicht möglich, da diese Angaben oft mit dem Empfinden und den Präferenzen der Nutzer:innen in Verbindung stehen. Eine konsistente Struktur der Detailangaben ermöglicht jedoch Vergleiche und Filterungen, bei denen persönliche Präferenzen berücksichtigt werden können.

5.3 Quantifizierende Eigenschaften

Quantifizierende Eigenschaften beschreiben eine Eigenschaft über eine Anzahl, etwa die Anzahl der App-Zugriffsberechtigungen oder die Menge

verarbeiteter Datentypen. Auch dies stellt keine Bewertung dar, ermöglicht aber einfache Vergleiche zwischen ähnlichen Produkten, weil Unterschiede numerisch darstellbar werden.

5.4 Bewertungen

Nur wenige Eigenschaften erlauben direkte Bewertungen eines Produktes. Dies steht häufig im Zusammenhang mit technischen und rechtlichen Mindeststandards. Beispielsweise gilt ein Produkt, das mit dem Internet verbunden wird, aber nicht geupdatet werden kann, nach heutigem Stand der Technik als unzureichend. Werden beim Betrieb eines IoT-Gerätes personenbezogene Daten durch einen Cloud-Dienst verarbeitet, ist die Implementierung der Betroffenenrechte nach DSGVO unverzichtbar. Ebenso ist unverschlüsselte Kommunikation mit Backend-Diensten ein Beispiel für eine Eigenschaft, die einen aufklärenden Hinweis nahelegt, weil sie grundlegende Schutzziele der IT-Sicherheit verletzt.

6. Individualisierbare Bewertungssystematik

Da das Privatheitsempfinden grundsätzlich subjektiv ist und IoT-Produkte sich über vielfältige Produktkategorien erstrecken, ergibt sich die Notwendigkeit einer flexiblen Bewertungssystematik. Diese sollte insbesondere berücksichtigen, dass vergleichende Aussagen nur innerhalb einer Produktkategorie sinnvoll sind, weil Eigenschaften spezifisch für bestimmte Kategorien sein können. So ist etwa die Verarbeitung von Videodaten bei Kameras trivial, während sie bei Kinderspielzeug in einem anderen Kontext steht und anders gewichtet werden kann.

Darüber hinaus sollte die Priorisierung der Darstellung privatheits-relevanter Eigenschaften personalisierbar sein, da das Empfinden der Sensitivität unterschiedlicher Eigenschaften individuell abweichen kann. Zusätzlich kann jeweils nur eine begrenzte Auswahl an privatheits-relevanten Eigenschaften gleichzeitig präsentiert werden. Dies macht eine Priorisierung erforderlich, die nicht allein allgemein festgelegt ist, sondern anpassbar gestaltet werden muss, um unterschiedlichen Bedürfnissen und Präferenzlagen gerecht zu werden. Um eine solche Priorisierung fundiert gestalten zu können, sollte die empfundene Relevanz privatheitsrelevanter Eigenschaften durch Nutzerstudien erhoben werden. Da diese Wahrnehmung subjek-

tiv und kontextabhängig ist, gilt dies sowohl allgemein als auch in Abhängigkeit von der jeweiligen Produktkategorie. Die Ergebnisse können als Grundlage dienen, um sinnvolle Standardpriorisierungen abzuleiten und den Gestaltungsraum für individuelle Anpassungen empirisch abzusichern.

7. Fazit

Die identifizierten Privatheitseigenschaften ermöglichen es Endnutzer:innen, fundierte Entscheidungen beim Erwerb von IoT-Geräten zu treffen. Darüber hinaus bieten sie Verbraucherzentralen eine belastbare Grundlage für die Erstellung von Berichten sowie für die Beratung im Kontext des IoT. Auch Aufsichtsbehörden und Gesetzgeber profitieren von diesen Erkenntnissen, da sie einen aktuellen Überblick über bestehende Datenschutzpraktiken erhalten und darauf aufbauend langfristige Maßnahmen zur Verbesserung des Datenschutzes und des Schutzes der Privatheit entwickeln können.

Gleichwohl ist die resultierende Sammlung von 54 Eigenschaften nicht als abschließend zu betrachten. Die gewählte Methodik – eine fokussierte Literaturrecherche, ergänzt durch Expertenworkshops – war darauf ausgelegt, eine breite und praxisrelevante Merkmalsbasis zu erarbeiten, nicht jedoch darauf, sämtliche denkbaren Eigenschaften lückenlos zu erheben, wie es etwa eine systematische Literaturrecherche anstreben würde. Der interdisziplinären Zusammensetzung der Expertengruppe kam dabei eine kompensierende Rolle zu, da die Vielfalt der fachlichen Perspektiven dazu beitrug, wesentliche Lücken im Eigenschaftsraum zu vermeiden. Darüber hinaus sind weiterführende Arbeiten erforderlich, um die Sammlung auf eine übersichtlichere und damit gebrauchstauglichere Anzahl zu reduzieren. Eine solche Reduktion sollte sowohl die Komplexität der Erhebung einzelner Eigenschaften berücksichtigen, die eng mit der konkreten Implementierung der jeweiligen Messinstrumente zusammenhängt, als auch die Perspektive von Nutzer:innen einbeziehen. Hierzu sind Nutzerstudien notwendig, die untersuchen, welche Eigenschaften als besonders relevant wahrgenommen werden und in welcher Priorisierung sie dargestellt werden sollten.

Literaturverzeichnis

- Alshehri, Ahmed Mohammed; Spielman, Joseph; Prasad, Amiya und Yue, Chuan (2022): Exploring the Privacy Concerns of Bystanders in Smart Homes from the Perspectives of Both Owners and Bystanders. *Proceedings on Privacy Enhancing Technologies*. URL: <https://petsymposium.org/popets/2022/popets-2022-0064.pdf>
- Anselmi, Gianluca; Vekaria, Yash; D'Souza, Alexander; Callejo, Patricia; Mandalari, Anna Maria und Shafiq, Zubair (2024): Watching TV with the Second-Party: A First Look at Automatic Content Recognition Tracking in Smart TVs. In: *Proceedings of the 2024 ACM on Internet Measurement Conference (IMC '24)*, S. 622–634. New York: Association for Computing Machinery. <https://doi.org/10.1145/3646547.3689013>
- Bermejo Fernandez, Carlos; Lee, Lik Hang; Nurmi, Petteri und Hui, Pan (2021): PARA: Privacy Management and Control in Emerging IoT Ecosystems Using Augmented Reality. In: *Proceedings of the International Conference on Multimodal Interaction*. <https://doi.org/10.1145/3462244.3479885>
- Chen, Jiahong; Edwards, Lilian; Urquhart, Lachlan und McAuley, Derek (2020): Who Is Responsible for Data Processing in Smart Homes? Reconsidering Joint Controller-ship and the Household Exemption. *International Data Privacy Law* 10 (4): S. 279–298. <https://doi.org/10.1093/idpl/ipaa011>
- Chen, Claire; Ravishankar, Hamsini; Shu, Dillon; Jain, Laxita; Zeng, Yifeng; Agarwal, Yuvraj und Cranor, Lorrie Faith (2023): Ask the Consumers: What Should Be On IoT Privacy & Security Labels. Poster at the *Symposium on Usable Privacy and Security (SOUPS)*.
- Chhetri, Chola und Motti, Vivian (2022): Designing and Evaluating a Prototype for Data-Related Privacy Controls in a Smart Home. In: *International Symposium on Human Aspects of Information Security and Assurance*. https://doi.org/10.1007/978-3-031-12172-2_19
- Cranor, Lorrie Faith; Agarwal, Yuvraj und Emami-Naeini, Pardis (2024): Internet of Things Security and Privacy Labels Should Empower Consumers. *Communications of the ACM* 67.
- Cruz, Stefany; Danek, Logan; Liu, Shinan; Kraemer, Christopher; Wang, Zixin; Feamster, Nick; Huang, Danny Yuxing; Yao, Yaxing und Hester, Josiah (2023): Augmented Reality's Potential for Identifying and Mitigating Home Privacy Leaks. In: *Symposium on Usable Security and Privacy (USEC)*. <https://doi.org/10.14722/usec.2023.238930>
- Emami-Naeini, Pardis; Dixon, Henry; Agarwal, Yuvraj und Cranor, Lorrie Faith (2019): Exploring How Privacy and Security Factor Into IoT Device Purchase Behavior. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3290605.3300764>
- Emami-Naeini, Pardis; Dheenadhayalan, Janarth; Agarwal, Yuvraj und Cranor, Lorrie Faith (2022): An Informative Security and Privacy "Nutrition" Label for Internet of Things Devices. *IEEE Security & Privacy* 20 (2): S. 31–39. <https://doi.org/10.1109/MS-EC.2021.3132398>

- Europäische Union (EU) (2016): Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates (Datenschutz-Grundverordnung, DSGVO).
- Europäische Union (EU) (2024): Verordnung (EU) 2024/2847 des Europäischen Parlaments und des Rates (Cyber Resilience Act, CRA).
- European Data Protection Board (EDPB) (2020): Guidelines 05/2020 on Consent Under Regulation 2016/679. Version 1.1.
- Fuhrken, Stefan (2026): BGH Verhandlungsterminhinweis I ZR 289/25 (Videoüberwachung in der Wohnküche) – Lünedatenschutz. Februar 9. URL: <https://xn--luedatenschutz-gsb.de/2026/02/09/bgh-verhandlungsterminhinweis-i-zr-289-25-videoueberwachung-in-der-wohnekueche/> (besucht am: 14.04.2026)
- Greveler, Ulrich; Justus, Benjamin und Löhr, Dennis (2012): Identifikation von Videoinhalten über granulare Stromverbrauchsdaten. *SICHERHEIT 2012 – Sicherheit, Schutz und Zuverlässigkeit*, S. 35–45. Bonn: Gesellschaft für Informatik e.V.
- Hudig, Anna Ida; Norval, Chris und Singh, Jatinder (2023): Transparency in the Consumer Internet of Things: Data Flows and Data Rights. London: Information Commissioner's Office.
- Hudig, Anna Ida; Mandalari, Anna Maria; Norval, Chris; Haddadi, Hamed; Binns, Reuben und Singh, Jatinder (2025): Rights Out of Sight: Data Practices and Transparency Gaps in Smart Consumer IoT Ecosystems. In: *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*, S. 2260–2273. New York: Association for Computing Machinery. <https://doi.org/10.1145/3715275.3732154>
- Jakaria, Md; Huang, Danny Yuxing und Das, Anupam (2024): Connecting the Dots: Tracing Data Endpoints in IoT Devices. *Proceedings on Privacy Enhancing Technologies*. URL: <https://petsymposium.org/popets/2024/popets-2024-0090.php>
- Johnson, Shane D.; Blythe, John M.; Manning, Matthew und Wong, Gabriel T. W. (2020): The Impact of IoT Security Labelling on Consumer Product Choice and Willingness to Pay. *PLOS ONE* 15 (1). <https://doi.org/10.1371/journal.pone.0227800>
- Khalesi, Ben (2025): I Thought I Had a Smart Home, but Then I Turned Off the Internet. *Android Police*, Juli 20. URL: <https://www.androidpolice.com/smart-home-vs-no-wifi/>
- Khetagourova, Natalia; Lundh, Christoffer und Kävrestad, Joakim (2025): Evaluating Security and Data Privacy in Smart Home Devices. In: *Socio-Technical Perspectives in Information Systems 2025: Proceedings of the 11th International Workshop on Socio-Technical Perspectives in Information Systems (STPIS 2025)*, CEUR Workshop Proceedings, Bd. 4134.
- Li, Tianshi; Cranor, Lorrie Faith; Agarwal, Yuvraj und Hong, Jason I. (2024): Matcha: An IDE Plugin for Creating Accurate Privacy Nutrition Labels. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 8 (1). <https://doi.org/10.1145/3643544>
- Lin, Yanzi; Juneja, Jaideep; Birrell, Eleanor und Cranor, Lorrie Faith (2023): Data Safety vs. App Privacy: Comparing the Usability of Android and iOS Privacy Labels. *arXiv preprint arXiv:2312.03918*.

- Manandhar, Sunil; Roundy, Kevin; Tamersoy, Acar; Memon, Nasir; Kirda, Engin; Yao, Daphne und Kang, Byungkoon (2022): Smart Home Privacy Policies Demystified: A Study of Availability, Content, and Coverage. In: *31st USENIX Security Symposium* (USENIX Security 22). USENIX Association. <https://www.usenix.org/conference/usenixsecurity22/presentation/manandhar>
- Norval, Chris und Singh, Jatinder (2023): A Room With an Overview: Toward Meaningful Transparency for the Consumer Internet of Things. *IEEE Internet of Things Journal* 11 (5). <https://doi.org/10.1109/JIOT.2023.3318369>
- Pan, Shidong; Tao, Zhen; Hoang, Thong; Zhang, Dawen; Li, Tianshi; Xing, Zhenchang; Xu, Xiwei; Staples, Mark; Rakotoarivelo, Thierry und Lo, David (2024): A New Hope: Contextual Privacy Policies for Mobile Applications and An Approach Toward Automated Generation. In *33rd USENIX Security Symposium* (USENIX Security 24), S. 5699–5716. Philadelphia: USENIX Association. <https://www.usenix.org/conference/usenixsecurity24/presentation/pan-shidong-hope>
- Railean, Alexandr und Reinhardt, Delphine (2018): Let There Be Lite: Design and Evaluation of a Label for IoT Transparency Enhancement. In: *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. <https://doi.org/10.1145/3236112.3236126>
- Railean, Alexandr und Reinhardt, Delphine (2021): OnLITE: On-Line Label for IoT Transparency Enhancement. In: Secure IT Systems, herausgegeben von Mikael Asplund und Simin Nadjm-Tehrani, Bd. 12556. *Lecture Notes in Computer Science*. Springer International Publishing. https://doi.org/10.1007/978-3-030-70852-8_14
- Rauti, Sampsa; Laato, Samuli und Pitkämäki, Tinja (2023): An Assessment of Privacy Policies for Smart Home Devices. In *Proceedings of the 24th International Conference on Computer Systems and Technologies*. New York: Association for Computing Machinery. <https://doi.org/10.1145/3606305.3606332>
- Reinhardt, Andreas; Englert, Frank und Christin, Delphine (2015): Averting the Privacy Risks of Smart Metering by Local Data Preprocessing. *Pervasive and Mobile Computing* 16 (Januar): S. 171–183. <https://doi.org/10.1016/j.pmcj.2014.10.002>
- Saqib, Eimaan; He, Shijing; Choy, Junghyun; Abu-Salma, Ruba; Such, Jose; Bernd, Julia und Javed, Mobin (2025): Bystander Privacy in Smart Homes: A Systematic Review of Concerns and Solutions. *ACM Transactions on Computer-Human Interaction*, Mai 8. <https://doi.org/10.1145/3731755>
- Shen, Yun und Vervier, Pierre-Antoine (2019): IoT Security and Privacy Labels. In: Privacy Technologies and Policy, APF 2019, herausgegeben von M. Naldi, G. F. Italiano, K. Rannenberg, M. Medina und A. Bourka, Bd. 11498. *Lecture Notes in Computer Science*. https://doi.org/10.1007/978-3-030-21752-5_9
- Subahi, Alanoud und Theodorakopoulos, George (2019): Detecting IoT User Behavior and Sensitive Information in Encrypted IoT-App Traffic. *Sensors* 19 (21). <https://doi.org/10.3390/s19214777>
- Urquhart, Lachlan und Chen, Jiahong (2021): On the Principle of Accountability: Challenges for Smart Homes & Cybersecurity. In: *Privacy by Design for the Internet of Things: Building Accountability and Security*, herausgegeben von Andrew Crabtree, Hamed Haddadi und Richard Mortier, S. 19–47. Stevenage: IET.

- Yang, Mian; Atluri, Vijayalakshmi; Sural, Shamik und Kundu, Ashish (2025): Automatized Privacy Policy Analysis Using Large Language Models. In: *Data and Applications Security and Privacy XXXIX: 39th IFIP WG 11.3 Annual Conference*, DBSec 2025, Gjøvik, Norway, June 23–24, 2025, Proceedings, S. 23–43. https://doi.org/10.1007/978-3-031-96590-6_2
- Zhang, Shikun und Sadeh, Norman (2023): Do Privacy Labels Answer Users' Privacy Questions? In *Workshop on Usable Security and Privacy*.

Einfache Sensoren, komplexe Folgen: Privatheit als soziale Praxis im vernetzten Haushalt

Tanja Lehmann, Franziska Baum, Albrecht Kurze, Arne Berger und Andreas Bischof

Zusammenfassung

Smart-Home-Systeme wie Kühlschränke, Thermostate und Rauchmelder bringen eine Vielzahl an Sensoren in die Häuslichkeit. Deren Funktionsweisen und Dimensionen der Datenerfassung sind häufig unbekannt. Eine Auseinandersetzung mit den Sensoren des *Smart Homes* und ihren Daten ist in Privatheitsdiskursen noch selten. Vor diesem Hintergrund zeigt der Artikel, dass *Smart-Home*-Sensoren (z. B. Temperatur, Luftfeuchte, Helligkeit, Bewegung, Lautstärke) erhebliche Implikationen für Privatheit im (in der Regel geteilten) Zuhause haben. In Haushalten werden diese Daten routinemäßig mit Alltagswissen verknüpft und zu Geschichten über Anwesenheit, Rhythmen, Aktivitäten oder Abweichungen verdichtet; Messwerte erhalten dadurch soziale Bedeutung und autoritativen Status. Das ermöglicht laterale Überwachung, Normdurchsetzung und Konflikte, auch ohne bössartige Absicht und verstärkt Informationsasymmetrien zwischen primären Administrator:innen und sekundären Nutzer:innen (Kinder, Partner:innen, Besucher:innen, Dienstleister:innen). Der Beitrag schlägt daher vor, Privatheit als relationalen Aushandlungsraum zu verstehen und zu gestalten. Methodisch nutzt die Studie, auf der dieser Beitrag wesentlich aufbaut, ein *Sensorkit*, das über zwei Wochen in Haushalten Daten erhebt, auf einem Tablet visualisiert und diese Daten in Aufgaben, Interviews sowie Gruppendiskussionen partizipativ interpretieren lässt. Die Auswertung der Interviews und Gruppendiskussionen erfolgt qualitativ-inhaltsanalytisch in Gruppeninterpretationen der interdisziplinären Forschendengruppe. So werden Interpretierbarkeit, Machtpraktiken, Akteurskonstellationen und mentale Modelle von Nutzer:innen und Betroffenen solcher nicht-audiovisueller Sensorik sichtbar. Daraus folgen Design- und Nutzungsempfehlungen: rollenbasierte Zugriffsrechte, transparente Datenerfassung, Zweckbindung, datensparsame Defaults, lokale Speicherung, einfache Pausen-/Löschfunktionen, sowie Aushandlungs- und Bildungsfor-

mate und regulatorische Mindeststandards, um Privatheit im vernetzten Zuhause situativ zu stabilisieren und Machtmissbrauch im Nahfeld wirksam vorzubeugen.

1. Einleitung

Das Zuhause gilt seit jeher als besonders geschützter Raum der Privatheit. Mit der zunehmenden Verbreitung von *Smart-Home*-Technologien gerät diese Sphäre unter Druck. *Smart-Home*-Systeme basieren ganz wesentlich auf Sensoren und der Erfassung verschiedener Umweltparameter. Solche Sensoren, die u. a. Temperatur, Luftfeuchtigkeit, Helligkeit oder Bewegung erfassen, erscheinen zunächst trivial und technisch unproblematisch, da sie auf den ersten Blick keine personenbezogenen Informationen liefern. Gleichwohl ermöglichen gerade diese scheinbar simplen Daten Rückschlüsse auf An- und Abwesenheiten, Routinen und Aktivitäten der Haushaltsmitglieder, insbesondere wenn sie im Zusammenspiel mit alltäglichen Praktiken interpretiert werden. Daten, die technisch neutral wirken, erhalten durch ihre soziale Einbettung eine privatheitsrelevante Bedeutung und können normative Bewertungen, Kontrolle oder Konflikte im Alltag hervorrufen (Fischer u. a. 2016; Morgner u. a. 2017; Tolmie u. a. 2016).

Eine exemplarische Episode verdeutlicht die Risiken solcher Datenpraktiken: Eine Teilnehmerin einer unserer Studien (s. u.) nutzte Sensordaten, um die Angaben ihres Partners zu überprüfen, der behauptet hatte, direkt nach der Arbeit in den Garten gegangen zu sein. Die Sensoren zeigten jedoch, dass er die Wohnung nicht verlassen hatte. Die Teilnehmerin konfrontierte ihn mit der Diskrepanz, ohne zu erklären, wie sie die Informationen erlangt hatte, wodurch der asymmetrische Informationszugang wirksam wurde. Dieses Beispiel illustriert, wie selbst einfache Sensordaten laterale Überwachung, d. h. die Überwachung eines Haushaltsmitglieds durch ein anderes, ermöglichen und Fehlinterpretationen hervorrufen können, die Privatheit und Sicherheit im Zuhause direkt beeinflussen (Kurze u. a. 2020; Richter u. a. 2018).

Privatheit muss im *Smart Home* daher als sozialer Aushandlungsraum verstanden werden, in dem verschiedene Akteure, wie Bewohner:innen, Besucher:innen, Vermieter:innen, Hersteller und Plattformen bzw. *Service-provider*, interagieren. In diesen komplizierten Akteurskonstellationen manifestieren sich Informationsasymmetrien und Konflikte, die vor allem über Kontrolle über Sensoren und den Zugang zu Daten verhandelt werden. Ins-

besondere Nutzer:innen, die Daten erzeugen, ohne Zugriff auf die *Smart-Home*-Daten zu erhalten, laufen Gefahr, in ihren Handlungsspielräumen eingeschränkt, überwacht oder sogar Opfer von Gewalt und Straftaten aus dem Nahfeld zu werden. *Smart-Home*-Systeme und die darin verbauten Sensoren verstärken und institutionalisieren somit soziale Machtverhältnisse im Zuhause.

Das Verbundprojekt *Simplications*¹ untersucht vor diesem Hintergrund die gesellschaftlichen und technischen Implikationen vermeintlich einfacher Sensordaten, wie sie häufig in *Smart-Home*-Systemen verbaut sind, deren Nutzung nicht intendierte bis potenziell „böartige“ Effekte entfalten kann (Kurze u. a. 2020; Kurze/Bischof 2021). Diese negativen Effekte entstehen nicht allein aus technischen Funktionsweisen oder ökonomischen Interessen „von außen“ (Saqib u. a. 2025), sondern aus der Verflechtung von Sensortechnologie, Nutzungspraxis, sozialen Beziehungen und interpretativen Praktiken. *Simplications* verfolgt zu deren Rekonstruktion einen interdisziplinären Ansatz, der Soziologie, *Human-Computer-Interaction* (HCI), Design und Medienbildung miteinander verknüpft. Von Interesse sind dabei folgende Fragen:

- Welche konkreten Konflikte und Informationsasymmetrien entstehen durch *Smart-Home*-Datenpraktiken innerhalb von Haushalten?
- Welche Konsequenzen bzw. *Implications for Use* ergeben sich daraus für eine informierte, souveräne Nutzung durch verschiedene Typen von Nutzer:innen?
- Welche *Implications for Design* ergeben sich für die bessere Ausgestaltung von diesen und ähnlichen Technologien?

Das Studiendesign verfolgt dabei einen Ansatz von *Privacy by Co-Design* und fragt gemeinsam mit Nutzer:innen danach, wie aus der Perspektive der tatsächlichen Anwendungen Perspektiven zur Ergänzung von *Privacy by Design* entwickelt werden können². Die gewonnenen Erkenntnisse fließen in die Entwicklung mediendidaktischer Formate ein, sodass technische Potenziale und soziale Praktiken gleichzeitig adressiert werden.

1 Das Projekt wurde vom BMFTR im Rahmen der Plattform Privatheit/ Vernetzung und Sicherheit digitaler Systeme gefördert (FKZ 16KIS1868K, Laufzeit: 06.2023 bis 12.2026). Projektpartner: Technische Universität Chemnitz, Hochschule Anhalt, Verbraucherzentrale Sachsen e. V., Leipzig.

2 Der Unterschied zwischen *Privacy by Design* und *Privacy by Co-Design* besteht in erster Linie darin, die Nutzenden in die Technikentwicklung einzubeziehen und nicht nur die *Developer*, vgl. Abschnitt 2, Forschungsstand.

Im vorliegenden Beitrag werden zunächst zentrale Forschungsbefunde zum Thema *Smart Home* und Privatheit mit Relevanz für die genannten Fragestellungen vorgestellt (2). Darauf aufbauend erfolgt die Erläuterung der methodischen Umsetzung der Studie. Es wird beschrieben, wie durch ein *Sensorkit* zur Erhebung einfacher Daten, Anwendungspraxen und idealen Einsatz- und Designvorstellungen in partizipativen Feldstudien erhoben wurden (3). Anschließend werden die Ergebnisse der Studie gegliedert nach vier zentralen Dimensionen präsentiert. Zunächst wird dargestellt, wie Sensordaten von Nutzer:innen interpretiert werden und welche Inferenzpotenziale sich daraus ergeben. Darauf folgt die Analyse, wie und wofür Daten als Ressource für soziale Kontrolle innerhalb von Haushalten genutzt werden. Im Anschluss werden Akteurskonstellationen betrachtet und die damit verbundene Diffusion von Verantwortung analysiert, bevor die Rolle von Privatheitsvorstellungen und des Problembewusstseins für die Nutzung und Reflexion von Sensordaten diskutiert wird (4). Aufbauend auf diesen Befunden erfolgt anschließend die Ableitung der Implikationen für eine informierte Anwendung und Design. Diese umfassen Empfehlungen für die Gestaltung technischer Systeme, Hinweise für einen informierten Einsatz im Alltag, Ansätze für Transfer und Intervention sowie möglicher Regulierungsansätze (5). Den Abschluss bildet eine Zusammenfassung der zentralen Erkenntnisse, in der die praktische Relevanz reflektiert und gezeigt wird, wie Technik, Nutzungspraxis und Bildung zusammenwirken, um die Privatheit im *Smart Home* relational und situativ zu stabilisieren (6).

2. Forschungsstand

Der aktuelle Forschungsdiskurs in der HCI zeigt, dass Privatheit im *Smart Home* zunehmend als situatives und relationales Phänomen verstanden werden muss. Frühere Studien konzentrierten sich vor allem auf abstrakte Datenschutzfragen und individuelle Einstellungen, sogenannte Datenschutzbedenken (*privacy concerns*) und die Identifikation von Nutzer:innen-Typen. Jüngere Studien diversifizieren zunehmend Typologien von Nutzer:innen und Abstufungen von Betroffenheit und Einfluss (Saqib u. a. 2025). Die Autor:innen unterscheiden beispielsweise zwischen Eigentümer, Hauptnutzer (*primary user*) und Beteiligten (*bystander*). *Bystander* werden weiter unterteilt in Besuchende, wie Gäste und Hausangestellte, im Haushalt lebende Personen wie Partner:in und Kinder und nicht involvierte Umstehende wie Nachbarn. Studien, die sich mit Genderfragen im

Smart Home befassen, stellen zudem die spezifischen Schwachstellen von Forschung und Entwicklung heraus, die ihren Fokus ausschließlich auf anschaffende und initialisierende Nutzer:innen – die Hauptnutzer:innen – legen (Lopez-Neira u. a. 2019, Schulz u. a. 2023). Während der Begriff der *bystander* zuletzt öfter aufgegriffen wurde, ist die Bezeichnung „sekundäre Nutzer:innen“ passender, da er den Grad der Betroffenheit von Haushaltsmitgliedern und Besuchenden, sowie deren Rolle für (oftmals nicht stattfindende) Aushandlungsprozesse aktiver setzt (Schulz u. a. 2023). Das problematische Verhältnis aus Betroffenheit bei mangelndem Bewusstsein für problematische Implikationen von *Smart-Home*-Systemen für die eigene Rolle bilden Studien ab, die (falsche) mentale Modelle von Nutzer:innen als Gefahr für Privatheit markieren (Marky u. a. 2021). Auf Grundlage einer Literaturübersicht identifizierten Saqib u. a. (2025) zehn spezifische Datenschutzbedenken sekundärer Nutzer:innen. Als häufigste Befürchtungen wurden bewusste Geheimhaltung, mangelndes Bewusstsein über die Implikationen der Daten, fehlende oder eingeschränkte Kontrolle, fehlende Einwilligung, Datenmissbrauch, auch für Belästigung oder diskriminierende oder unfaire Behandlung, Manipulation von Werten, die Datenverarbeitung durch Hersteller, sowie externe Bedrohungen rekonstruiert (ebd.).

Die HCI-Forschung zu *Data Legibility* und *Sense-Making* innerhalb von Haushalten zeigt, dass Menschen aus Sensordaten Rückschlüsse auf Ereignisse ziehen, die über den unmittelbaren Messzweck hinausgehen – und dass auch Fehlinterpretationen dabei eine zentrale privatheitsrelevante Rolle spielen (Fischer u. a. 2016; Kurze u. a. 2020; Tolmie u. a. 2016). Die oben angesprochenen Bedenken der sekundären Nutzer:innen sind also durchaus berechtigt. Selbst vermeintlich einfache Sensoren erzeugen komplexe Inferenzpotenziale. Bewegungs-, Temperatur- oder Luftfeuchtigkeitsdaten erlauben Rückschlüsse auf Anwesenheit, Schlafrhythmen, Alltagsaktivitäten oder Besuchsmuster. Forschungsarbeiten zum Einsatz und zur Nutzung von Sensoren zeigen, dass diese Daten jenseits ihres Messzwecks nutzbar sind und durch Kontextwissen und moralische und normative Ordnungen im Haushalt erst Bedeutung erhalten (Laput u. a. 2017). Diese Bedeutungszuweisungen sind entscheidend für Konflikte und Informationsasymmetrien im häuslichen Alltag. Die Interpretation vermeintlich einfacher Daten ist also nicht technisch determiniert, sondern entsteht im Zusammenspiel von Alltagspraxis, sozialen Normen und individuellen Deutungsmustern (Kurze/Bischof 2021; Morgner u. a. 2017).

Ein weiterer Schwerpunkt der Forschung an der Schnittstelle von Soziologie, *Gender Studies* und HCI liegt auf Macht, Kontrolle und potenziel-

ler digital vermittelter Gewalt im zunehmend vernetzten Zuhause (Lopez-Neira u. a. 2019; Tanczer 2021). Die Sensordaten ermöglichen laterale Überwachung und die subtile Kontrolle von Haushaltsmitgliedern. Die Forschung zu *Gender* und *Internet of Things* zeigt, dass insbesondere sekundäre und vermeintlich passive Nutzer:innen, die Daten erzeugen, ohne in die Nutzung eingebunden zu sein, eingeschränkte Handlungsspielräume haben und potenziell kontrolliert werden können (Marky u. a. 2020; Richter u. a. 2018). Fälle häuslicher Gewalt oder zwanghafte Kontrolle zeigen die Tragweite solcher daten-vermittelter Informationsasymmetrie, die zur Machtressource wird: *Smarte* Geräte eignen sich insbesondere für methodische psychische Gewalt wie *Gaslighting* (Tanczer 2021). Weniger intentional hat soziologische Forschung zu Subjektivierung durch und mit Technik gezeigt, dass intime Überwachung zur Normalität wird, so dass z. B. Kinder und deren Nannys selbstverständlich überwacht werden (Dereymaeker u. a. 2024; Lupton 2016; Mascheroni 2020; Rogers u. a. 2023). Eltern, Partner:innen oder technisch versiertere Haushaltsmitglieder geraten in die Position des permanenten Beobachtens. Das Zuhause wird als Schutzraum definiert, in dem vernetzte Technik den Nutzer:innen eine Ressource bieten, um persönliche Sicherheit durch Überwachung herzustellen – die „sorgende Konsumentin“ wird zur Leitfigur und Überwachung als Sorgepraxis verstanden (Stark/Levy 2018; Mascheroni 2020). So entstehen gerade im Familienalltag und ohne intentional bössartigen Einsatz von Sensorik Grauzonen, in denen Komfort, Kontrolle und normative Bewertungen von Verhalten miteinander verknüpft werden. *Simplications* richtet den Fokus genau auf diese Grenzbereiche zwischen unbedarften Einsatz und problematischer Anwendung.

Die Unterstützung informierter Nutzung und differenzierter Einwilligungsmöglichkeiten hat sich als weiteres Diskursfeld zwischen Verbrauch erforschung und HCI etabliert. So wurden bereits verschiedene Varianten zur Sichtbarmachung von Privatheitsrisiken vorgeschlagen. *Privacy Labels*, bzw. *Label* zur Produktbewertung, enthalten häufig zumindest eine *Privacy*-Komponente (u. a. Caven u. a. 2024; Garg 2021; Ramokapane u. a. 2024; Reinhardt/Railean 2020). Daneben gibt es *community*-basierte bzw. *Co-Design*- Ansätze (Adryan 2015), wie auch das *Open IoTmark* (später *Better IoTmark*) von Deschamps-Sonsino (2018) oder das *ThingsCon/Mozilla Open IoT Studio „Trustmark for IoT“* (Bihr 2017). All diesen Vorschlägen ist gemeinsam, dass sie zwar Nutzer:innen- und Design-Perspektiven einbinden, aber bisher nicht über das Stadium eines konzeptionellen Vorschlags

hinausgekommen sind, und noch keinen Eingang in Produkte und Services rund um das *Smart Home* gefunden haben.

Partizipative Forschung und datenbasierte Reflexion haben sich als zentrale methodische Ansätze vor allem in der HCI erwiesen, um die privatsphärenrelevanten Implikationen von *Smart-Home*-Sensordaten sichtbar und erfahrbar zu machen und auch *IoT* bzw. *Privacy Labels* zu gestalten. Im Kontext von *Privacy-by-Co-Design* werden Nutzer:innen aktiv in die Interpretation eigener Daten einbezogen, wodurch nicht nur Problembewusstsein und Reflexion gefördert werden, sondern auch normative Annahmen über angemessenes Verhalten und Kontrolle im Haushalt aufgedeckt werden. Methoden wie *Design Fiction*, *Critical Design* oder praxisnahe Bildungsformate zeigen, dass die unmittelbare Erfahrung von Daten, ihre Visualisierung, Annotation und die gezielte Irritation durch widersprüchliche Szenarien wirksamer sind als rein kognitive Aufklärung. Solche Interventionen eröffnen Räume, in denen Nutzer:innen ihre mentalen Modelle prüfen, die Wahrnehmung von Risiken und Machtbeziehungen reflektieren und Rückschlüsse für die eigene Praxis ziehen können. Studien im HCI- und Design-Kontext zeigen, dass solche Formate nicht nur die Interpretation und Sensibilisierung erleichtern, sondern auch konkrete Hinweise für Gestaltung, *Usability* und mediendidaktische Anwendungen liefern (Köpferl u. a. 2024; Mucha u. a. 2022). *Co-Design* für Privatheit besteht dabei zumeist auf der Entwicklung von gewünschten oder besser passenden Anwendungsszenarien und Interfaces, wobei die Teilnehmenden in der Regel Verbraucher:innen sind (Leitão 2019; Yao u. a. 2019; Jin u. a. 2022; Hahn u. a. 2024; Zufferey u. a. 2024; Wang u. a. 2025). Forschung, die Privatsphäre-Implikationen partizipativ untersucht, zeigt aber, dass das Bewusstsein für problematische Daten und Konstellationen sowie der eigenen Betroffenheit auch in solchen Kontexten nicht allen Studienteilnehmer:innen klar ist. Vielmehr zeigen unsere Vorarbeiten, dass die Verquickung aus technischer Datensammlung, Wissen über Haushaltsaktivitäten und eine intersubjektive Rekonstruktion in der Gruppe ein informiertes *Privacy-by-Co-Design* ermöglichen (Kurze u. a. 2020, 2025).

Insgesamt verdeutlicht die Forschung die Notwendigkeit eines integrativen Ansatzes, der technische Gestaltung, Nutzungspraxis und partizipative Reflexion kombiniert. Rein technische Lösungen reichen nicht aus, um Privatheit im *Smart Home* zu sichern, da Daten immer relational und situativ interpretiert werden. Bisher beziehen sich *Co-Design Workshops* auf den Gebrauch von Kameras und *Smart Speaker* und nicht auf die vermeintlich weniger privatsphärenrelevante Sensorik des *Smart Homes* (Becker u. a.

2025; Saqib u. a. 2025). Diese Erkenntnisse bzw. Forschungslücken bilden die Grundlage für die Konzeption unseres Studiendesigns, welches wir seit 2017 in mehreren Iterationen entwickelt haben, wie wir im nächsten Abschnitt erläutern.

3. Erforschbarkeit von Datenpraktiken im Smart Home: *Sensorkit* und Studiendesign

Für die empirische Forschung ergeben sich durch *Smart-Home*-Systeme methodische Herausforderungen. Es reicht nicht aus, technische Systeme isoliert auf ihr Gefahrenpotenzial zu analysieren oder hypothetische Szenarien abzufragen. Um die sozialen Implikationen von Sensordaten aus dem Zuhause zu untersuchen, müssen Datenpraktiken im tatsächlichen häuslichen Alltag erhoben werden und zugleich Möglichkeiten geschaffen werden, die Interpretationen und Aushandlungen der Nutzer:innen sichtbar machen zu können. Vor diesem Hintergrund wurde ein Forschungsansatz entwickelt, der Datenerhebung, Visualisierung und partizipative Analyse miteinander verbindet.

Kommerzielle *Smart-Home*-Systeme sind hierfür nur eingeschränkt geeignet. Rohdaten sind meist nicht ohne Weiteres zugänglich, sie werden proprietär und an externe Server übermittelt, wodurch forschungsethische Risiken entstehen. Um diese Abhängigkeiten zu vermeiden, wurde eine eigenständige Forschungsinfrastruktur entwickelt, die Erhebung, Visualisierung und Reflexion von Sensordaten in einem geschlossenen System vereint und vollständig unter Kontrolle der Teilnehmenden lässt.

Das Forschungsdesign zielt auf die Erhebung von diversen Konstellationen und Perspektiven, die in der Forschung noch selten empirisch erhoben wurden.

3.1 Erhebungsinstrument *Sensorkit*

Zentrales Instrument der Forschungen im Projektkontext ist ein *Sensorkit*, ein kleines, einfach zu installierendes Set aus vernetzten, batteriebetriebenen technisch einfachen Sensoren, das über rund zwei Wochen in Haushalten eingesetzt wird. Das *Sensorkit* dient dazu, typische Umweltdaten im Haushalt kontinuierlich zu erfassen und sichtbar zu machen. Die Sensoren erfassen Temperatur, Luftfeuchtigkeit, Helligkeit, Bewegung, Luftqualität

und Lautstärke. Dies wurde in mehreren Studien erprobt, angewendet und stetig weiterentwickelt (Kurze 2022; Kurze/Bischof 2021; Kurze/Reuter 2024; Kurze u. a. 2020; Kurze u. a. 2026; Köpferl u. a. 2024; Kurze u. a. 2025; Reuter 2025; Richter u. a. 2018). Die Auswahl der Messgrößen orientiert sich an den typischen im *Smart-Home*-System verbauten Komponenten, die zudem typischerweise als wenig intrusiv wahrgenommen werden – Anders als Mikrofone und Kameras, wie im bereits gut informierten Diskurs über so genannte *Smart Speaker*.

Die Visualisierung der erhobenen Daten erfolgt lokal auf einem Tablet in Form von Zeitreihen, die Veränderungen über den Tagesverlauf abbilden. Die Darstellung ist so gestaltet, dass auch weniger technisch versierte Personen Muster und Veränderungen im Alltag leicht erkennen können. Alle Daten verbleiben im Haushalt, können jederzeit eingesehen oder gelöscht werden, und es findet keine externe Datenübertragung statt. Damit wird sowohl Datensouveränität gewährleistet als auch die Reproduktion von proprietären Überwachungspraktiken verhindert (Kurze 2022; Kurze u. a. 2025). Abb. 1 zeigt das *Sensorkit* mit seinen verschiedenen technischen Elementen.



Abbildung 1: *Sensorkit* (eigenes Foto).

Die Teilnehmenden der Studien wurden durch verschiedene, an *Cultural Probes* orientierten Aufgaben (Berger u. a. 2018) schrittweise an die Datenerfassung, die Beobachtung der Daten am Tablet, und Möglichkeiten der Verknüpfung von Ereignissen und Daten herangeführt. Die Sensoren des *Sensorkits* lassen sich dabei flexibel im Haushalt platzieren, etwa in Wohnräumen oder in der Nähe von Fenstern oder Heizkörpern, um typische Alltagsbedingungen zu erfassen. Zu Beginn der Studie wurden die Sensoren in allen Haushalten zusammen mit den Teilnehmenden an denselben Orten – Eingangstür, Küche, Wohnzimmer – platziert. Dazu fanden Aufbaugespräche statt, in denen die Funktionsweise, Erwartungen und eventuell vorhandene Bedenken hinsichtlich des Einsatzes von Sensoren und *Smart Home* generell thematisiert werden. Während der Feldphase bearbeiteten die Teilnehmenden täglich kurze Impulsaufgaben zur Selbstbeobachtung bzw. Interpretation der eigenen Daten, wie die Beschreibung von Mustern, die Formulierung von Ursachenhypothesen oder die Identifikation von Abweichungen von Routinen.

Ziel dieser Aufgaben ist nicht die ‚richtige‘ Interpretation, sondern die Förderung von Aufmerksamkeit und Bewusstsein für die Aussagekraft der Sensordaten und Reflexion darüber, was über den eigenen Haushalt durch die Datenmitschnitte erfahrbar wird. Das *Sensorkit* übernimmt dabei lediglich eine Messfunktion und ist nicht mit anderen Objekten im Haushalt verknüpft. Mit zunehmender Vertrautheit wird der Grad der Strukturierung reduziert, um eigene Fragestellungen der Teilnehmenden zu ermöglichen. Der *Privacy-by-Co-Design*-Aspekt in dieser Phase der Studie besteht darin, dass die Teilnehmenden selbst entscheiden können, wo, wofür und wie sie die Sensoren einsetzen – Was durchaus problematische Nutzungsweisen hervorgebracht hat (*siehe unten*). In der Nutzung der Teilnehmenden werden jedoch Einsatzgebiete wie der Abgleich mit anderen, eigenen *Smart-Home*-Sensoren bzw. Systemen oder das Abbilden von Heiz- und Lüftungsroutinen ersichtlich, so dass in der Nutzung Funktionalität entstehen kann.

3.2 Erhebungsinstrument Interviews und Gruppendiskussion

Ergänzend fanden am Ende der Datenerhebung mittels *Sensorkit* qualitative Interviews und *Debriefings* statt, um die Wahrnehmung, Bedenken und Erfahrungen der Teilnehmer:innen zu erheben. In anschließenden Gruppendiskussionen wurden anonymisierte Datenausschnitte mit Teilnehmenden aus jeweils zwei bis drei Haushalten gemeinsam analysiert bzw. disku-

tiert. Dies knüpft direkt an das „*Guess-the-Data-Verfahren*“ an (Kurze u. a. 2020), mit dem gezeigt werden konnte, dass Personen aus einfachen *Smart Home*-Sensordaten nicht nur eigene Routinen, sondern auch das Verhalten anderer Haushaltsmitglieder und Haushalten rekonstruieren können. Zusätzlich macht das Format kollektive Deutungsmuster sichtbar und zeigt, wie stark Interpretationen vom jeweiligen Alltagswissen abhängen. Nur über diese Praxisnähe werden nicht-intendierte Nutzungen, Informationsasymmetrien, Konflikte, Missbrauchspotenziale und Fehlinterpretationen sichtbar. Die Teilnehmenden fungieren so nicht nur als ‚Datenlieferant:innen‘, sondern als doppelte Ko-Interpret:innen der eigenen Daten – sowohl in der individuellen Exploration des *Sensorkits* als auch im kollektiven Deuten im „Daten-Raten“. Indem die Methode die Teilnehmenden dazu anregt, eigene und fremde Daten zu interpretieren, wird implizites Alltagswissen über vermeintlich einfache und nicht datenschutzrelevante Sensorwerte sichtbar.

3.3 Sampling

Das Sampling der Studien folgte einer Minimal-/Maximalkontrast-Logik. In die jüngste Durchführung, die wir hier berichten, wurden Haushalte mit unterschiedlichen sozialen Konstellationen, Technikaffinität und Wohnformen einbezogen. Es nahmen sowohl Einpersonenhaushalte, Paare, Familien mit Kindern als auch Wohngemeinschaften teil. Miet- und Eigentumsverhältnisse wurden ebenfalls berücksichtigt, da sie Einfluss auf Kontrolle und Verantwortungszuschreibung haben können^{3,4}. Die Teilnehmenden wurden über Zeitungsanzeigen, digitale Verteiler und soziale Medien zur Teilnahme aufgefordert. Die aktuelle Studie fand im Raum Chemnitz und Umgebung statt, umfasste 18 Haushalte und wurde in zwei Feldphasen im Frühjahr 2024 (F1) und Frühjahr 2025 (F2) erhoben.

Das Sample der jüngsten Studie besteht insgesamt aus acht Paaren über 60 Jahren, zwei Müttern mit Kindern und Partnern, fünf alleinlebenden Männern im mittleren Alter (teilweise verpartnert), einer WG aus zwei

3 In der Erhebung zeigt sich die Präsenz von *Smart-Home*-Systemen mit ähnlicher Sensorik bspw. zur Steuerung des Energieverbrauchs besonders bei Eigenheimbesitzer:innen.

4 In Mietwohnungen wurde der Einsatz solcher Systeme tendenziell als zu aufwendig, teilweise als Luxus verstanden. In diesen ergibt sich außerdem zunehmend die Problematik des ‚unfreiwilligen‘ Einsatzes von Sensorik (Wolfangel 2024).

männlichen Personen unter 30 Jahren, einem jungen Paar unter 30 Jahren, sowie einer alleinlebenden Frau über 60 Jahren. Bei den älteren Paaren (+60 Jahre) findet sich vermehrt die Konstellation eines männlichen Partners, der über eine berufliche Nähe zur IT-Branche verfügt und bereits ein eigenes *Smart-Home*-System betreibt. In 12 der 18 Haushalte waren *smarte* Geräte wie Alexa, Thermomix bis hin zu komplexen *Smart-Home*-Systemen vorhanden. Das spricht für einen Selbstselektionseffekt anhand des Interesses an *Smart-Home*-Technik, was sich für das Forschungsinteresse zu Bewusstsein und Reflexion über Implikationen der Nutzung als wertvoll erwies.

3.4 Angewendete Auswertungsmethoden

Die Auswertung der Daten erfolgte interdisziplinär und im Methodenmix und jeweils angepasst auf die Form der vorliegenden Daten. So wurden bspw. die Annotationen der Teilnehmenden separat und inhaltsanalytisch ausgewertet (Reuter/Kurze 2025). Die während der Installation und beim Abbau des *Sensorkits* erhobenen Interviews und Gruppendiskussionen wurden transkribiert und interpretativ-rekonstruktiv sowie sequenzanalytisch in Gruppeninterpretationen in der interdisziplinären Forscher:innen-Gruppe ausgewertet. Das Vorgehen orientiert sich an der integrativen Auswertungsmethode von Kruse (2015), die verschiedenen Strategien qualitativer-rekonstruktiver Verfahren vereint und sich daher für die unterschiedlichen Daten (problemzentriertes Interview, Paarinterview, Gruppendiskussionen) besonders eignet.

3.5 Methodisch-ethische Reflexion und Limitationen

Die Untersuchung wurde iterativ durchgeführt. Nach dem ersten Feldstudienzyklus erfolgte eine Auswertung sowie eine Anpassung des Systems und der Aufgabenstellungen, bevor der zweite Feldstudienzyklus durchgeführt wurde. Dadurch konnten insbesondere technische Verbesserungen und zuvor bestehende Verständnishürden berücksichtigt werden. Die Forschungsethik und das bewusste Risikomanagement bilden einen wichtigen Rahmen für den gewählten Ansatz. Ziel ist es, Privatheitsrisiken den Teilnehmenden sichtbar zu machen, ohne sie in der Studie zu erzeugen oder zu verschärfen. Dazu werden Datenminimierung, Anonymisierung und lokale Speicherung

konsequent umgesetzt. Teilnehmende behalten jederzeit Kontrolle über ihre Daten; einschließlich der Möglichkeit zur Löschung. Eine audiovisuelle Erfassung wurde bewusst vermieden. Alle Haushaltsmitglieder und Besucher:innen werden in den Prozess einbezogen, um die informierte Einwilligung und den Schutz aller möglichen Datenlieferant:innen sicherzustellen. Auf diese Weise werden ethische Prinzipien und Designentscheidungen eng miteinander verknüpft, sodass der wissenschaftliche Erkenntnisgewinn und der Schutz der Teilnehmenden zugleich gewährleistet sind.

Insgesamt kombiniert der methodische Ansatz technische Datenerhebung, soziale Interpretation und partizipative Reflexion. Er ermöglicht es, Praktiken sichtbar zu machen, in denen Sensordaten zu sozialen Ressourcen werden, Wissen erzeugen, Entscheidungen legitimieren und Machtverhältnisse mitstrukturieren. Erst durch die Verbindung von kontinuierlicher Datenerhebung, individueller und kollektiver Interpretation sowie qualitativer Auswertung lassen sich die Dynamiken rekonstruieren, in denen aus einfachen Sensordaten soziale Wirklichkeit entsteht. Dieses Vorgehen bildet die Grundlage für die Darstellung der Ergebnisse im nächsten Kapitel, in dem Nutzung, Bedeutung und Konflikte rund um die Sensordaten analysiert werden.

Durch forschungspraktische und explorative Gründe weist die Studie mehrere Limitationen auf: Die Erhebung in 18 Haushalten im Raum Chemnitz bildet lokale soziokulturelle Strukturen ab und schränkt die Generalisierbarkeit ein. Zudem führt die Rekrutierung über soziale Medien zu einem selektiven Sample, wodurch bestimmte Gruppen (z. B. nicht deutschsprachige oder konfliktreiche Haushalte) unterrepräsentiert sind. Das Studiendesign erzeugt außerdem Interventionseffekte, da das *Sensorkit* zur Reflexion anregt und damit Daten aktiv mitgestaltet werden; dies entspricht zwar dem *Co-Design*-Ansatz, kann aber auch unbeabsichtigte Effekte haben, etwa die Anregung neuer Überwachungspraktiken. Teilnehmende interpretierten die Sensoren teils als Grundlage für Produktentwicklungen statt als Reflexionsinstrument. Zusätzlich ist die Vergleichbarkeit mit proprietären und *Open Source* basierten *Smart-Home*-Systemen bewusst eingeschränkt, da reale Nutzungsdaten schwer zugänglich und anders verwendet werden. Trotz dieser Einschränkungen bietet der Ansatz wertvolle Einblicke in soziale Praktiken, Verantwortlichkeiten und Normalitätskonstruktionen in vernetzten Haushalten mit Fokus auf sozial verankerte Sinnstrukturen statt universeller Verallgemeinerung.

4. Aus Sensordaten werden soziale Wirklichkeiten

Unsere Studie zeigt, wie vermeintlich einfache Sensordaten interpretiert, mit Routinen verknüpft und zur Begründung von Entscheidungen herangezogen werden. Auf diese Weise entstehen aus Messwerten soziale Bedeutungen, die Konflikte, Machtverschiebungen und Verantwortlichkeiten erzeugen. Wie das geschieht, zeigen wir im Folgenden entlang vier zentraler Dimensionen dieser Interpretations- und Aushandlungsprozesse: Interpretierbarkeit, Datenpraktiken als Machtpraktiken, Akteurskonstellationen sowie mentale Modelle und *Awareness* (Sensibilisierung). Wir nutzen vornehmlich Beispiele aus einer von sechs Gruppendiskussionen zur Veranschaulichung⁵.

4.1 Interpretierbarkeit: Aus Daten werden Geschichten

Die Studie zeigt, dass die Teilnehmenden aus vermeintlich einfachen Sensorwerten vielfältige Rückschlüsse über Alltag, Routinen und Interaktionen sowohl in ihrem als auch in anderen Haushalten ziehen können. Sie zeigt nicht nur, dass die Teilnehmer:innen anhand von Daten, die z. B. Türöffnungen anzeigen, oder Änderungen von Temperatur und Luftfeuchtigkeit, prinzipiell in der Lage sind, diese mit zeitlichen Mustern und bekannten Gewohnheiten zu verknüpfen. Unsere Ergebnisse zeigen auch, dass Daten aus dem Zuhause immer mit alltagsweltlichen Kategorien wie Anwesenheit, Schlaf- und Arbeitsrhythmen oder Besuchssituationen korreliert werden. Die Sensordaten werden durch die Teilnehmenden standardmäßig *nicht* als technische Werte, sondern als Ausdruck von alltagspraktischen Vollzügen interpretiert. Das wird besonders deutlich, wenn Abweichungen von erwarteten Mustern nicht als technische Artefakte oder Messfehler interpretiert werden, sondern als Indikatoren für unerwartetes, erklärungsbedürftiges Verhalten. Die Teilnehmer:innen haben stets den Messwerten geglaubt. So fragt z. B. eine Teilnehmerin den Partner scherzhaft, ob er sich wohl rausgeschlichen habe, denn nur so wäre die nächtliche Bewegung in der Küche zu erklären. Eine wesentliche privatheitsrelevante Implikation von Sensordaten aus dem Zuhause entsteht demnach durch ihre Interpretierbarkeit auf Basis von eigenem alltagsweltlichem Wissen, und einer Normalitätser-

5 In den Gruppendiskussionen im *Guess-the-Data*-Verfahren wissen die Teilnehmenden nicht, aus welchem Haushalt die Daten stammen. Sie raten und plausibilisieren daher, ob bzw. in welchem Kontext die vorliegenden Daten entstanden sein könnten.

wartung, dass angezeigte Daten direkter Ausdruck von zuvor vorgängigem Verhalten sind.

4.2 Datenpraktiken als Machtpraktiken

In diesem Zusammenhang liegt es nahe, dass Daten im Haushalt eine geradezu autoritative Rolle einnehmen: Sie stützen Vermutungen, schaffen Legitimation für Handlungen und Bewertungen, aber schaffen auch Konfrontationen zwischen Haushaltsmitgliedern. Sensordaten können im Haushalt als Instrumente von Kontrolle und Normdurchsetzung eingesetzt werden. Dies zeigte sich in einem Haushalt beim Lüften und der Kühl-schranköffnung. So unterließen es mehrere Teilnehmerinnen, an den Kühl-schrank zu gehen, oder waren sich bewusst, dass nun sichtbar ist, wie kurz der Spaziergang mit dem Hund ausfällt oder wie gelüftet wurde. Im Studienkontext erfolgte eine Kontrolle von Verhalten nie mit schädigender Absicht. Dennoch zeigte sich auch in unserer Feldforschung, dass der Zugriff auf Daten bzw. Datenausschnitte laterale Überwachung innerhalb sozialer Nahbeziehungen, konkret zwischen Partner:innen, Eltern und Kindern und in Wohngemeinschaften, ermöglicht hat. Ein Kind explizierte dies und bewertete die Erhebungsphase als dauerhafte Überwachung durch die Mutter im Privatraum Kinderzimmer. *Smart-Home*-Daten dienen nicht nur der Beobachtung der Umwelt oder der Selbstbeobachtung, sondern auch der Beobachtung Dritter, etwa, wenn damit die nächtliche Geräuschkulisse im Hausflur eines Mietshauses erfasst wurde. Dass dies nicht unproblematisch ist, wurde darüber sichtbar, dass ein Teilnehmer sich aktiv darum bemühte, dies so klandestin wie möglich zu gestalten.

Innerhalb von Haushalten und in den Gruppendiskussionen wurden die Sensordaten regelmäßig als Grundlage für Verbesserungs- und Normierungspraktiken genutzt, z. B. in Bezug auf Heizen, Lüften oder Energieverbrauch. In den Gruppendiskussionen zeigt sich, wie sehr die Bewertungen dieser Aspekte normativ aufgeladen sind und Vorstellungen von Verantwortung, Fürsorge oder Nachhaltigkeit transportieren. So expliziert eine Teilnehmerin, sie hoffe, die Werte stammten nicht aus ihrem Haushalt, da ihr die Werte der schlechten Luftqualität peinlich wären. Heiz- und Lüftungsverhalten wurde regelmäßig als falsch beschrieben, wobei falsches Lüften und Heizen bspw. zu Schimmel führen könnten.

Diese Dimension der Implikationen zeigt, dass hinsichtlich Privatheit im *Smart Home* neue Herausforderungen entstehen. Es kann nicht nur darum

gehen, wie das vernetzte Zuhause *nach außen* abgesichert ist. Vielmehr muss neben dem kollektiven Erfordernis von Datenschutz (Véliz 2021) auch beachtet werden, wie Privatheit *innerhalb* von Familien- bzw. Haushaltsgefügen ausgestaltet wird. Privatheit gerät so als relationales, situiertes Geschehen – nicht als Individualpräferenz bzw. individuelle Abwehraufgabe in den Blick. Im Haushaltskontext gilt zu fragen: Wer darf Daten einsehen, wer interpretiert sie, wessen Deutungen gelten als maßgeblich und wessen Komfort wird berücksichtigt? Datenpraktiken können bestehende Machtasymmetrien verstärken, insbesondere wenn einzelne dauerhaft die Rolle der Beobachtenden übernehmen, während andere zu Objekten der Interpretation werden.

4.3 Akteurskonstellationen

Der Haushalt selbst darf nicht als abgeschlossene Einheit mit den immer gleichen Personen verstanden werden. Das vernetzte Zuhause beobachtet und speichert nicht nur die Daten von Personen, die sich für die Installation entschieden haben und sich der Implikationen der Nutzung (vermeintlich) bewusst sind. Neben den direkt beteiligten Haushaltsmitgliedern sind in unserer Studie eine Reihe von sekundären Nutzer:innen betroffen gewesen, u. a. waren Besucher:innen, Dienstleister:innen oder Handwerker:innen zeitweise in den Daten präsent. So rekonstruiert ein Paar in der ‚Daten-Raten‘ Sitzung genau, wann der Besuch gekommen ist, was er gemacht hat und wann dieser wieder gegangen ist. In diesem Kontext wird besonders deutlich, dass Privatheit im *Smart Home* nicht allein durch individuelle Zustimmung bzw. Ablehnung der Anschaffenden gewährleistet werden kann. Zudem wurden mehrfach Datenpraktiken über einzelne Haushalte hinaus berichtet. Häufig installierte ein:e technisch versierte Nutzer:in älteren Angehörigen *Smart-Home*-Geräte und wurde so zu einer Art Administrator aus der Ferne. Neben handlungspraktischen Implikationen, dass der Administrator bei Problemen eingreifen muss, ergaben sich daraus auch privatsphärerelevante Aushandlungen, wie z. B. um *smarte* Schließsysteme und die Frage, ob der Schwiegersohn dann einfach unangekündigt in der Wohnung stehen könne. Dabei wird die mögliche laterale Überwachung durch solche zugriffsberechtigten Haushaltsmitglieder zugunsten des Aspekts der Sicherheit im Notfall tendenziell ausgeblendet.

4.4 Mentale Modelle und Awareness

Unsere Studie zeigt, dass nicht-audiovisuelle Sensoren vor und während der Studie häufig als unproblematisch wahrgenommen werden. Dies zeigt zum Beispiel der Umgang mit der Messung der durchschnittlichen Lautstärke. Diese wird von den Teilnehmer:innen im Vergleich zur Sprachaufzeichnung als unproblematisch gerahmt. Paradox erscheint, dass es in den Haushalten, in denen die primären und sekundären Teilnehmenden Sorge vor Sprachaufzeichnungen im Studienkontext hatten, dennoch auch der Sprachassistent Amazon Alexa genutzt wurde, da dieser Teil der eigens geschaffenen häuslichen Komfortinfrastruktur war. Temperatur, Luftfeuchtigkeit oder Lautstärke erscheinen zunächst als harmlos, da sie keinen direkten Personenbezug aufweisen. Die Annahme der Harmlosigkeit von Umweltmessungen führt dazu, dass die Aussagekraft der Daten, die gerade aus der Überlagerung der einzelnen Messwerte entsteht, unterschätzt wird.

Ein avisiertes Ziel der Studie war, dass die Teilnehmer:innen durch die praktische Auseinandersetzung mit den eigenen Messwerten ein Bewusstsein dafür entwickeln, welche Rückschlüsse auf Anwesenheit, Routinen oder Aktivitäten möglich sind. Durch die Kombination aus individueller Exploration, strukturierten Aufgaben und kollektiver Interpretation in der Gruppendiskussion ‚Daten-Raten‘ wurde den Teilnehmenden deutlich, dass dieselben Werte unterschiedliche Bedeutungen annehmen können. Besonders anschaulich wird das in wiederkehrenden Diskussionen zur Rolle von Luftfeuchtigkeit und richtigem Lüften – und wie ein Vermieter im Vergleich zu einem Mieter diese Daten lesen würde. So zeigte sich im ‚Daten-Raten‘ bereits, dass die Teilnehmenden die Daten einer Person als Hinweis werten, dass bei den vorliegenden niedrigen Temperaturen bzw. Dauerlüften Schimmel vorprogrammiert sei. An dieser Stelle wurden die Implikationen von *Smart-Home*-Daten, insbesondere vor dem Hintergrund von durch Vermieter installierte Systeme, für die Teilnehmer:innen besonders nachvollziehbar.⁶

5. Implikationen für Gestaltung und Nutzung von Smart-Home-Daten

Die empirischen Befunde zeigen, dass Privatheitskonflikte im *Smart Home* nicht allein durch die eingesetzte Technik, sondern durch die konkrete

6 In der Erhebungsphase begann eine öffentliche Diskussion um den Einsatz von Sensorik in „smarten Rauchmeldern“ durch Vermieter (Wolfangel 2024).

Nutzung, institutionelle Rahmenbedingungen und gestalterische Entscheidungen entstehen. Daraus ergeben sich praktische Implikationen für Design, Nutzung, Transfer und Intervention sowie für die politische Rahmung und Regulierung.

5.1 Implikationen für die Gestaltung technischer Systeme

Smart-Home-Technologien können Privatheit nur dann wirksam gewährleisten, wenn ihre Gestaltung die sozialen Konstellationen innerhalb eines Haushalts berücksichtigt. Das betrifft vor allem die Ausgestaltung von Zugriffsrechten. Sowohl Nutzungsbedingungen, die Optionen zur Einwilligung in diese als auch Schnittstellen und Apps zur Steuerung differenzieren bislang nicht oder nicht ausreichend zwischen unterschiedlichen Nutzer:innen und deren Privatheitsbedürfnissen. Es gibt zur Steuerung von Geräten in der Regel ein Benutzerkonto. Das Vorhandensein weiterer Personen, die ebenfalls zustimmen müssten, wird technisch nicht beachtet. Dies gilt es genau andersherum zu gestalten. Unterschiedliche Rollen wie Erwachsene, Kinder oder temporäre Haushaltsmitglieder und Besucher:innen müssten in den *Interfaces* und Einstellungsmöglichkeiten berücksichtigt werden. Die Erfassung der Daten muss insgesamt nach außen sichtbarer werden, insbesondere für Personen, die sich nicht intensiv mit dem konkreten *Smart-Home*-System beschäftigen und über ihre Datenerfassung informiert sind. Zudem ließe sich bei der Datenerfassung eine Zweckbindung bzw. zeitliche Beschränkung der Datenerhebungen durch Gesetze oder Richtlinien zur Technikgestaltung einführen.

Hierbei kommt es wesentlich darauf an, ob das System auf *Open-Source*-Anwendungen basiert, oder kommerzielle Produkte zum Einsatz kommen. Generell sollten *Default*-Einstellungen die minimale Datenerfassung und lokale Speicherung priorisieren. Bei den meisten *Open-Source*-Systemen ist dies bereits der Fall bzw. oberste Priorität der primären Nutzenden. Dennoch sollten auch in diesen Fällen Möglichkeiten bereitstehen, um gerade sekundären Nutzer:innen mehr Handlungsspielräume zu eröffnen, ohne dass dafür ein Einstieg in die Programmebene der Systeme notwendig ist.

Funktionen wie ein temporäres Anhalten bzw. Stummschalten, eine einfache Möglichkeit zur Löschung von bestimmten Zeitabschnitten, eine einfache *exit-option* sowie Migration und Export von Daten eröffnen Handlungsspielräume, ohne die Nutzung zu unterbinden (Kurze u. a. 2025).

Gerade um sekundären Nutzer:innen den Zugang zu erleichtern, sollten Möglichkeiten zur Annotation, Kommentierung und Kontextualisierung bestehen, damit Nutzer:innen eigene Erklärungen, Irritationen oder alternative Deutungen einbringen und sichtbar machen können (Kurze/Reuter 2024; Reuter 2025; Reuter/Kurze 2025).

Festgelegte Transparenzmechanismen sollten für Dritte verständlich darstellen, welche Daten entstehen, wo sie verarbeitet werden und wer Zugriff darauf hat und welche Interpretationen und Rückschlüsse ggf. daraus möglich sind. Zugleich müssen *Interfaces* und Visualisierungen handhabbar sein und Aushandlungsprozesse erleichtern. Irritationen können produktiv gemacht und bestehende Machtasymmetrien als Ergebnis von Aushandlungen sichtbar werden, anstatt sie zu verschleiern. Erst wenn die technische Gestaltung sozial anschlussfähig ist, können die Schutzwirkungen der Systeme im Alltag realisiert werden.

5.2 Implikationen für die Nutzung

Privatheit wird nicht allein durch vernetzte Systeme, sondern wesentlich durch deren Nutzung und fehlende Aushandlungsprozesse innerhalb des Haushalts gefährdet. Entscheidungen über Sensorplatzierung, Datensichtbarkeit und -interpretation sollten offen kommuniziert, diskutiert und gemeinsam getroffen werden, sodass alle betroffenen Personen einbezogen sind. Praktische Checklisten oder Pfaddiagramme können dabei helfen, zentrale Fragen vor der Installation eines Sensors zu klären, zum Beispiel: Welche Daten werden wirklich benötigt? Wer ist davon betroffen? Welche Alternativen gibt es? Solche Reflexionen verwandeln technische in soziale Entscheidungen und fördern ein Bewusstsein für die Auswirkungen der Datenerhebung auf alle Beteiligten. Die Gruppendiskussionen bezeugen, dass der eigentliche Wunsch hinter dem Einsatz von *Smart Home* nicht selten der Wunsch nach einfachen Abschaltmechanismen, bspw. für Strom und Herd im Notfall, ist. Zudem verdeutlichen die Ergebnisse die Grenzen individueller Verantwortung und technischer Kompetenz für *Smart-Home*-Systeme. In Mehrpersonenhaushalten, bei *Cloud*-Diensten oder bei ungleichen Zugangsmöglichkeiten reicht individuelles Wissen allein nicht aus, um alle Risiken zu überblicken. So wird bspw. immer wieder von technisch versierten und fürs *Smart-Home*-System verantwortlichen Personen betont, dass sie die alleinigen Administrator:innen sind, die das System nach außen vollständig absichern. Doch auch in solchen Konstellationen gehen Risiken

einher. So erwähnte ein Teilnehmer in der Diskussion, das z. B. seine Ehefrau keinen Zugriff auf die Steuerung der Heizungsanlage hätte. Effektiver Schutz der individuellen Privatheit auch und gerade im eigenen Zuhause erfordert daher, dass Nutzung in technische, rechtliche und institutionelle Rahmenbedingungen eingebettet wird und die Aushandlung im Haushalt durch unterstützende Formate und geeignete Gestaltung begleitet wird.

5.3 Implikationen für weitere Forschung und Wissenstransfer

Die privatheitsrelevante Wirkung einfacher Sensordaten wird besonders dann erfahrbar, wenn alltagsnahe Formate und Methoden eingesetzt werden. Daten-Vignetten, Simulationen, *Workshops* und Elemente aus *Design Fiction* konfrontieren die Teilnehmenden mit irritierenden oder widersprüchlichen Szenarien und machen die Inferenzpotenziale der Daten nachvollziehbar. Als besonders effektiv hat sich der *Zugang zu den eigenen Daten* gezeigt. Personen, die sonst vor allem in einer sekundären Position sind, wird durch den Zugriff die eigene Abbildbarkeit bewusst, und es regt zur Reflexion an. Aber auch administrativ agierende primäre Nutzende überdachten ihre Position. Beispielsweise wurde im Studienkontext in einem Haushalt davon Abstand genommen, einen Amazon-Echo-Lautsprecher – wie ursprünglich geplant – im Schlafzimmer zu installieren. Da Selbsterfahrungsformate extrem aufwendig sind, hat sich auch das Format von Daten-Raten-*Workshops* als effektiv erwiesen, in denen fremde anonymisierte Datenausschnitte auf Basis kurzer Impulse gemeinsam interpretiert werden. Abstrakte Risiken werden dadurch in konkrete Alltagsszenarien übersetzt und Diskussionen über Angemessenheit, Fehlinterpretation sowie Daten als soziale und machtvoll Ressource ermöglicht. Verbraucherzentralen und andere Bildungsträger können diesbezüglich eine Multiplikator:innen-Rolle übernehmen, indem sie wissenschaftliche Befunde in alltagstaugliche Bildungsressourcen übersetzen und Nutzer:innen aktiv in Deutung und Reflexion einbeziehen. Formate, die den Zugang bzw. die Einsicht in die eigenen Daten durch andere simulieren, können eine ähnliche Auseinandersetzung ermöglichen (*Serious Games*).

Privacy-by-Co-Design erweist sich insgesamt als wertvolles und zentrales Konzept, da Partizipation zusätzliche Aspekte sichtbar macht, die allein durch antizipierende Planung von Produkten oder Fallstricken nicht erkennbar sind. Dies sind nicht antizipierte Bedrohungspotenziale, alltagspraktische Zumutungen und die Bedienbarkeit von Datenschutzmaßnah-

men im Umfeld von *Smart-Home*-Systemen. Im Studienkontext hat sich hier der Zugang zu den Daten als solcher Aspekt erwiesen. Die dabei erzeugte methodische Irritation ist kein Störfaktor, sondern ein Werkzeug, um implizite Annahmen über Technik, Normalität und Kontrolle offenzulegen und reflexive Distanz bei den Teilnehmenden zu fördern. Neben den Teilnehmenden sind auch die Gestalter:innen von *Smart-Home*-Technik eine Zielgruppe des Forschungsansatzes: Zwar kann in der Entwicklungspraxis nicht für jede neue *Smart-Home*-Anwendung so eine Feldstudie vorgelagert werden, aber eine grundlegende Sensibilisierung der Entwickler:innen für *Smart-Home*-Privatheit als soziale Aushandlung auf Haushaltsebene hat sich sowohl als Leerstelle als auch machbar erwiesen. Eine aus dem Projekt entwickelte Workshopmethode dazu berichten Becker u. a. (2026).

5.4 Implikationen für die politische Regulierung und Steuerung

Unsere Forschung zeigt, dass Privatheit für alle betroffenen Personen im *Smart Home* nicht allein durch individuelles Verhalten oder technische Gestaltung abgesichert werden kann, sondern eine normative und regulatorische Rahmung erforderlich erscheint. Zentral ist dabei das Transparenz-Erfordernis: *Smart-Home*-Systeme müssen klar aufzeigen, welche Daten sie erfassen, wie diese verarbeitet und aggregiert werden können und nicht zuletzt, wer Zugriff darauf hat. Ebenso wesentlich sind wirksame *Default*-Vorgaben. Sensoren sollten standardmäßig nur notwendige Daten erfassen, lokal speichern und leicht von außen deaktivierbar sein. Die Datenhaltung sollte standardisiert zeitlich sowie zweckgebunden begrenzt werden. Darüber hinaus müssen die unterschiedlichen Haushaltsrollen berücksichtigt werden. Erwachsene, Kinder und Dritte sollten klar definierte Zugriffsrechte erhalten, und Entscheidungen über die Nutzung oder Freigabe von Daten sollten gemeinschaftlich im Haushalt abgestimmt werden, um Informationsasymmetrien zu vermeiden. Dies kann bspw. über Vorgaben zu Nutzungs-IDs und Zustimmungserfordernissen unterstützt werden. Diese Vorgaben müssen insgesamt in technische und organisatorische Anforderungen übersetzt werden, etwa durch lokale Speicheroptionen, standardisierte Visualisierungen von Datenflüssen, verständliche Rollenmodelle sowie systematische Bildungs- und Reflexionsangebote. Auf diese Weise kann Privatheit nicht als statische Norm verstanden werden, sondern als dynamisches, relationales Gut, das situativ stabilisiert werden muss. Diese

Diskurs-, und Interpretationsprozesse in technische Abläufe zu übersetzen, so dass die Technik bei der Inbetriebnahme viele der oben erwähnten Aspekte in die Konfiguration übernehmen, ist das nächste Forschungsziel aus dem Projektkontext.

6. Fazit

Der Beitrag zeigt, dass vermeintlich einfache Sensordaten im häuslichen Umfeld komplexe Privatheitsimplikationen erzeugen, weil sie im Kontext sozialer Beziehungen interpretiert werden und als Ressource für Wissen, Kontrolle und Normdurchsetzung dienen. Sensoren, die Temperatur, Bewegung, Licht oder Luftfeuchtigkeit messen, erscheinen zunächst unproblematisch, entfalten ihre Bedeutung jedoch in Verbindung mit alltäglichen Routinen, Rollenverteilungen und der Organisation von Haushalten. Sensordaten werden nicht nur erfasst, sondern in Geschichten über An- und Abwesenheit, Verhalten und Abweichungen übersetzt und sind in soziale Aushandlungsprozesse eingebettet. Sie legitimieren Wertungen, stabilisieren normative Erwartungen und strukturieren Machtverhältnisse im Haushalt mit.

Die Daten und Visualisierungen fungieren nicht als technisch erfasste Informationen, sondern als Belege für bestehende Wert- und Normvorstellungen und werden mit autoritativer Aussagekraft, im Sinne von objektiven Fakten, eingesetzt, um Positionen Gewicht zu verleihen. Dies zeigt sich insbesondere hinsichtlich der Fragen von richtigem und falschem Lüften, zum Umgang mit Eigentum bzw. Mietsachen oder bei Fragen der Lärmbelästigung.

Durch die Verbindung von Feldforschung, partizipativen *Workshops* und reflexiver Analyse wird sichtbar, wie einfache Sensordaten komplexe soziale Dynamiken, Informationsasymmetrien und normative Aushandlungsprozesse auslösen. Sensordaten werden zudem leichtfertig als Mittel lateraler Überwachung genutzt. Sie erlauben die Beobachtung anderer Haushaltsmitglieder, werden für die Bewertung von deren Verhalten eingesetzt. Privatheit erweist sich in diesem Kontext nicht als individuelles Recht, sondern als relationales und situatives Gut, das in Aushandlungsprozessen, Konflikten und impliziten Hierarchien im Haushalt – und teils darüber hinaus – entsteht. Neben Haushaltsmitgliedern betrifft die Datenerhebung Besucher:innen, Handwerker:innen, und Pflegedienste. Schon auf Ebene der Studie war diese Zustimmung durch alle Betroffenen – trotz schrift-

licher und mündlicher Aufklärung und Ausgabe entsprechender Einwilligungsbögen – nicht immer gegeben. In der empirischen Welt außerhalb der Studie ist dies noch seltener der Fall.

Die Studie verdeutlicht, dass Problembewusstsein und Sensibilität gerade für nicht-audiovisuelle Sensorik schwach ausgeprägt sind. Temperatur-, Luftfeuchtigkeits- oder Geräuschdaten werden häufig als harmlos eingeschätzt, obwohl sie in Kombination mit Alltagswissen hochgradig aussagekräftig sind. Erst partizipative, datenbasierte Konfrontationen mit den eigenen Daten machten Risiken, Fehlinterpretationen und Konfliktpotenziale sichtbar. *Awareness* entsteht somit nicht primär durch Information, sondern durch Erfahrung und kollektive Reflexion.

Aus praktischer Perspektive ergibt sich, dass Privatheit im *Smart Home* nur durch das Zusammenspiel von Gestaltung, Nutzung und Bildung gesichert werden kann. Technische Systeme müssen Rollen- und Rechtskonzepte, Transparenzmechanismen und datensparsame Voreinstellungen bieten. Nutzung muss als sozialer Prozess verstanden werden, der Kommunikation, Aushandlung und gemeinsame Regeln einschließt. Bildung und Transfer befähigen Nutzer:innen, Daten zu interpretieren, Fehlinterpretationen zu erkennen und normative Implikationen zu reflektieren. Schutz entsteht demnach durch die Kombination von Technik, Praxis und Kompetenzinfrastruktur.

Regulatorische und pädagogische Ansätze schließen sich dabei nicht aus, sondern ergänzen sich. Vorgaben für Hersteller und Plattformen sichern Mindeststandards, während Aufklärung und Reflexion der Nutzer:innen die sozialen Dynamiken adressieren. Laterale Überwachung, normative Kontrolle und konfliktreiche Deutungen entstehen nicht primär durch Marktversagen, sondern durch alltägliche Nutzungskonstellationen. Nutzer:innen sind zugleich Adressat:innen und aktive Akteur:innen bei der Herstellung oder Unterminierung von Privatheit.

In der doppelten Bedeutung der ‚Nutzer:innen‘, einerseits als technische Bedienende und andererseits als soziale Akteur:innen, zeigt sich eine grundlegende Ambivalenz. Schutz erfordert sowohl Regulierung der Anbieter als auch Aufklärung derjenigen, die Technik einsetzen. Partizipative Aufklärung, datenbasierte Irritation und reflexive Nutzung sind kein Ersatz für Regulierung, sondern eine notwendige Ergänzung, um Privatheit im vernetzten Zuhause als relationales, situiertes und gestaltbares Gut zu stabilisieren.

Literatur

- Adryan, Boris (2015): Connected product labelling github.io URL: <https://badryan.github.io/2015/09/30/connected-product-labelling.html> (besucht am 02.06.2019)
- Becker, Alexa; Schäfer, Mirko; Kurze, Albrecht; Bischof, Andreas; Baum, Franziska; Pierce, James und Berger, Arne (2026): Design Strategies for Mitigating Tensions in Using Simple Smart Home Sensors. In: ACM (Hrsg.): *Proceedings of Designing Interactive Systems Conference (DIS) 2026, Singapore*. New York, NY: ACM, S. 722–733. <https://doi.org/10.1145/3800645.381298>
- Berger, Arne; Kurze, Albrecht; Totzauer, Sören; Storz, Michael; Lefevre, Kevin; Bischof, Andreas und Freiermuth, Mira (2018): Sensing Home: Designing an Open Tool That Lets People Collect and Interpret Simple Sensor Data from Their Homes. *i-com*, 17(2), S. 153–167.
- Bühr, Peter (2017): Report: A Trustmark for IoT. thingscon.com. URL: <https://www.thingscon.com/report-a-trustmark-for-iot/> (besucht am 24.06.2018)
- Caven, Peter; Zhang, Zitao; Abbott, Jacob; Ma, Xinyao und Camp, Ljean (2024): Comparing the Use and Usefulness of Four IoT Security Labels. In: ACM (Hrsg.): *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI '24, Honolulu, HI*. New York, NY: ACM, S. 1–31. <https://dl.acm.org/doi/10.1145/3613904.3642951>
- Deschamps-Sonsino, Alexandra (2018): *Smarter Homes: How Technology Will Change Your Home Life*. Berkeley, CA: Apress.
- Dereymaeker, Julie; Schokkenbroek, Janneke M.; Martens, Marijn und De Wolf, Ralf (2024): Smarter homes, smarter surveillance? Exploring intimate surveillance practices in modern day households. *New Media & Society*, 27(11), S. 6119–6138. <https://doi.org/10.1177/14614448241263757>
- Fischer, Joel E.; Crabtree, Andy; Rodden, Tom; Colley, James A.; Costanza, Enrico; Jewell, Michael O. und Ramchurn, Sarvapali D. (2016): „Just Whack It on Until It Gets Hot“: Working with IoT Data in the Home. In: ACM (Hrsg.): *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI'16, San Jose*. New York, NY: ACM, S. 5933–5944. <https://dl.acm.org/doi/10.1145/2858036.2858518>
- Garg, Vaibhav (2021): A Lemon by Any Other Label: In: Mori, Paolo, Lenzini, Gabriele und Furnell, Steven (Hrsg.): *Proceedings of the 7th International Conference on Information Systems Security and Privacy, ICISSP, Paolo Mori*, Vol. 1. Setúbal: Science and Technology Publications, S. 558–565. <https://doi.org/10.5220/0010295205580565>
- Hahn, Sara; Al-Amrani, Nusaibah; Allagah, Sara; Huang, He Rong; Kircher, Eva; Rehs, Cornelia; Vogt, Joachim; Marky, Karola und Gerber, Nina (2024): How to Respect Bystanders' Privacy in Smart Homes - A Co-Creation Study. In: *Proceedings of the 13th Nordic Conference on Human-Computer Interaction, NordiCHI'24, Upsala*, Artikel Nr. 7. New York, NY: ACM, S. 1–19. <https://doi.org/10.1145/3679318.3685340>

- Jin, Haojian; Guo, Boyuan; Roychoudhury, Rituparna; Yao, Yaxing; Kumar, Swarn; Agarwal, Yuvraj und Hong, Jason I. (2022): Exploring the Needs of Users for Supporting Privacy-Protective Behaviors in Smart Homes. In: ACM (Hrsg.): *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI'22, New Orleans, LA*, Artikel Nr. 449. New York, NY: ACM, S. 1–19. <https://doi.org/10.1145/3491102.3517602>
- Köpferl, Karola; Lehmann, Tanja; Börner, Andy; Becker, Alexa; Berger, Arne; Bischof, Andreas und Kurze, Albrecht (2024): Ich sehe das, was Du nicht siehst: Spuren einfacher Sensordaten im smarten Zuhause erleben und reflektieren. In: Gesellschaft für Informatik e.V., (Hrsg.): *Mensch und Computer 2024 – Workshopband, Karlsruhe*. Bonn: Gesellschaft für Informatik e.V. <https://doi.org/10.18420/muc2024-mci-ws17-153>
- Kruse, Jan (2015): *Qualitative Interviewforschung: ein integrativer Ansatz*, (Grundlagentexte Methoden). 2., überarbeitete und ergänzte Auflage. Weinheim Basel: Beltz Juventa.
- Kurze, Albrecht; Bischof, Andreas; Totzauer, Sören; Storz, Michael; Eibl, Maximilian; Brereton, Margot und Berger, Arne (2020): Guess the Data: Data Work to Understand How People Make Sense of and Use Simple Sensor Data from Homes. In: ACM (Hrsg.): *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI'20, Honolulu, HI*. New York, NY: ACM, S. 1–12. <https://doi.org/10.1145/3313831.33762>
- Kurze, Albrecht und Bischof, Andreas (9. Mai 2021): Wicked Implications for Human Interaction with IoT Sensor Data. Workshop Human-Data Interaction through Design at Conference on Human Factors in Computing Systems (CHI21), Yokohama. <https://dl.acm.org/doi/pdf/10.1145/3613905>
- Kurze, Albrecht (2022): Senseful Sensors: Learnings and Optimizations for IoT Sensors in HCI Research In: IEEE (Hrsg.): *2022 IEEE 9th International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, Chemnitz. Setúbal: Science and Technology Publications. <https://doi.org/10.1109/CIVEMSA53371.2022.9853698>
- Kurze, Albrecht und Reuter, Christin (2024): Understanding and addressing user needs for annotation of simple sensor data: Bridging the gap between human sensemaking and machine interpretation. In: Gesellschaft für Informatik e.V. (Hrsg.): *Proceedings of INFORMATIK 2024, Wiesbaden*. Bonn: Gesellschaft für Informatik e.V., S. 465–472. https://doi.org/10.18420/inf2024_34
- Kurze, Albrecht; Bischof, Andreas und Berger, Arne (2025): Smart, simple, sincere - Why and how we might rethink connected things in our smart homes. In: Stiftung ThingsCon (Hrsg.): *ThingsCon Report: The State of Responsible Technology 2025*. Amsterdam: ThingsCon, S. 24–30. <https://doi.org/10.5281/zenodo.16498330>
- Kurze, Albrecht; Becker, Alexa; Bergk, Jana und Reuter, Christin (2026): (De)Coding Lay Users' Annotations of Time-series Sensor Data from the Home. In: ACM (Hrsg.): *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems, CHI EA '26, Barcelona*. New York, NY: ACM, S. 1-7. <https://dl.acm.org/doi/10.1145/3772363.3798962>

- Laput, Gierad; Zhang, Yang und Harrison, Chris (2017): Synthetic Sensors: Towards General-Purpose Sensing. In: ACM (Hrsg.): *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, Denver, CO*. New York, NY: ACM, S. 3986–3999. <http://dx.doi.org/10.1145/3025453.3025773>
- Leitão, Roxanne (2019): Anticipating Smart Home Security and Privacy Threats with Survivors of Intimate Partner Abuse. In: *Proceedings of the 2019 on Designing Interactive Systems Conference. DIS' 19, San Diego*. New York, NY: ACM, S. 527–539. <https://doi.org/10.1145/3322276.3322366>
- Lopez-Neira, Isabel; Patel, Trupt; Parkin, Simon; Danezis, George und Tanczer, Leonie (2019): 'Internet of Things': How Abuse is Getting Smarter. In: *Safe – The Domestic Abuse Quarterly*, (63), S. 22–26. <http://dx.doi.org/10.2139/ssrn.3350615>
- Lupton, Deborah (2016): *The quantified self: A sociology of self-tracking*. Cambridge, Malden, MA: Polity.
- Marky, Karola; Voit, Alexandra; Stöver, Alina; Kunze, Kai; Schröder, Svenja und Mühlhäuser, Max (2020): „I don't know how to protect myself“: Understanding Privacy Perceptions Resulting from the Presence of Bystanders in Smart Environments. In: ACM (Hrsg.): *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society. Nordic Chi'20, Talin*, Artikel Nr. 4. New York, NY: ACM, S. 1–11. <https://doi.org/10.1145/3419249.3420164>
- Marky, Karola; Prange, Sarah; Mühlhäuser, Max und Alt, Florian (2021): Roles Matter! Understanding Differences in the Privacy Mental Models of Smart Home Visitors and Residents. In: ACM (Hrsg.): *20th International Conference on Mobile and Ubiquitous Multimedia. Association for Computing Machinery, Lueven*. New York, NY: ACM, S. 108–122. <https://doi.org/10.1145/3490632.3490664>
- Mascheroni, Giovanna (2020): Datafied childhoods: Contextualising datafication in everyday life. *Current Sociology*, 68(6), S. 798–781. <https://doi.org/10.1177/001139211880753>
- Morgner, Philipp; Müller, Christian; Ring, Matthias; Eskofier, Björn; Riess, Christian; Armknecht, Frederik und Benenson, Zinaida (2017): Privacy Implications of Room Climate Data. Computer Security. In: Foley, Simon; Gollmann, Dieter und Snekkenes, Einar (Hrsg.): *Proceedings of Computer Security, ESORICS 2017, Oslo*. Cham: Springer, S. 324–343. https://doi.org/10.1007/978-3-319-66399-9_18
- Mucha, Henrik; Barros, Ana; Benjamin, Jesse; Benz Müller, Christoph; Bischof, Andreas; Buchmüller, Sandra; Carvalho, Alexandra; Dhungel, Anna-Katharina; Draude, Claude; Fleck, Marc-Julian; Jarke, Julian; Klein, Stefanie; Kortekaas, Caroline; Kurze, Albrecht; Linke, Diane; Maas, Franzisca; Marsden, Nicola; Melo, Ricardo; Michel, Susanne und Berger, Arne (2022): Collaborative Speculations on Future Themes for Participatory Design in Germany. *i-com*, 21(2), S. 283–298. <https://doi.org/10.1515/icom-2021-0030>
- Ramokapane, Kopo M.; Sameen, Maria und Dkaidek, Zaina (2024): Inclusive Internet of Things Privacy Labels. *IEEE Security & Privacy*, 22(5), S. 32–39. <https://doi.org/10.1109/MSEC.2024.3417819>

- Reinhardt, Delphine und Railean, Alexandr (2020): OnLITE: On-line Label for IoT Transparency Enhancement. In: Asplund, Mikael und Nadjm-Tehrani, Simin (Hrsg.): *Proceedings of the 25th Nordic Conference on Secure IT Systems (NordSec), 2020: Secure IT Systems*. Cham: Springer Nature, S. 229–245. https://doi.org/10.1007/978-3-030-70852-8_14
- Reuter, Christin (2025): Eine Nutzerfreundliche Annotationslösung für Einfache Sensordaten im Zuhause. 18. *Internationales Symposium für Informationswissenschaft (ISI 2025)*, Chemnitz. <https://doi.org/10.5281/zenodo.15511487>
- Reuter, Christin und Kurze, Albrecht (2025): AnnoTool: A User-Friendly Annotation Tool for Simple Sensor Data in the Home. In: Gesellschaft für Informatik e.V. (Hrsg.): *Mensch und Computer 2025 – Workshopband, Dresden*. Bonn: Gesellschaft für Informatik e.V. <https://doi.org/10.18420/muc2025-mci-wip-284>
- Richter, Johanna; Bischof, Andreas; Kurze, Albrecht; Totzauer, Sören; Freiermuth, Mira; Storz, Michael; Lefeuve, Kevin und Berger, Arne (2018): Machtförmige Praktiken durch Sensordaten in Wohnungen. In: Gesellschaft für Informatik e.V. (Hrsg.): *Mensch und Computer 2018 – Tagungsband, Dresden*. Bonn: Gesellschaft für Informatik e.V. <https://doi.org/10.18420/muc2018-mci-0253>
- Rogers, Michaela M.; Fisher, Coleen; Ali, Parveen; Allmark, Peter und Fontes, Lisa (2023): Technology-Facilitated Abuse in Intimate Relationships: A Scoping Review. *Trauma, Violence, & Abuse*, 24(4), S. 2210–2226. <https://doi.org/10.1177/15248380221090218>
- Saqib, Eimaan; He, Shijing; Choy, Junghyun; Abu-Salma, Ruba; Such, Jose; Bernd, Julia und Javed, Mobin (2025): Bystander Privacy in Smart Homes: A Systematic Review of Concerns and Solutions. *ACM Transactions on Computer-Human Interaction*, 32(5), S. 1–43. <https://doi.org/10.1145/3731755>
- Schulz, Annika Sabrina; Müller, Johanna und Beruscha, Frank (2023): Experience by Cohabitation: Living in a Smart Home Initiated by Your Partner. In: Abdelnour Nocera, Jose; Kristín Lárusdóttir, Marta; Petrie, Helen; Piccinno, Antonio und Winckler, Marco (Hrsg.): *Human-Computer Interaction – INTERACT 2023 (Lecture Notes in Computer Science, Bd. 14144)*. Cham: Springer Nature, S. 304–323. https://doi.org/10.1007/978-3-031-42286-7_17
- Stark, Luke und Levy, Karen (2018): The Surveillant Consumer. *Media, Culture & Society*, 40(8), S. 1202–1220. <https://doi.org/10.1177/0163443718781985>
- Tanczer, Leonie Maria (2021): Das Internet der Dinge: Die Auswirkungen »smarter« Geräte auf häusliche Gewalt. In: bff: Bundesverband Frauenberatungsstellen und Frauennotrufe; Prasad, N. (Hrsg.): *Geschlechtsspezifische Gewalt in Zeiten der Digitalisierung. Formen und Interventionsstrategien (Gender Studies)*. Bielefeld: transcript, S. 205–226. <https://doi.org/10.14361/9783839452813>
- Tolmie, Peter; Crabtree, Andy; Rodden, Tom; Colley, James und Luger, Ewa (2016): “This has to be the cats”: Personal Data Legibility in Networked Sensing Systems. In: ACM (Hrsg.): *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, CSCW’16, Vienna*. New York, NY: ACM, S. 491–502. <https://dl.acm.org/doi/10.1145/2818048.2819992>
- Véliz, Carissa (2021): *Privacy is power: why and how you should take back control of your data*. London: Corgi Books.

- Wang, Ge; Zhao, Jun; Van Kleek, Max; Pea, Roy und Shadbolt, Nigel (2025): FamiData Hub: A Speculative Design Exploration with Families on Smart Home Datafication. In: ACM (Hrsg.): *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. Chi'25, Yokohama*. New York, NY: ACM, Artikel Nr. 940, S. 1–17. <https://dl.acm.org/doi/10.1145/3706598.3713494>
- Wolfangel, Eva (2024): Rauchmelder bei Vonovia: Der Spion an meiner Decke. In: *Zeit Online vom 24.10.2024*. Zeit.de URL: <https://www.zeit.de/digital/datenschutz/2024-10/rauchmelder-vonovia-mieter-datenschutz-smart> (besucht am 12.03.2026)
- Yao, Yaxing; Basdeo, Justin Reed; Kaushik, Smirity und Wang, Yang (2019): Defending My Castle: A Co-Design Study of Privacy Mechanisms for Smart Homes. In ACM (Hrsg.): *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI'19, Glasgow*, Artikel Nr. 198. New York, NY: ACM, S. 1–12. <https://doi.org/10.1145/3290605.3300428>
- Zufferey, No  ; Salehzadeh Niksirat, Kavous; Humbert, Mathias und Huguenin, K  vin (2024): Our Data, Our Solutions: A Participatory Approach for Enhancing Privacy in Wearable Activity Tracker Third-Party Apps. *Proceedings on Privacy Enhancing Technologies*, 2024(4), S. 734–754. <https://doi.org/10.56553/popets-2024-0139>

Governance-Strukturen und staatliches Handeln im technologischen Wandel

Mehr Klarheit im Datenraum: Die Abgrenzung des Datenvermittlungsdienstes zum Infrastrukturanbieter

Katharina Schilke

Zusammenfassung

Die rechtliche Einordnung von Datenvermittlungsdiensten nach dem Data Governance Act ist auch nach vier Jahren noch nicht abschließend geklärt. Rechtliche Unsicherheiten ergeben sich insbesondere bei der Abgrenzung verschiedener Dienste in sogenannten Datenräumen. Dabei ist es für Diensteanbieter keinesfalls unbedeutend, ob sie nach dem Data Governance Act als Anbieter von Datenvermittlungsdiensten gelten – knüpft hier doch das strenge Regelungsregime des Art. 12 DGA an, das Auswirkungen auf die Erbringung solcher Dienste hat. Zur weiteren Systematisierung des Begriffs des Datenvermittlungsdienstes analysiert der Beitrag das Merkmal der „Absicht, eine Geschäftsbeziehung herzustellen“. Im Lichte dieser Analyse zeigt sich, inwiefern bisherige Auslegungen zum Begriff der Geschäftsbeziehung überdacht werden müssen und wo sich weitere Abgrenzungsfragen in der Praxis ergeben. Daran schließt eine forschungsleitende Fragestellung des Beitrages an: Wie lassen sich Infrastrukturanbieter im Komplex der Datenvermittlungsdienste einordnen? Als Lösung wird ein Unmittelbarkeitskriterium vorgeschlagen, das danach fragt, wie direkt ein Dienst an der Anbahnung wirtschaftlicher Datenbeziehungen beteiligt ist.

1. Einleitung¹

Daten sind zur Schlüsselressource in der Wirtschaft aufgestiegen.² Als Basis für vielfältige Wertschöpfungsmöglichkeiten stellen Daten längst einen

1 Der Beitrag ist im Rahmen des BMFTR-geförderten Projekts ScaleTrust (Förderkennzeichen 16DTM312B) entstanden.

2 Mitteilung der Europäischen Kommission über die europäische Datenstrategie <https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52020DC0066>, Link zuletzt abgerufen am 17.04.2026, Brüssel 2020 (nachfolgend: Europäische Datenstrategie), S. 3.

eigenen volkswirtschaftlichen Produktionsfaktor dar.³ Die Europäische Datenstrategie zielt darauf ab, diese „Lebensader der wirtschaftlichen Entwicklung“ im Einklang mit den europäischen Werten nutzbar zu machen.⁴ Voraussetzung dafür ist das Vertrauen der betroffenen Personen und Dateneinhaber in die Datenwirtschaft, insbesondere in den Datenaustausch. Neben der strikten Einhaltung der datenschutzrechtlichen Vorgaben des EU-Rechts sind *Datenvermittlungsdienste* ein wichtiger Baustein zur Stärkung dieses Vertrauens.

Mit dem Data Governance Act⁵ (nachfolgend: DGA) hat der europäische Gesetzgeber einen regulatorischen Rahmen für die Datenverwaltung geschaffen. Zur Förderung des Datenaustauschs in der Union findet sich in Kapitel III DGA ein Anmelde- und Aufsichtsrahmen für die Erbringung von Datenvermittlungsdiensten. Diese Dienste fungieren als Mittler zwischen Dateneinhaber und betroffener Person auf der einen Seite und Datennutzer auf der anderen Seite. Durch ihre „Matchmaking-Tätigkeit“ sollen die Datenvermittlungsdienste das Auffinden passender datenbezogener Angebote erleichtern. Daneben sollen die Dienste das Vertrauen in die Datenwirtschaft und den Datenaustausch stärken.⁶ Gerade das bestärkte Vertrauen der Akteure spielt eine Schlüsselrolle für die Wertschöpfung mit Daten. Nur bei vorhandenem Vertrauen werden die Akteure gewillt sein, Daten weiterzugeben.⁷ Hier setzt der europäische Gesetzgeber auf ein strenges Regelungsregime. Durch die in Art. 12 DGA normierten Bedingungen werden etwa die Neutralität der Dienste, die Sicherheit der Daten, der faire, diskriminierungsfreie und transparente Zugang zu den Diensten und die angemessene Weiterführung im Insolvenzfall sichergestellt.⁸

Neben dem Vertrauensaufbau haben Datenvermittlungsdienste auch das Potential, die Datensouveränität zu stärken. Der Datenvermittlungsdienst kann betroffene Personen und Rechteinhaber bei der Wahrung und Durchsetzung ihrer Rechte unterstützen und das Ungleichgewicht von Verhand-

3 Duhr/Horn/Pfau/Schaefer/Zimaj, in: Schwartmann/Benedikt/Reif, 1. Auflage 2025, Kapitel 21 Rn. 15.

4 Europäische Datenstrategie, S. 1, 3.

5 Verordnung (EU) 2022/868 des Europäischen Parlaments und des Rates vom 30. Mai 2022 über europäische Daten Governance und zur Änderung der Verordnung (EU) 2018/1724 (Daten-Governance-Rechtsakt).

6 Specht-Riemenschneider, in: HK-DatenR, 2. Auflage 2025, DGA Art. 10 Rn. 11.

7 Vgl. Europäische Datenstrategie, S. 1, 8 f.

8 Vgl. Richter, Datenvermittlungsdienste gem. Art. 2 Nr. 11 DGA, EuDIR 2025, 264.

lungspositionen in der Datenwirtschaft ausgleichen.⁹ Den Akteuren kann so zu mehr Kontrolle über ihre Daten verholfen werden.¹⁰ Teilweise sind auch die Bedingungen in Art. 12 DGA zur Erbringung von Datenvermittlungsdiensten souveränitätsstützend. So ist beispielsweise die Verwendung von zusätzlichen spezifischen Werkzeugen und Diensten zur Erleichterung des Datenaustauschs nach Art. 12 lit. e DGA an den ausdrücklichen Antrag oder die Zustimmung des Dateninhabers oder der betroffenen Person geknüpft.

Die Anwendung dieser vertrauens- und souveränitätswahrenden Bedingungen hängt davon ab, ob ein Dienst als Datenvermittlungsdienst im Sinne des Art. 2 Nr. 11 i. V. m. Art. 10 DGA einzuordnen ist. Anbieter von Datenvermittlungsdiensten im Sinne des Art. 2 Nr. 11 i. V. m. Art. 10 DGA dürfen gemäß Art. 11 Abs. 9 DGA das Label „in der Union anerkannter Anbieter von Datenvermittlungsdiensten“ führen. Dadurch kann Vertrauen bei den Nutzenden gewonnen werden. Gleichzeitig bedeutet das Regelungsregime des Art. 12 DGA einen häufig kritisierten Hemmungseffekt für den Aufbau und die Bereitstellung solcher Dienste.¹¹ Gerade vor dem Hintergrund der Vor- und Nachteile der bestehenden Regelungen ist es für den Rechtsanwender von erheblicher Bedeutung, welche Dienste den Anforderungen des Kapitel III DGA unterfallen.

Gerade hier liegt ein zentrales Problem: Die Begriffsbestimmung ist bewusst weit und technologieoffen gefasst. Dies führt jedoch zu rechtlicher Unsicherheit¹²: Wer unterfällt dem Begriff und wer unterliegt damit den bekannten Pflichten? Insbesondere sonstige datenbezogene Dienste – wie die Bereitstellung der technischen Infrastruktur zum Datenteilen¹³ – können je nach Ausgestaltung und Kombination der Dienste Grenzfälle darstellen. Dabei ist die richtige Einordnung von erheblicher Bedeutung für die Akteure der Datenwirtschaft und die Zielsetzungen der Europäischen Union. Die strengen Bedingungen des Kapitel III DGA sollten nur diejenigen Dienste treffen, die nach der Vorstellung des Gesetzgebers beim Vertrauensaufbau

9 Beise, Datensouveränität und Datentreuhand, RD i 2021, 597 (604).

10 Europäische Datenstrategie, S. 12.

11 Beise, Datensouveränität und Datentreuhand, RD i 2021, 597 (604); Schwamberger, in: Spindler/Schuster/Kaesling, 5. Auflage 2026, DGA Art. 12 Rn. 1; Specht-Riemenschneider, in: HK-DatenR, 2. Auflage 2025, DGA Art. 12 Rn. 25.

12 Richter, Datenvermittlungsdienste gem. Art. 2 Nr. 11 DGA, EuDIR 2025, 264; Tolks, Die finale Fassung des Data Governance Act, MMR 2022, 444 (446).

13 Schleimer u. a., in: Proff (Hrsg.), New Players in Mobility, Wissenschaftsforum Mobilität, 2025, 551 (558).

und einem offenen Datenzugang eine Schlüsselrolle einnehmen. Zweifelsannahmen – wie etwa beim Personenbezug nach der DSGVO – aufgrund von Rechtsunsicherheit würden wegen der strengen Bedingungen zu einem noch umfassenderen Hemmungseffekt in der Datenwirtschaft führen, als er ohnehin schon in Bezug auf Art. 12 DGA befürchtet wird.

Der Beitrag widmet sich dieser begrifflichen Unbestimmtheit. Ziel ist insbesondere die Abgrenzung von Datenvermittlungsdiensten und anderen datenbezogenen Diensten an der Schwelle zum Datenvermittlungsdienst. Ausgangspunkt ist die Frage, wann ein Dienst überhaupt als Datenvermittlungsdienst im Sinne des Art. 2 Nr. 11 DGA einzuordnen ist. Im Zentrum steht dabei zunächst der Begriff der *Geschäftsbeziehung*, der anhand einer beziehungsorientierten Auslegung konkretisiert wird. Darauf aufbauend wird untersucht, wie das subjektive Element der *Absicht* Geschäftsbeziehungen herzustellen objektivierbar ist, insbesondere ob die Herstellung von Geschäftsbeziehungen den Hauptzweck des Dienstes darstellen muss. Auf dieser Grundlage wendet sich der Beitrag der Abgrenzung von sonstigen datenbezogenen Diensten zu und erwägt hierfür ein Unmittelbarkeitskriterium. In einem weiteren Schritt wird die Einordnung von zugehörigen Diensten im Sinne des Art. 10 lit. a DGA und Art. 12 lit. e DGA untersucht.

2. Der Begriff der Geschäftsbeziehungen gemäß Art. 2 Nr. 11 DGA

Ausgangspunkt der Analyse ist die Begriffsbestimmung in Art. 2 Nr. 11 DGA. Danach ist ein Datenvermittlungsdienst „[ein] Dienst, mit dem durch technische, rechtliche oder sonstige Mittel Geschäftsbeziehungen zwischen einer unbestimmten Anzahl von betroffenen Personen oder Dateninhabern einerseits und Datennutzern andererseits hergestellt werden sollen, um die gemeinsame Datennutzung [...] zu ermöglichen [...]“.

Die zitierte Positivdefinition wird durch Negativbeispiele zumindest teilweise konkretisiert.¹⁴ Der Anwendungsbereich des Kapitels III wird zudem durch die Auflistung bestimmter Datenvermittlungsdienste in Art. 10 DGA präzisiert.¹⁵ Art. 10 DGA führt allerdings zu keinen wesentlichen Beschrän-

14 Specht-Riemenschneider, in: HK-DatenR, 2. Auflage 2025, DGA Art. 2 Rn. 64.

15 Vgl. Schwamberger, in: Spindler/Schuster/Kaesling, 5. Auflage 2026, DGA Art. 10 Rn. 4.

kungen, da Dienste im Sinne des Art. 2 Nr. 11 DGA in aller Regel auch den Bestimmungen des Art. 10 DGA unterfallen.¹⁶

Von wesentlicher Bedeutung für die Einordnung eines Dienstes als Datenvermittlungsdienst, und damit auch für die Abgrenzungsfragen dieses Beitrags, ist die *Absicht Geschäftsbeziehungen zum Zweck der gemeinsamen Datennutzung herzustellen*. Dieses Tatbestandsmerkmal enthält wiederum einzelne ausfüllungsbedürftige Rechtsbegriffe: Geschäftsbeziehung, gemeinsame Datennutzung und Herstellungsabsicht. Die einschlägige Literatur hat bereits einige Anstrengungen unternommen, die begriffliche Unbestimmtheit zu konkretisieren. Daran knüpft dieser Beitrag an, greift zentrale Positionen auf und ordnet diese kritisch ein, um eine konsistente Grundlage für nachfolgende Überlegungen zu schaffen.

Zunächst stellt sich die Frage, was unter einer *Geschäftsbeziehung* zu verstehen ist. In diesem Zusammenhang muss auch betrachtet werden, wie die Zwecksetzung – die „Ermöglichung der gemeinsamen Datennutzung“ – zur begrifflichen Konkretisierung beiträgt.

Eine Legaldefinition des Begriffs der *Geschäftsbeziehung* findet sich im DGA nicht. Die Literatur hat hier in dem Versuch, diesen unbestimmten Rechtsbegriff zu schärfen, verschiedene Ansichten entwickelt. Dabei divergieren die Ansichten dazu, was Bezugspunkt des „Geschäftlichen“ ist. Teilweise wird darauf abgestellt, ob die beteiligten Datenbereitsteller und Datennutzer geschäftlich handeln.¹⁷ Teilweise wird aber auch der Zweck der an die Datenbereitstellung anschließenden Datennutzung als maßgebliches Kriterium benannt.¹⁸

2.1 Die „akteursbezogene“ Ansicht

Die erstgenannte Ansicht geht davon aus, dass eine Geschäftsbeziehung eine Interaktion zwischen Parteien in Ausübung ihrer gewerblichen oder selbstständigen beruflichen Tätigkeit erfordert.¹⁹ Diese Ansicht bezieht die „Geschäftlichkeit“ somit auf die Akteure. Bezugspunkt ist eine gewerbliche oder selbstständig berufliche Tätigkeit der Datenbereitsteller (Dateninha-

16 Specht-Riemenschneider, in: HK-DatenR, 2. Auflage 2025, DGA Art. 10 Rn. 16.

17 Schild u. a., in: BeckOK Datenschutzrecht, 54. Edition 2025, DGA Art. 2 Rn. 93.

18 von Ditfurth 2024, S. 235.

19 Richter, Datenvermittlungsdienste gem. Art. 2 Nr. 11 DGA, EuDIR 2025, 264 (267); Schild u. a., in: BeckOK Datenschutzrecht, 54. Edition 2025, DGA Art. 2 Rn. 93.

ber oder betroffene Person) und Datennutzer. Die Interaktion, die der Dienst durch seine Vermittlung ermöglicht hat, muss in Ausübung dieser gewerblichen bzw. selbstständigen Tätigkeit erfolgen. Dadurch werden Vermittlungen ausgeschlossen, die zwischen privat oder im Schwerpunkt nicht gewerblich oder selbstständig beruflich handelnden Akteuren herbeigeführt werden. So werden beispielsweise betroffene Personen, die ihre personenbezogenen Daten anbieten, in aller Regel nicht erfasst. Diese *akteursbezogene* Auffassung wird damit begründet, dass mit Kapitel III DGA die Förderung des Datenaustauschs zwischen Unternehmen bezweckt werde. Dabei verweist diese Ansicht auch auf ErwG 27 DGA.²⁰ Auch öffentliche Stellen könnten gewerblich handeln, sodass auch deren Datenbereitstellung erfasst sei.²¹

Richtig ist, dass in ErwG 27 S.1 DGA den Datenvermittlungsdiensten „insbesondere [...] die Unterstützung und Förderung freiwilliger Verfahren zur gemeinsamen Datennutzung zwischen Unternehmen“ beigemessen wird. Wie das Wort „insbesondere“ zeigt, handelt es sich hierbei jedoch nur um eine beispielhafte Aufzählung. Damit ist die Schlussfolgerung, nur gewerblich bzw. selbstständig beruflich handelnde Akteure könnten von einem Datenvermittlungsdienst in diesem Sinne adressiert werden, nicht zwingend.

Weiter heißt es in ErwG 27 S.3 DGA, dass Datenvermittlungsdienste die verschiedenen Akteure miteinander verbinden und so zur effizienten Bündelung von Daten sowie zur Erleichterung des bilateralen Datenaustauschs beitragen können. Danach liegt der Zweck von Datenvermittlungsdiensten im Zusammenführen verschiedenster Akteure, um Daten *als wirtschaftliches Gut* verfügbar zu machen. Der Bezugspunkt ist hier das Ermöglichen des Datenteilens durch das Verbinden der Akteure und nicht die Einordnung der Akteure als Gewerbe oder Unternehmen.

Zudem kann man dieser Ansicht den Wortlaut des Gesetzes entgegenhalten. Danach sind die zu vermittelnden Akteure nicht nur gewerblich oder selbstständig beruflich handelnde Personen. So können betroffene Personen nach Art. 2 Nr. 11 DGA i. V. m. Art. 2 Nr. 7 DGA und natürliche Personen nach Art. 2 Nr. 11 DGA i. V. m. Art. 2 Nr. 8 DGA als Bereitsteller ihre Daten teilen und nach Art. 2 Nr. 11 DGA i. V. m. Art. 2 Nr. 9 DGA können auch natürliche Personen Datennutzer sein. Der Begriff der betrof-

20 Richter, Datenvermittlungsdienste gem. Art. 2 Nr. 11 DGA, EuDIR 2025, 264 (267).

21 Richter, Datenvermittlungsdienste gem. Art. 2 Nr. 11 DGA, EuDIR 2025, 264 (267).

fenen Person im Sinne des Art. 2 Nr. 7 DGA i. V. m. Art. 4 Nr. 1 DSGVO²² knüpft nicht an eine gewerbliche oder selbstständige Tätigkeit an.²³ Vielmehr steht der Begriff unabhängig von einer solchen Tätigkeit. Da der DGA ausdrücklich an das Begriffsverständnis der DSGVO anknüpft, ist eine gewerbliche oder selbstständige Tätigkeit aufseiten der Bereitsteller nicht zwingend.

Dafür spricht auch, dass natürliche oder juristische Personen ohne gewerblichen Hintergrund Datenbereitsteller und -nutzer sein können.²⁴ Auch diese Begriffe stehen in der unionsrechtsautonomen Auslegung unabhängig von einer gewerblichen oder selbstständig beruflichen Tätigkeit. Der Begriff der juristischen Person umfasst etwa im Kontext des Vertrags über die Arbeitsweise der Europäischen Union (AEUV) „alle Einheiten des öffentlichen und privaten Rechts, die von einer zuständigen Rechtsordnung (innerhalb oder außerhalb der EU) mit Rechtspersönlichkeit versehen worden sind“²⁵. Hier wird nicht Bezug zu einem Gewerbe oder einer selbstständigen beruflichen Tätigkeit genommen. Dies kann als Anhaltspunkt dienen, da der DGA auf Grundlage des AEUV, insbesondere des Art. 114 AEUV erlassen wurde (Präambel, DGA). Zudem fasst auch Art. 2 Nr. 8 DGA öffentliche Stellen und internationale Organisationen („einschließlich“) unter den Begriff des „Dateninhabers“, die oftmals nicht gewerblich oder selbstständig beruflich handeln.²⁶ Folglich richtet sich der DGA allgemein an verschiedenste Akteure, die nicht zwingend gewerblich oder selbstständig beruflich handeln.

2.2 Die „nutzungsbezogene“ Ansicht

Die zweite Ansicht stellt auf die Zwecksetzung durch den Datennutzer ab. Nach dieser *nutzungsbezogenen* Auffassung ist maßgeblich, dass die

22 Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung). ABl. L 119 vom 4.5.2016.

23 Gola, in: Gola/Heckmann, 3. Auflage 2022, DSGVO Art. 4 Rn. 26.

24 Bobev u. a., White Paper on the Definition of Data Intermediation Services, 2023, S. 38.

25 Wendenburg, in: Europäisches Unionsrecht, 8. Auflage 2026, AEUV Art. 263 Rn. 27; ähnlich: Forsthoff, in: Das Recht der EU, 86. EL 2025, AEUV Art. 54 Rn. 6.

26 von Ditzfurth 2024, S. 234.

gemeinsame Datennutzung einem kommerziellen Zweck dient.²⁷ Dies sei der Fall, wenn der Datennutzer eine Datennutzung bezweckt, die im Zusammenhang mit seinen Geschäften erfolgt und zumindest mittelbar auf die Erzielung von Gewinn gerichtet ist.²⁸ Dies wäre etwa bei Forschungsvorhaben fraglich.

Diese Auffassung argumentiert ebenfalls mit dem Wortlaut des ErwG 27 DGA. Bezugspunkt ist hierbei, dass die Datenvermittlungsdienste Datenverwendungspraktiken zwischen Unternehmen fördern sollen. Dies zeige, dass der Gesetzgeber mit dem Begriff der „Geschäftsbeziehung“ nur Datenweitergaben und -nutzungen mit wirtschaftlicher Zielsetzung erfassen wollte. Dadurch würden Anbieter, deren Zweck in der Anbahnung von Datenweitergaben zu nicht kommerziellen Zwecken besteht, von B2B-Datenvermittlern im Sinne des Art. 10 lit. a DGA abgegrenzt.²⁹

Diese Auffassung ist nicht zwingend. Zunächst spricht der Wortlaut des Art. 2 Nr. 11 DGA gegen den Nutzungszweck als Bezugspunkt der Geschäftsbeziehung. Das „Geschäft“ bezieht sich schon nach dem Wortlaut der Norm auf die Beziehung. Es ist nicht die Rede von einer geschäftlichen Nutzung, sondern einer „geschäftlichen Beziehung“ (engl. „commercial relationship“ oder frz. „relations commerciales“). Betrachtet man die anderen Sprachfassungen, ergibt sich aus dem Wortlaut kein Bezug zur Nutzung der Daten („commercial relationships for the purposes of data sharing“).

An dieser Stelle ist auch eine nähere Betrachtung des Zwecks der „gemeinsamen Datennutzung“ erforderlich. In der deutschen Sprachfassung des Art. 2 Nr. 11 DGA heißt es, dass „Geschäftsbeziehungen [...] hergestellt werden sollen, um die gemeinsame Datennutzung [...] zu ermöglichen“. Nach der Begriffsbestimmung der „gemeinsamen Datennutzung“ in Art. 2 Nr. 10 DGA handelt es sich dabei um „die entgeltliche oder unentgeltliche Bereitstellung von Daten [...]“. Gemeint ist nicht primär die Nutzung selbst, sondern die Bereitstellung der Daten für eine dann weitergehende Nutzung.³⁰ Der Begriff der „gemeinsamen Datennutzung“ in Art. 2 Nr. 10 DGA ist insofern missverständlich. Gegenstand ist nicht die an den Austausch der Daten anschließende Nutzung, sondern eine „gemeinsame Nutzung“ im Sinne eines „Geben und Nehmens“ – klarer ausgedrückt: das Teilen von Daten. Dies kommt in den anderen Sprachfassungen besser zum Ausdruck,

27 von Ditfurth 2024, S. 235.

28 von Ditfurth 2024, S. 235.

29 von Ditfurth 2024, S. 236.

30 Schild u. a., in: BeckOK Datenschutzrecht, 54. Edition 2025, DGA Art. 2 Rn. 77f.

die etwa von „data sharing“ (Englisch) und „partage de données“ (Französisch) sprechen. Die Festlegung des Zwecks „für die gemeinschaftliche oder individuelle Nutzung dieser Daten“ dient lediglich der Klarstellung, dass sämtliche Formen der Weiterverwendung erfasst sind.³¹ Im Zusammenhang mit dem Begriff der Geschäftsbeziehung wird deutlich, dass nicht die Datennutzung Bezugspunkt des Geschäfts ist, sondern das Datenteilen. Betrachtet man die einzelnen Abschnitte als Ausschnitt des Daten-Lebenszyklus, wird die Nutzung in der Sache nicht erfasst, sondern ist nur Teil der Zwecksetzung.

Darüber hinaus wird die Datennutzung durch den Datennutzer in Kapitel III DGA nicht adressiert oder reguliert. Eine „nutzungsbezogene“ Auslegung erkennt die im DGA angelegte Trennung zwischen der Bereitstellung, der Vermittlung und der Nutzung von Daten (ErwG 32 S. 5 DGA). Diese Trennung dient dem Aufbau von Vertrauen in die Datenvermittlungsdienste (ErwG 32 S. 1 DGA). Wie Art. 12 lit. a DGA zeigt, treten die Dienste ausschließlich bei der Vermittlung von Daten auf und dürfen selbst keine weitergehende Nutzung der Daten verfolgen. Der Gesetzgeber reguliert nur das Vorgehen der Datenvermittlungsdienste, nicht aber das der Nutzer. Folglich darf auch der Begriff des Datenvermittlungsdienstes – insbesondere der Geschäftsgegenstand – nicht anhand einer an das Datenteilen anschließenden Datennutzung ausgelegt werden.³² Die Vermittlung der Beziehung der Akteure zum Zweck des Datenteilens ist somit von der weitergehenden Nutzung zu differenzieren.

Die „nutzungsbezogene“ Ansicht verlagert die Zielsetzung des Dienstes zudem auf die Nutzerseite. Damit wird die Einordnung als Datenvermittlungsdienst von subjektiven, dienstfremden Motiven abhängig gemacht. Da die Gewinnerzielungsabsicht des Datennutzers für den Dienst weder verlässlich erkennbar noch überprüfbar ist, entstehen Rechtsunsicherheiten für den Diensteanbieter. Dies widerspricht zudem der ex-ante-Regulierung³³ von Datenvermittlungsdiensten, da es sich für die Dienste und Aufsichtsbehörden schwierig gestalten dürfte, die Zwecksetzungen der Datennutzer rechtssicher im Vorfeld der Dienstleistung festzustellen. Daher kann eine „nutzungsbezogene“ Auslegung nicht überzeugen.

31 Bobev u. a., White Paper on the Definition of Data Intermediation Services, 2023, S. 39.

32 Bobev u. a., White Paper on the Definition of Data Intermediation Services, 2023, S. 40.

33 von Ditfurth 2024, S. 232.

2.3 Eine „beziehungsorientierte“ Ansicht

Im Ergebnis hängt die „Geschäftlichkeit“ von der Beziehung zwischen Datenbereitsteller und Datennutzer ab.³⁴ Dafür spricht der Wortlaut des Art. 2 Nr. 11 und ErwG 28 S. 1 DGA („geschäftliche Beziehung“) ebenso wie die Systematik des Gesetzes – Regelung der Vermittlung von Daten – und der Sinn und Zweck des Datenvermittlungsdienstes – das Verfügbarmachen von Daten durch das Zusammenbringen der Akteure.

Eng damit verbunden ist die Frage, wann die Beziehung geschäftlicher Natur ist – wobei „geschäftlich“ europarechtlich autonom und im Kontext des DGA auszulegen ist. Dies ist der Fall, wenn eine Gegenleistung vereinbart wird, die über eine bloße Entschädigung der Bereitstellungskosten hinausgeht.

Gerade hierin liegt die zentrale Abgrenzung zum *Dataltruismus*. Dieser wird in Art. 2 Nr. 16 DGA legaldefiniert als „die freiwillige gemeinsame Nutzung von Daten [...], ohne hierfür ein Entgelt zu fordern oder zu erhalten, das über eine Entschädigung für die [den betroffenen Personen und Dateninhabern] durch die Bereitstellung ihrer Daten entstandenen Kosten hinausgeht, für Ziele von allgemeinem Interesse [...]“. Auch hier ist die freiwillige gemeinsame Nutzung wieder als das freiwillige Teilen von Daten (engl. „the voluntary sharing of data“) zu verstehen. Der Datenaltruismus hängt im Übrigen von zwei wesentlichen Merkmalen ab: dem Fehlen eines Entgelts (engl. „reward“) und der Zielsetzung von allgemeinem Interesse. Im Umkehrschluss hängt eine „Geschäftsbeziehung zur Ermöglichung des Datenteilens“ von der Erbringung einer Gegenleistung ab, die über eine bloße Entschädigung der Bereitstellungskosten hinausgeht.³⁵ Dafür spricht auch ErwG 28 S. 4 des DGA, der im Kontext von Datenbeständen ausdrücklich von einer Gegenleistung aller Teilnehmer für ihren Beitrag spricht. Im Ergebnis ist die „Geschäftsbeziehung“ als eine Beziehung wirtschaftlicher Art zu verstehen, wobei die Wirtschaftlichkeit vorliegt, wenn für die Datenbereitstellung ein Entgelt bzw. eine Gegenleistung erbracht wird.

34 Bobev u. a., White Paper on the Definition of Data Intermediation Services, 2023, S. 39.

35 Bobev u. a., White Paper on the Definition of Data Intermediation Services, 2023, S. 40f.

3. Die Absicht Geschäftsbeziehungen herzustellen, Art. 2 Nr. 11 DGA

Doch inwieweit ist diese Geschäftsbeziehung herzustellen? Die Formulierung „hergestellt werden sollen“ in Art. 2 Nr. 11 DGA zeigt, dass kein tatsächlicher Erfolg erforderlich ist. Die Geschäftsbeziehung muss noch nicht hergestellt worden sein. Maßgeblich ist, dass der Diensteanbieter darauf abzielt, Geschäftsbeziehungen herzustellen.³⁶ Deutlicher ist die englische Fassung – „a service which aims to“, da sie ausdrücklich auf die Zielsetzung des Dienstes abstellt. Für die Zielsetzung des Dienstes als Kriterium spricht auch der systematische Zusammenhang mit Art. 11 Abs. 1 DGA, wonach „[j]eder Anbieter von Datenvermittlungsdiensten, der beabsichtigt, die in Artikel 10 genannten Datenvermittlungsdienste „zu erbringen“, einem Anmeldeerfordernis unterliegt. Damit stellt der Gesetzgeber auf die Sicht bzw. subjektive Zielsetzung des Diensteanbieters ab. Der „zielgerichtete Wille des Diensteanbieters“³⁷ birgt als subjektives Kriterium allerdings Probleme in der Praxis. Hier sind objektive Kriterien zur Bestimmung eines solchen Willens erforderlich. Daher wird die Anwendung eines Unmittelbarkeitskriteriums vorgeschlagen. Dieses Kriterium soll die Subsumtion des Einzelfalls unter die abstrakte Begriffsbestimmung des Art. 2 Nr. 11 DGA verlässlich ermöglichen. Zum Verständnis der Erforderlichkeit eines solchen Kriteriums wird zunächst die Abgrenzungsproblematik im Kontext von Infrastrukturanbietern dargelegt. Anschließend wird auf die Ausgestaltung eines Unmittelbarkeitskriteriums eingegangen. Das so entwickelte Kriterium wird sodann anhand von Grenzfällen evaluiert.

3.1 Problemendarstellung

Bei der Abgrenzung von Datenvermittlungsdiensten und sonstigen datenbezogenen Diensten, wie Infrastrukturanbietern, geht es im Kern um die Frage, wann ein Dienst ein „ausschließlich technisches Werkzeug zur gemeinsamen Datennutzung“ im Sinne des ErwG 28 S. 3 DGA ist. Diese Problematik wird am Beispiel sogenannter Konnektoren untersucht.

Um das Problem darstellen zu können, müssen zunächst die relevanten Begrifflichkeiten erläutert werden. Mangels einheitlicher Definition wird

36 Schild u. a., in: BeckOK Datenschutzrecht, 54. Edition 2025, DGA Art. 2 Rn. 77f.

37 Schild u. a., in: BeckOK Datenschutzrecht, 54. Edition 2025, DGA Art. 2 Rn. 94.

für den Begriff des *Datenraums*³⁸ die Legaldefinition von „gemeinsamen europäischen Datenräumen“ herangezogen. Auf eine vertiefere Untersuchung des Begriffs des Datenraums wird in diesem Beitrag verzichtet. Gemäß Art. 30 lit. h DGA handelt es sich dabei um „zweck- oder sektorspezifisch[e] oder auch sektorübergreifend[e] interoperabl[e] Rahmen mit gemeinsamen Normen und Praktiken für die gemeinsame Nutzung oder Verarbeitung von Daten [...]“. Kennzeichnend sind eine föderierte, dezentrale und offene Infrastruktur sowie gemeinsame Regeln und Verfahren für den Datenaustausch, denen sich die Teilnehmenden unterwerfen.³⁹

„Konnektoren“ sind Softwarekomponenten, die den technischen Zugang zu den Diensten des Datenraums ermöglichen. Sie bilden damit das „Tor“ zum Datenraum und sollen als sicheres Gateway sowohl den Austausch der geteilten Daten als auch die Kommunikation mit anderen Diensten und Teilnehmern im Datenraum ermöglichen.⁴⁰ Daraus ergibt sich folgende Abgrenzungsfrage: Wenn ein Anbieter eine Infrastruktur für den Datenraum bereitstellt und die Teilnehmenden durch ihre Konnektoren mit Datenkatalogen, Datenbeständen oder ähnlichen Vermittlungsdiensten kommunizieren und unmittelbar Daten austauschen bzw. Zugang zu Daten gewähren können, stellt dann auch der Anbieter der Konnektoren einen Datenvermittlungsdienst bereit? Die Frage kann sich noch zugespitzter stellen, wenn als Teil des Konnektors auch Katalogdienste angeboten werden, „die sowohl Datensätze der eigenen Datenbestände speichern als auch externe Daten crawlen und cachen können“.⁴¹

Eng damit verknüpft ist auch die Abgrenzung zwischen Infrastrukturdiensten im Sinne des Art. 10 lit. a DGA und zusätzlichen Diensten zur Erleichterung des Datenaustauschs im Sinne des Art. 12 lit. e DGA⁴². Grundsätzlich können Datenvermittlungsdienste solche weiteren Tools und Dienste umfassen. Sollte das Bereitstellen von Konnektoren keinen Datenvermittlungsdienst im Sinne des DGA darstellen, würde sich die Überlegung anschließen, ob die Konnektoren als sonstiger Dienst Teil des Datenvermittlungsdienstes anzusehen sind. Dabei steht die Frage im Raum, ob auch bei sonstigen Diensten das Pflichtenregime des DGA gelten soll.

38 Sorge/Wiedemann, in: HK-DatenR, 2. Auflage 2025, Art. 33 DA Rn. 29.

39 Reiberg u. a., ‘What is a Data Space?’, Gaia-X Hub Germany, White Paper 1/2022, S. 11; Sorge/Wiedemann, in: HK-DatenR, 2. Auflage 2025, Art. 33 DA Rn. 30.

40 Siska u. a., Aufbau eines Datenraums, 2023, S. 4, 14.

41 Siska u. a., Aufbau eines Datenraums, 2023, S. 8.

42 Hartmann, Daten als Leistung, 2023, S. 294f.

Um diese Problematik aufzutrennen, wird zunächst ein Kriterium zur Abgrenzung der Datenvermittlungsdienste von sonstigen datenbezogenen Diensten untersucht. Auf der zweiten Ebene – unter der Voraussetzung, dass ein Datenvermittlungsdienst vorliegt – wird die Abgrenzung von dem Datenvermittlungsdienst zugehörigen Diensten im Sinne von Art. 10 lit. a DGA bzw. Art. 12 lit. e DGA untersucht, um ein handhabbares Prüfungssystem für die Praxis zu erörtern.

3.2 Die Abgrenzung über die Zielsetzung des Dienstes

Auf der ersten Ebene wird untersucht, wie sonstige datenbezogene Dienste von datenvermittelnden Tätigkeiten abzugrenzen sind. Wie oben dargelegt, können sich Grenzfälle ergeben, die in der Datenwirtschaft nicht immer klar einzuordnen sind. Subsumiert man etwa das Angebot von Konnektoren unter die Begriffsbestimmung des Art. 2 Nr. 11 DGA, könnte eine oberflächliche Prüfung hinsichtlich der grundsätzlichen Ermöglichung der Vermittlung von Geschäftsbeziehungen zu dem Ergebnis führen, dass ein Datenvermittlungsdienst vorläge.

Wesentlich ist bei der Subsumtion die *Absicht geschäftliche Beziehungen zum Zwecke des Datenteilens herzustellen*. Auch Anbieter von Konnektoren können eine solche Absicht verfolgen, indem die Anbieter die technische Infrastruktur und ggf. Katalogdienste zur Vernetzung der Teilnehmenden bereitstellen. Damit wird den Teilnehmenden des Datenraums ermöglicht, über weitere Dienste wie Plattformen, Marktplätze oder Kataloge Geschäfte anzubahnen.

Dennoch ist fraglich, ob der Gesetzgeber auch solche Infrastrukturdienste erfassen will. Hier reiht sich die Frage ein, ob Hauptzweck des Dienstes die Herstellung von Geschäftsbeziehungen sein muss – nicht zu verwechseln mit der Frage, ob Hauptzweck der Geschäftsbeziehung die Bereitstellung und Nutzung von Daten sein muss. Damit rückt die subjektive Zwecksetzung des Dienstes als subjektives Kriterium der Begriffsbestimmung in Art. 2 Nr. 11 DGA in den Fokus. Gegenstand ist die untergeordnete Frage, welchen Grad die Absicht des Dienstes Geschäftsbeziehungen herzustellen aufweisen muss. Genügt ein bloßer Nebenzweck oder muss die Herstellung von Geschäftsbeziehungen das Hauptziel des Dienstes betreffen?⁴³ Sachge-

43 Bobev u. a., White Paper on the Definition of Data Intermediation Services, 2023, S. 42.

recht erscheint es, die „Absicht“ anhand objektiv überprüfbarer Faktoren wie Tätigkeitsprofil und Geschäftsmodell des Anbieters zu bestimmen.⁴⁴

Teilweise wird – auch vor dem Hintergrund der Anpassung in ErwG 22 DGA-E – eine Differenzierung zwischen Haupt- und Nebenzweck abgelehnt, um Abgrenzungsprobleme und denkbare Ausweichbewegungen bei der Geschäftsmodellentwicklung zu vermeiden.⁴⁵ Dies überzeugt nur bedingt: Gerade in dezentralen Datenräumen mit eng verzahnten Diensten würde eine solche Ausweitung die rechtssichere Einordnung der Dienste weiter erschweren. Die Frage der Abgrenzung würde sich lediglich vom Haupt- auf den Nebenzweck verlagern. Vor dem Hintergrund der spezifischen Zwecke und des strengen Pflichtenregimes für Datenvermittlungsdienste erscheint es fragwürdig, Dienste zu erfassen, die Geschäftsbeziehungen nur mittelbar oder nebenbei fördern. Dies gilt etwa bei Cloud-Marktplätzen oder App-Ökosystemen, bei denen Hauptzweck die Bereitstellung von Software oder Services ist und der Austausch von Nutzungs- oder Betriebsdaten nur ein Nebeneffekt ist. Erforderlich sollte vielmehr sein, dass der Dienst gezielt für die Herstellung geschäftlicher Beziehungen zum Zwecke des Datenteilens konstruiert ist.⁴⁶ Dies ergibt sich auch durch Heranziehung von ErwG 28 S. 2 DGA. Danach soll der DGA nur solche Dienste erfassen, die *unmittelbar* die Bereitstellung von Datenvermittlungsdiensten betreffen. Die Unmittelbarkeit deutet stark auf eine direkte Herstellung hin bzw. einen „hohen Grad“ an Nähe zur Geschäftsvermittlung und damit auf einen Hauptzweck.

Es bleibt noch die Frage: Wann liegt ein Tätigkeitsprofil vor, das die Herstellung von Geschäftsbeziehungen zum Hauptzweck hat? Auch hier kann die Unmittelbarkeit bei der Bereitstellung von Datenvermittlungsdiensten möglicherweise weiterhelfen.

3.3 Das Kriterium der Unmittelbarkeit

Der Gesetzgeber bietet in ErwG 28 S. 2 DGA mit der Unmittelbarkeit ein weiteres Indiz für die Abgrenzung des Datenvermittlungsdienstes von sons-

44 Bobev u. a., White Paper on the Definition of Data Intermediation Services, 2023, S. 41f.; von Ditfurth 2024, S. 232.

45 Richter, Europäisches Datenprivatrecht, ZEuP 2021, 634 (655); von Ditfurth 2024, S. 232.

46 Vgl. auch Bobev u. a., White Paper on the Definition of Data Intermediation Services, 2023, S. 42.

tigen datenbezogenen Diensten eines Unternehmers oder sonstiger Stellen an. Damit adressiert ErWG 28 S.2 DGA primär die Differenzierung zwischen verschiedenen Diensten desselben Anbieters. Dennoch kann dieses Kriterium auch als allgemeine Auslegungshilfe für Dienste unterschiedlicher Akteure herangezogen werden.

Die Unmittelbarkeit ist dabei funktional zu verstehen: Ein Anbieter beabsichtigt die Herstellung von Geschäftsbeziehungen, wenn der angebotene Dienst die Akteure in die Lage versetzt, eine wirtschaftliche Interaktion mit anderen anzubahnen, ohne dass es weiterer wesentlicher Zwischenschritte der Akteure in Form von erheblichen Such- oder Vermittlungshandlungen bedarf.

Dieser Ansatz widerspricht auch nicht der Auslegung des Beziehungsbegriffs des Art. 2 Nr. 11 DGA, wonach noch kein konkreter Vertragsabschluss erforderlich ist. Es genügt, dass ein geschäftlicher Kontakt mit dem Ziel des Datenteilens hergestellt wird.⁴⁷ Danach kann sich die Unmittelbarkeit nicht auf einen Geschäftsabschluss beziehen, sondern nur auf die Anbahnung eines solchen. Daraus folgt, dass die Tätigkeit des Dienstes unmittelbar an der Geschäftsanbahnung ansetzen muss, aber noch keinen Vertragsabschluss herbeizuführen braucht.

Es reicht jedoch nicht aus, wenn der Dienst nur die Umstände oder Voraussetzungen dafür schafft, dass die Datenbereitsteller und Datennutzer eine Umgebung vorfinden, in der sie Daten austauschen können. Solange die beteiligten Akteure mit eigenen Mitteln wesentliche Schritte unternehmen müssen, um potenzielle Geschäftspartner zu finden, handelt es sich nicht um einen Datenvermittlungsdienst.

Dies ergibt sich insbesondere aus dem Sinn und Zweck der Dienste, das Auffinden von nutzbaren Daten zu erleichtern und die bestehenden Informationsasymmetrien in der Datenwirtschaft zu beseitigen.⁴⁸ Das Auffinden passender Vertragspartner gestaltet sich bisher schwierig. Durch die Dienste sollen die Such- und Transaktionskosten der Wirtschaftsakteure verringert werden⁴⁹, indem die Auffindbarkeit der Daten im Datenraum verbessert wird. In der Konsequenz soll der Austausch (nutzbarer) Daten gesteigert werden. Diese Zwecke sind nur dann zu erreichen, wenn der Dienst auch tatsächlich die Informationsdefizite hinsichtlich der Verfügbar-

47 von Ditfurth 2024, S. 233.

48 Specht-Riemenschneider, in: HK-DatenR, 2. Auflage 2025, DGA Art. 10 Rn. 11.

49 von Ditfurth 2024, S. 138; Schumacher/Lück, in: Paschke/Rücker, Data Governance Act: DGA, 2024, Art. 10 Rn. 4.

keit und Qualität von Daten verringert, indem er den Datenbereitstellern und -nutzern unmittelbar eine Vermittlung geschäftlicher Kontakte anbietet.

Daneben ist es Zweck des Datenvermittlungsdienstes, einen nichtdiskriminierenden Zugang zur Datenwirtschaft zu schaffen, um der ungleichen Marktmacht in der Datenwirtschaft entgegenzuwirken. Dadurch sollen „Unternehmen aller Größenordnungen – insbesondere KMU und Start-up-Unternehmen mit eingeschränkten finanziellen, rechtlichen oder administrativen Möglichkeiten“ Zugang zur Datenwirtschaft erhalten (ErwG 27 S. 4 DGA). Auch dies setzt voraus, dass der Dienst unmittelbar an der Herstellung von Geschäftsbeziehungen konkreter Parteien ansetzt. Die Ermöglichung der Geschäftsbeziehung darf also nicht zu weit im Vorfeld eines konkreten Kontakts liegen. Ansonsten wäre der Einfluss des Dienstes auf einen diskriminierungsfreien Zugang zu den datengebenden Akteuren und ihren Angeboten gehemmt. Ein weiterer Aspekt ist das strenge Pflichtenregime des Art. 12 DGA, das an die Qualifikation als Datenvermittlungsdienst anknüpft. Die Auferlegung scheint nur dann geboten, wenn der Dienst auch unmittelbar diejenigen Gefahren und Risiken schafft, die durch die Bedingungen des Art. 12 DGA abgefangen werden sollen. Die Anwendung des Art. 12 DGA erscheint nicht geboten, wenn die Herstellung von Geschäftsbeziehungen noch nicht unmittelbar durch den Dienst erbracht wird.

Im Ergebnis kann für die Frage auf der ersten Ebene, ob ein Dienst überhaupt Datenvermittlungsdienst im engeren Sinne ist – also nach Art. 2 Nr. 11 DGA Geschäftsbeziehungen herstellt – das Kriterium der Unmittelbarkeit zur Überprüfung herangezogen werden. Dabei muss eine Betrachtung des Dienstes am Einzelfall erfolgen. Der Anbieter der technischen Infrastruktur, der etwa Konnektoren bereitstellt, ermöglicht zwar mittelbar die Herstellung von Geschäftsbeziehungen durch technische Mittel, allerdings übertritt er nicht die Unmittelbarkeitsschwelle. Die Akteure müssen weiterhin eigenen Suchaufwand betreiben und sind der Gefahr ausgesetzt, von anbotsvermittelnden Diensten ausgeschlossen zu bleiben.⁵⁰

50 Ein anderes Ergebnis wäre im Hinblick auf das Angebot von Katalogdiensten bei Konnektoren denkbar. Das ist Gegenstand weiterer Forschung und kann an dieser Stelle nicht näher ausgeführt werden.

3.4 Die Abgrenzung von zugehörigen Diensten

Bei der Frage, wie die einzelnen Dienste und Werkzeuge der Datenwirtschaft im Verhältnis zum DGA stehen, wurde die erste Ebene bereits betrachtet: Was ist ein Datenvermittlungsdienst im (engeren) Sinne – also wann wird unmittelbar die Herstellung von Geschäftsbeziehungen angeboten?

Daran schließt sich die zweite noch zu betrachtende Ebene an: Der Gesetzgeber erwähnt an mehreren Stellen (Art. 10 lit. a, b DGA, Art. 12 lit. e DGA), dass zu dem Vermittlungsdienst auch weitere Mittel und Dienste vom Datenvermittlungsdienst gehören bzw. von diesem umfasst sind. Damit sind andere (datenbezogene) Dienste nicht nur von Art. 2 Nr. 11 DGA abzugrenzen, sondern auch von weiteren Diensten, die Teil des Datenvermittlungsdienstes sein können (zugehörige bzw. umfasste Dienste) und damit möglicherweise dem Pflichtenkatalog des DGA unterfallen.

Zunächst werden in Art. 2 Nr. 11 DGA technische, rechtliche oder sonstige Mittel aufgezählt. Diese beziehen sich nach dem Wortlaut der Norm unmittelbar auf die Herstellung der Geschäftsbeziehung. Folglich ist der „Geschäftsbeziehungs-Vermittlungsdienst“ im engeren Sinne gemeint. In Art. 10 lit. a, b DGA stellt der Gesetzgeber klar, dass Vermittlungsdienste auch die „Bereitstellung der technischen oder sonstigen Mittel als Voraussetzung solcher Dienste“ einschließen können. Dies ist wie im Rahmen des Art. 2 Nr. 11 DGA als Bereitstellung des Vermittlungsdienstes im engeren Sinne zu verstehen.⁵¹ In der englischen Fassung wird „to enable“ verwendet; gemeint sind solche technischen und sonstigen Mittel, die den Vermittlungsdienst im engeren Sinne ermöglichen.

Die Vermittlungsdienste können nach Art. 10 lit. a DGA zudem den „zwei- oder mehrseitige[n] Austausch von Daten oder die Einrichtung von Plattformen oder Datenbanken, die den Austausch oder die gemeinsame Nutzung von Daten ermöglichen sowie die Einrichtung anderer spezieller Infrastrukturen für die Vernetzung von Dateninhabern mit Datennutzern“ beinhalten.

In Bezug auf den zwei- oder mehrseitigen Austausch von Daten ist zu klären, was der Gesetzgeber hiermit ausdrücken wollte. In der Literatur wird unter dem Begriff des „Austauschs“ eine synallagmatische Leistungspflicht verstanden. Diskutiert wird hierbei die Frage, ob es sich um eine

51 Schwamberger, in: Spindler/Schuster/Kaesling, 5. Auflage 2026, DGA Art. 10 Rn. 8.

Klarstellung dahingehend handelt, dass sich die Leistungspflicht auch auf einen Tausch „Daten gegen Daten“ beziehen kann.⁵² Diese Auslegung des Begriffs deutet darauf hin, dass sich die Klarstellung auf die Vertragsbeziehung zwischen Datenbereitsteller und Datennutzer bezieht.

Dies erscheint nicht zwingend. Der „zwei- oder mehrseitigen Austausch von Daten“ kann auch als faktischer Akt im Sinne einer Übermittlung der Daten bzw. eines Zugangs zu Daten verstanden werden. Dafür spricht die Systematik des Art. 10 lit. a Hs. 2 DGA. Sowohl die Einrichtung von Plattformen und Datenbanken zur Ermöglichung des Datenaustauschs als auch die Einrichtung sonstiger Infrastrukturen als Auffangtatbestand beziehen sich auf die Bereitstellung technischer (oder sonstiger) Infrastruktur durch den Datenvermittlungsdienst. Die Klarstellung, auch der „Austausch von Daten“ könne eingeschlossen sein („may include“), legt nahe, dass auf die Unterstützung bei der Übermittlung der Daten Bezug genommen wird. Gegenstand der Datenvermittlungsdienste kann somit auch die Ermöglichung des faktischen Akts der Übermittlung sein, nicht nur die Vermittlung der geschäftlichen Beziehung. Im Ergebnis können Datenvermittlungsdienste auch sonstige (Infrastruktur-)Dienste beinhalten, die entweder den Austausch oder die gemeinsame Nutzung von Daten ermöglichen oder Dateninhaber und -nutzer miteinander vernetzen. Damit handelt es sich um akzessorische Dienste, die mit der Erbringung des Datenvermittlungsdienstes verbunden sind, aber nicht unmittelbar Geschäftsbeziehungen herstellen.

Hieran schließt sich die untergeordnete Frage an, ob diese zugehörigen Dienste als Teil des Datenvermittlungsdienstes den strengen Regelungen des DGA unterfallen sollen. Jedenfalls Art. 10 lit. a Hs. 2 DGA bringt dies zum Ausdruck, indem die aufgezählten Einrichtungen den Datenvermittlungsdiensten zugeordnet werden („zu diesen Diensten können [...] gehören“ bzw. „those services may include“). Eine Einschränkung ergibt sich, wenn man ErwG 28 S. 3 des DGA bei der Auslegung berücksichtigt. Danach sollen ausschließlich technische Werkzeuge zum Datenteilen nicht als Datenvermittlungsdienst gelten, sofern die Bereitstellung dieser Werkzeuge weder auf die Herstellung einer geschäftlichen Beziehung abzielt noch dem Anbieter von Datenvermittlungsdiensten ermöglicht, Informationen über die Herstellung geschäftlicher Beziehungen zu erlangen. Jedenfalls am zweiten Kriterium – der Informationsermöglichung – dürfte ein Ausschluss der Dienste im Regelfall scheitern.

52 Schwamberger, in: Spindler/Schuster/Kaesling, 5. Auflage 2026, DGA Art. 10 Rn. 11.

Dies kann am Fall des Konnektors veranschaulicht werden. Der Anbieter der technischen Infrastruktur qualifiziert sich allein nicht als Datenvermittlungsdienst. Bietet jedoch ein Anbieter neben einem Datenvermittlungsdienst auch die technische Infrastruktur – insbesondere Konnektoren – an, um den zwei- oder mehrseitigen Austausch von Daten zu ermöglichen, werden auf der Steuerungsebene des Konnektors weitergehende Informationen u.a. über die geschäftlichen Beziehungen beobachtet. Dazu zählen etwa Identitäts- und Zugriffsmanagement, Bearbeitung von Angeboten, Erstellen, Aushandeln und Abrechnen von Verträgen sowie Protokollierung.⁵³ Damit wird dem Anbieter des Datenvermittlungsdienstes durch den Infrastrukturdienst die Erlangung von Informationen über die Herstellung geschäftlicher Beziehungen ermöglicht.

Schließlich gestattet der Gesetzgeber den Datenvermittlungsdiensten in Art. 12 lit. e Hs. 1 DGA, dass sie „ein Angebot zusätzlicher spezifischer Werkzeuge und Dienste für Dateninhaber oder betroffene Personen umfassen [können], insbesondere um den Datenaustausch zu erleichtern“. Hinzuweisen ist hier auf ein weiteres Redaktionsversehen des Gesetzgebers. Die englische Sprachfassung schränkt die Werkzeuge und Dienste auf den spezifischen Zweck der Erleichterung des Datenaustauschs ein („for the specific purpose of facilitating the exchange of data“). Die deutsche Fassung erklärt hingegen mit dem Wort „insbesondere“, dass auch andere Zwecke verfolgt werden können.⁵⁴ Dies kann vom Gesetzgeber kaum gewollt sein, wird der Datenvermittlungsdienst doch sonst einer strengen Zweckbindung unterworfen und dient dieser doch der Vermittlung potenzieller Akteure zum Zweck des Datenaustauschs.

Weiterhin stellt sich die Frage, in welchem Verhältnis die zusätzlichen spezifischen Dienste nach Art. 12 lit. e Hs. 1 DGA zu den datenbezogenen Diensten im Sinne des ErwG 28 S. 2 des DGA stehen.⁵⁵

Man kann den DGA so auslegen, dass die zusätzlichen Dienste und Werkzeuge im Sinne des Art. 12 lit. e DGA als Teil des Datenvermittlungsdienstes anzusehen sind.⁵⁶ Dafür spricht zum einen, dass Art. 12 lit. e Hs. 2

53 Siska u. a., *Aufbau eines Datenraums*, 2023, S. 4.

54 Schröder, in: BeckOK Datenschutzrecht, 54. Edition 2025, DGA Art. 12 Rn. 45.1; Schumacher/Lück, in: Paschke/Rücker, *Data Governance Act: DGA*, 2024, Art. 12 Rn. 23.

55 Bobev u. a., *White Paper on the Definition of Data Intermediation Services*, 2023, S. 45.

56 Bobev u. a., *White Paper on the Definition of Data Intermediation Services*, 2023, S. 45.

DGA nur in Bezug auf die (vom Vermittlungsdienst) angebotenen Werkzeuge Dritter eine Zweckbindung normiert, um die strenge Zweckbindung des Art. 12 lit. a DGA auch in diesem Fall aufrechtzuerhalten.⁵⁷ Zum anderen kann hier der Wortlaut des DGA angeführt werden, der von „können umfassen“ („may include“) spricht.⁵⁸

Die Abgrenzung zu den zugehörigen Diensten im Sinne des Art. 10 lit. a Hs. 2 DGA ergibt sich schließlich aus der Zwecksetzung der zusätzlichen spezifischen Werkzeuge, die gerade nicht der Ermöglichung – wie in Art. 10 DGA –, sondern nur der Erleichterung des Datenaustauschs dienen.

4. Fazit und Ausblick

Die Analyse der Begriffsbestimmung des Art. 2 Nr. 11 DGA hat gezeigt, dass der Begriff des *Datenvermittlungsdienstes* weiterhin erheblichen Auslegungsbedarf aufweist. Insbesondere im Hinblick auf das Merkmal der *Geschäftsbeziehungen* wird deutlich, dass die offene und unklare Formulierung des Gesetzgebers erheblichen Konkretisierungsbedarf hervorruft. Zur weiteren Konkretisierung ist eine beziehungsorientierte Auslegung geboten. Bei der Abgrenzung zwischen Datenvermittlungsdiensten und Infrastrukturanbietern ist darüber hinaus der Fokus auf die Auslegung der „Herstellungsabsicht“ zu setzen, die anhand der Unmittelbarkeit bei der Bereitstellung von Datenvermittlungsdiensten bestimmt werden kann. Insofern wird ein Beitrag zur weiteren Konkretisierung des Begriffs „Datenvermittlungsdienst“ geleistet.

Die Digital-Omnibus-Verordnung-E⁵⁹ erkennt nun ausdrücklich die Schwächen der Begriffsbestimmung „Anbieter von Datenvermittlungsdiensten“ an (ErwG 7 Digital-Omnibus-Verordnung-E). Eine wesentliche Klarstellung könnte in der Anpassung des Begriffs der „Geschäftsbeziehung“ liegen, der nun als „Geschäftsbeziehungen wirtschaftlicher Natur“ in Art. 2 Nr. 38a Digital-Omnibus-Verordnung-E zu finden ist. Erneut ist die englische Version deutlicher, die „relationships of an economic character“ lautet und sich damit von dem „Geschäft“ („commercial“) trennt. Hier wird

57 Schröder, in: BeckOK Datenschutzrecht, 54. Edition 2025, DGA Art. 12 Rn. 48.

58 Bobev u. a., White Paper on the Definition of Data Intermediation Services, 2023, S. 45.

59 Vorschlag für eine Digital-Omnibus-Verordnung, COM(2025) 837 final, 2025/0360 (COD), eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52025PC0837 (Link zuletzt abgerufen am 30.01.26).

der Bezug der Wirtschaftlichkeit zur Beziehung – und nicht etwa zu den Akteuren und der Datennutzung – deutlicher.

Der Gesetzgeber sollte die deutsche Sprachfassung hier nicht vernachlässigen und sich vom missverständlichen Begriff der „Geschäftsbeziehung“ verabschieden. Wünschenswert wäre insbesondere eine mit den anderen Fassungen kohärente deutsche Sprachfassung.

Literatur

- Beise, Clara (2021): Datensouveränität und Datentreuhand. *Recht Digital* 2021, S. 597–604.
- Bobev, Tervel; Dessers, Viltè Kristina; Ducuing, Charlotte; Fierens, Michiel; Palumbo, Andrea; Peeters, Bert und Stähler, Leander (2023): White Paper on the Definition of Data Intermediation Services. CiTiP White Paper 2023. Leuven: KU Leuven Centre for IT & IP Law. <https://doi.org/10.2139/ssrn.5397118>.
- von Ditfurth, Lukas (2024): *Datenmärkte, Datenintermediäre und der Data Governance Act – Eine Analyse der europäischen Regulierung von B2B-Datenvermittlungsdienste*, Berlin/Boston: de Gruyter.
- Gola, Peter und Heckmann, Dirk (Hrsg.) (2022): *Datenschutz-Grundverordnung – Bundesdatenschutzgesetz*, 3. Auflage, München: C.H.Beck (zitiert: Bearbeiter, in: Gola/Heckmann, 3. Auflage 2022).
- Grabitz, Eberhard (Begr.); Hilf, Meinhard; Nettesheim, Martin (Hrsg.) (2025): *Das Recht der Europäischen Union Band I EUV/AEUV*, 86. Ergänzungslieferung, München: C.H.Beck (zitiert: Bearbeiter, in: Das Recht der EU, 86. EL 2025).
- Hartmann, Matthias (2023): Daten als Leistung: Datenverträge. In: Beatrix Weber (Hrsg.): *Data Governance: Nachhaltige Geschäftsmodelle und Technologien im europäischen Rechtsrahmen*. Berlin, Heidelberg: Springer, S. 267–317. https://doi.org/10.1007/978-3-662-67556-4_10.
- Hoppe, Thomas und Fricke, Rolf (Hrsg.) (2025): *Resiliente krisenrelevante Versorgungsnetze Einsatz von KI in der Krisen- und Katastrophenbewältigung*, Wiesbaden: Springer. <https://doi.org/10.1007/978-3-658-48639-6>
- Paschke, Anne und Rücker, Daniel (Hrsg.) (2024): *Data Governance Act: DGA*, 1. Auflage, München: C.H.Beck (zitiert: Bearbeiter, in: Paschke/Rücker, Data Governance Act: DGA, 2024).
- Reiberg, Abel; Niebel, Crispin und Kraemer, Peter (2022): What is a Data Space? Definition of the concept Data Space, Whitepaper, Gaia-X Hub Germany 2022.
- Richter, Heiko (2021): Europäisches Datenprivatrecht: Lehren aus dem Kommissionsvorschlag für eine „Verordnung über europäische Daten-Governance“. *Zeitschrift für Europäisches Privatrecht* 2021, Heft 3, S. 635–667.
- Richter, Heiko (2025): Datenvermittlungsdienste gem. Art. 2 Nr. 11 DGA: Begriffsschärfung mithilfe eines Seitenblicks auf öffentliche Stellen. *Zeitschrift für Europäisches Daten- und Informationsrecht*, 2025, Heft 5, S. 264–270.

- Schleimer, Anna Maria; Altendeitering, Marcel; Hupperz, Marius und Steinert, Michael (2025): Datenräume als Basis für neue Akteure in der Mobilität. In: Proff, Heike (Hrsg.): *New Players in Mobility Technische und betriebswirtschaftliche Aspekte*, 16. *Wissenschaftsforum Mobilität*, S. 552–565.
- Schwartmann, Rolf; Benedikt, Kristin und Reif, Yvette (Hrsg.) (2025): *Datenschutz im Internet*, 1. Auflage, München: C.H.Beck (zitiert: Bearbeiter, in: Schwartmann/Benedikt/Reif, 1. Auflage 2025).
- Schwarze, Jürgen; Hatje, Armin und Terhechte, Jörg Philipp (Hrsg.) (2026): *Europäisches Unionsrecht*, 8. Auflage, Baden-Baden: Nomos (zitiert: Bearbeiter, in: Europäisches Unionsrecht, 8. Auflage 2026).
- Siska, Veronika; Karagiannis, Vasileios und Drobits, Mario (2023): Aufbau eines Datenraums: Technischer Überblick, Whitepaper, Gaia-X Hub Austria 2023.
- Specht-Riemenschneider, Louisa und Hennemann, Moritz (Hrsg.) (2025): *Data Act Data Governance Act*, 2. Auflage, Baden-Baden: Nomos (zitiert: Bearbeiter, in: HK-DatenR, 2. Auflage 2025).
- Spindler, Gerald; Schuster, Fabian und Kaesling, Katharina (Hrsg.) (2026): *Recht der elektronischen Medien*, 5. Auflage, München: C.H.Beck (zitiert: Bearbeiter, in: Spindler/Schuster/Kaesling, 5. Auflage 2026).
- Tolks, Daniel (2022): Die finale Fassung des Data Governance Act. Erste Schritte in Richtung einer europäischen Datenwirtschaft. *Zeitschrift für IT-Recht und Recht der Digitalisierung (MMR)*, 2022, Heft 6, S. 444–449.
- Wolff, Heinrich Amadeus; Brink, Stefan und v. Ungern-Sternberg, Antje (Hrsg.) (2025): *BeckOK Datenschutzrecht DS-GVO, DA, DGA, BDSG. Datenschutz und Datennutzung*, 54. Edition, München: C.H.Beck (zitiert: Autor, in: BeckOK Datenschutzrecht, 54. Edition 2025).

Maßnahmen zur Altersverifikation in Metaversen: Ethische, rechtliche und sicherheitstechnische Aspekte

*Jessica Heesen, Thomas Kunz, Matthias Meitzler, Maria Pawelec,
Marina Steenbergen und Ulrich Waldmann*

Zusammenfassung

Metaversen ermöglichen immersive, digitale Interaktionen, bergen jedoch vor allem für Kinder und Jugendliche besondere Risiken, da sie u. a. einen unkontrollierten Zugang zu nicht altersgerechten Inhalten und Interaktionen begünstigen. Der Beitrag beleuchtet Altersverifikationen in Metaversen aus ethischer sowie rechtlicher Perspektive und diskutiert die Anforderungen des Jugend- und Datenschutzrechts sowie grundlegende Prinzipien wie informationelle Selbstbestimmung, digitale Souveränität und Nichtdiskriminierung. Vorgestellt wird ein sicherheitstechnischer Ansatz auf Basis selbstbestimmter digitaler Identitäten, der pseudonyme Nachweise von Altersgruppen ermöglicht, ohne in Metaversen reale Identitäten offenzulegen. Er schafft damit einen angemessenen Rahmen für den Schutz Minderjähriger in transnationalen Metaversen.

1. Einleitung

Metaversen sind persistente, virtuelle Welten, die – häufig mit Hilfe von Virtual-Reality (VR)-Zubehör – eine neue Form der digitalen Vernetzung und Interaktion ermöglichen. In ihnen agieren Nutzende als *Avatare*, d. h. als verkörperte virtuelle Repräsentationen ihrer selbst, die sie in Echtzeit steuern. Typische Anwendungen umfassen u. a. Gaming, virtuelle Events, Bildungsangebote, simulationsgestützte Produktentwicklung und interaktiven Austausch im Allgemeinen. Charakteristisch für Metaversen ist das hohe Maß an *Immersion*, wodurch Nutzende in die virtuelle Welt eintauchen.

Aufgrund ihrer innovativen Form virtueller Präsenz, Immersion und Interaktion sowie fehlender Moderationsstrukturen bergen Metaversen neuartige, den Datenschutz sowie die (informationelle) Selbstbestimmung, die Identität und die Sicherheit der Nutzenden in besonderem Maße betreffen-

de Risiken (Al-kfairy u. a. 2025), welche über bereits bekannte Problemlagen des Internets und sozialer Medien hinausgehen (Croll/Krause 2024). Hierzu gehört beispielsweise die illegitime Aneignung von Avataren oder der Einsatz von rein KI-gesteuerten Avataren, die sich als menschliche Akteure ausgeben und Nutzende gezielt manipulieren können.

Verifikationsmaßnahmen können einen wichtigen Beitrag leisten, um diesen Risiken zu begegnen. Sie eignen sich etwa zur Einschätzung anderer Nutzenden, u. a. ob es sich um den Avatar einer realen Person oder einen KI-gesteuerten Avatar (Kunz/Waldmann 2025a) handelt, oder besonders des Alters eines Nutzenden durch eine entsprechende Altersverifikation.

Schätzungsweise zwei Drittel aller Nutzenden von Metaversen sind derzeit jünger als 16 Jahre (Standards Australia 2024). Als „Early Adopters“ agieren Kinder und Jugendliche mit Blick auf Medien häufig offener und risikobereiter und sind dabei mitunter weniger stark für mögliche Gefahren sensibilisiert als Erwachsene. In der Folge können sie von den negativen Auswirkungen der Mediennutzung stärker betroffen sein (vgl. Livingstone u. a. 2019, S. 4; Stapf/Heesen 2026). Die Vulnerabilität und die besondere Schutzbedürftigkeit von Kindern und Jugendlichen rücken vor diesem Hintergrund die Frage nach strukturellen Vorkehrungen im Sinne altersangemessener Zugangs- und Interaktionskontrollen in den Fokus.

Auf der Grundlage spezieller Risiken von Metaversen für Kinder und Jugendliche sowie qualitativer Interviews mit Expertinnen und Experten wird in diesem Beitrag zunächst die Legitimität entsprechender (Alters-)Verifikationsmaßnahmen aus ethischer Perspektive diskutiert. Im Anschluss folgen eine Darstellung der rechtlichen Anforderungen an (Alters-)Verifikationsmaßnahmen und die Beschreibung eines technischen Lösungsansatzes, mit dem das Alter der Nutzenden sicher und datensparsam verifiziert, der Zugang zu bestimmten Angeboten wirksam eingeschränkt und der Schutz von Privatheit verbessert werden kann. Den Abschluss bilden eine Berücksichtigung möglicher Grenzen von Altersverifikationen, die Ermittlung von Verantwortlichkeiten zur Umsetzung der Maßnahmen sowie ein Fazit.

2. Empirische und methodische Grundlagen

Der vorliegende Beitrag basiert sowohl auf umfassenden ethischen, rechtlichen und technischen Literaturstudien als auch semi-strukturierten, qualitativen Interviews mit Expertinnen und Experten, die im Rahmen des For-

schungsprojekts „Privatheit, Ethik und IT-Sicherheit in Metaversen“ (Pri-Meta; 2024-2027) durchgeführt wurden.* Leitfaden-gestützte Interviews ermöglichen eine Erfassung des „Bezugsrahmen[s] des Befragten“ und erlauben Einsichten in deren „Relevanzstrukturen und die Erfahrungshintergründe“ solcher (Schnell u. a. 2008, S. 387).

Die insgesamt 20 Interviews wurden unter einer explorativen, qualitativen Herangehensweise zwischen Oktober 2024 und November 2025 geführt, u. a. mit Expertinnen und Experten für Kinder- und Jugendmedienschutz, Medienkompetenzbildung sowie mit Personen, die sich auf die Nutzung von virtuellen Realitäten durch vulnerable Gruppen wie alte Menschen oder Menschen mit Behinderung spezialisiert haben. Die Befragten sollten ein möglichst breites Feld an Stakeholdern abdecken, die gesellschaftlich relevante Anwendungen in Metaversen (weiter-)entwickeln oder selbst nutzen. Kernthemen der Interviews waren Erfahrungen der Befragten mit Metaversen, ihre Einschätzung aktueller und künftiger technischer Entwicklungen, Chancen und Gefahren der Anwendungen, u. a. für den Datenschutz und gesellschaftliche Teilhabe. Die Interviewleitfäden wurden dabei teilweise an die jeweiligen Schwerpunkte der Befragten angepasst. Tab. 1 gibt einen Überblick über die geführten Interviews.

Tab. 1: Qualitative Interviews mit Expertinnen und Experten

Interview	Befragte Person	Hauptthema	Datum
I1	Vertreterin aus dem Bereich Tod und Digitalisierung	Entwicklung virtueller Trauerräume	08.10.2024
I2	Vertreter einer gemeinnützigen Stiftung	Entwicklung von Angeboten für ältere Menschen in VR	15.10.2024
I3	Experte für digitale Barrierefreiheit	Inklusion in und Zugänglichkeit von Metaversen	22.10.2024
I4	Unternehmensberaterin	Rolle von VR in Unternehmensentwicklung	05.12.2024
I5	Ev. Pfarrer	Digitalisierung und VR im Bereich Kirche und Trauer	06.12.2024
I6	Online-Rollenspielerin	Identität und Körperlichkeit im Kontext von Avataren	14.01.2025

* Die betreffenden Forschungsarbeiten wurden vom Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) sowie vom BMFTR und dem Hessischen Ministerium für Wissenschaft und Kunst (HMWK) im Rahmen ihrer gemeinsamen Förderung für das Nationale Forschungszentrum für angewandte Cybersicherheit ATHENE unterstützt.

Interview	Befragte Person	Hauptthema	Datum
I7	VR-Designerin	Entwicklung von VR-Anwendungen	31.01.2025
I8	Ev. Pfarrer für Kirche im digitalen Raum	Kirche und Spiritualität in Minecraft	13.02.2025
I9	Experte für Lernbehinderungen	VR in der (Aus-)Bildung und Chancen für Inklusion	25.02.2025
I10	Kommunalvertreter	Kommunen in Metaversen	28.04.2025
I11	VR-Entwickler	Kommunen in Metaversen	23.05.2025
I12	Experte für Bürgerbeteiligung	Bürgerbeteiligung und Partizipation in Metaversen	22.07.2025
I13	Experte für Kinder- und Jugendmedienschutz	Kinder- und Jugendmedienschutz in Metaversen	07.08.2025
I14	Labengineer im Bereich Augmented and Virtual Reality Design	Chancen und Risiken von AR und VR	07.08.2025
I15	VR-Künstlerin	Kunst und Medienbildung zu VR und Metaversen	04.09.2025
I16	Vertreter eines Branchenverbands	Industrie und Metaversen	17.10.2025
I17	Digitalpolitiker	Politische Implikationen und Aufmerksamkeit für Metaversen	21.10.2025
I18	Zwei Expert:innen für Kinder- und Jugendmedienschutz	Kinder- und Jugendmedienschutz in Metaversen	19.11.2025
I19	Experte für Datenschutz	Privatheit und Datenschutz in Metaversen	21.11.2025
I20	Digital- und Politikberater	Politik und demokratische Prozesse in Metaversen	24.11.2025

Die Interviews wurden mit Zustimmung der Befragten aufgezeichnet, transkribiert und mit Hilfe der qualitativen Analysesoftware *MAXQDA* codiert. Die Entwicklung der insgesamt 155 Kategorien (Codes) erfolgte sowohl deduktiv auf Grundlage der Interviewleitfäden als auch induktiv durch die Auswertung des Materials. Die Codes beziehen sich insbesondere auf technische Grundlagen und Herausforderungen von Metaversen, ebenso wie auf aktuelle Trends, die Akzeptanz, das Immersionserleben, gesellschaftliche, wirtschaftliche und politische Chancen und Gefahren (vor allem im Hinblick auf Gerechtigkeit und Inklusion), die Sicherheit, psychische Integrität, Privatheit und Datenschutz, Identität, Körperlichkeit und Avatargestaltung, Zukunftsvisionen sowie Handlungsempfehlungen. Insgesamt

wurden über 1.400 Textstellen codiert. Der vorliegende Beitrag bezieht sich hauptsächlich auf die Aussagen der Befragten zur Altersverifikation und zur Vulnerabilität von Kindern und Jugendlichen in Metaversen, aber auch auf die Chancen, die solche immersiven Umgebungen jungen Menschen bieten.

3. Nutzungsrisiken von Metaversen für Kinder und Jugendliche

Die Nutzung von Metaversen birgt verschiedene Gefahren für Kinder und Jugendliche, die teils auch aus den sozialen Medien oder Online-Games bekannt sind. So sind Kinder und Jugendliche aufgrund ihrer noch nicht vollständig ausgeprägten Fähigkeit zur Selbstregulation besonders anfällig für exzessive und suchtähnliche Nutzungsweisen (Kim 2023). In Metaversen und Online-Gaming-Umgebungen ist hier bedeutsam, dass sich die Auswahl und Gestaltung von Avataren auch auf die Identitätsgestaltung junger Nutzenden auswirken (Fu u. a. 2025). Dies macht den potenziell unkontrollierten Erwerb virtueller Güter wie Kleidung oder Accessoires in Metaversen attraktiv.

Weitere mögliche Risiken sind Mobbing, Cyberbullying, Grooming und sexuelle Übergriffe, die sich aus den Interaktionen in Metaversen ergeben (Fiani u. a. 2024). Zwar sind diese Risiken auch aus sozialen Netzwerken bekannt (vgl. Europäische Kommission 2025, S.1), doch werden sie in Metaversen durch affektiv wirkende Avatar-Interaktionen, die Immersion und das Embodiment (also die Verschmelzung des menschlichen Bewusstseins mit dem virtuellen Körper in der VR; vgl. Wienrich 2023, S. 59) als besonders eindringlich erlebt.

In Metaversen ist zudem weder das tatsächliche Alter der Nutzenden von Avataren erkennbar, noch ist klar ersichtlich, ob sich dahinter überhaupt eine reale Person verbirgt. Insbesondere für Kinder und Jugendliche besteht dadurch die Gefahr, intensive parasoziale Beziehungen zu KI-basierten Avataren aufzubauen. Dies kann emotionale Abhängigkeiten fördern und manipulative Einflussnahmen, beispielsweise in Bezug auf Konsum oder politische bzw. weltanschauliche Einstellungen, begünstigen.

Kinder und Jugendliche weisen darüber hinaus häufig nur ein geringes Bewusstsein für Datenschutzrisiken auf. Dies ist in Bezug auf Metaversen besonders problematisch, da bei Nutzung von VR-Zubehör neben Verhaltensdaten auch biometrische Daten (z. B. Eye-Tracking, Stimm- und Be-

wegungsprofile, Herzfrequenz) sowie Informationen über die physische Umgebung erfasst werden (vgl. UNICEF 2023, S. 20).

4. Ethische Aspekte der Altersverifikation

Aus ethischer Sicht dienen Systeme zur Altersverifikation nicht nur dem Schutz von Kindern und Jugendlichen, sondern auch ihrer Teilhabe und Befähigung. Kinder sollten dazu befähigt werden, eine Kommunikationssituation bewerten zu können, um selbstbestimmt und kreativ in Metaversen aktiv zu sein. Altersverifikation kann, neben anderen Maßnahmen, dazu einen Beitrag leisten. Sie muss jedoch bestimmte ethisch begründete Kriterien erfüllen, um individuelle Rechte und gesellschaftliche Zielvorstellungen für ein sicheres und selbstbestimmtes Aufwachsen in digitalen Umgebungen zu wahren.

4.1 Schutz, Befähigung und Teilhabe

Vor dem Hintergrund der oben genannten Nutzungsrisiken gilt es, Kinder und Jugendliche in Metaversen bestmöglich zu schützen. Laut UN-Kinderrechtskonvention haben Kinder jedoch nicht nur ein Recht auf Schutz und Privatheit, sondern auch auf Befähigung und Teilhabe (Stapf u. a. 2023).

Medienmündigkeit ist ein wesentlicher Kern von Selbstbestimmung in digitalen Kontexten und umfasst die selbstbestimmte Nutzung, Produktion und kritische Reflexion medialer Angebote, Plattformen und Inhalte durch Heranwachsende. Daraus folgt, dass Kinder und Jugendliche zu einem selbstbestimmten Umgang mit neuen Technologien und der Partizipation an Angeboten wie Metaversen befähigt werden sollten.

In diesem Sinne werden aktuell verschiedene Lösungen diskutiert, u. a. Nachweise über die Herkunft von Avataren (Council of Europe and IEEE 2024; Witzleb 2025), altersgerechte Räume mit beschränktem Erwachsenenzugang oder die Einführung von Altersgrenzen für spezifische Angebote (Niestadt 2024). Die letztgenannten Ansätze sollen den unterschiedlichen Entwicklungsständen von Kindern und Jugendlichen ebenso gerecht wer-

den wie dem durchaus berechtigten Interesse, mit Erwachsenen in Kontakt zu treten (Brailovskaia u. a. 2025).¹

Einige Metaversen verfügen bereits über Bereiche, die nur von Personen einer gewissen Altersgruppe genutzt werden sollen, beispielsweise „ab 18 Jahren“-markierte Räume in *Metas Horizon Worlds*, in denen u. a. „intensive Gewalt, sexuelle Inhalte, vulgäre Sprache und/oder reglementierte Güter wie Tabak oder Alkohol“ präsent sein können (Meta 2025). In *VRChat* finden sich wiederum Clubs zum Feiern und Tanzen, die ebenfalls nur ab 18 Jahren zugänglich sind (Krell/Wettmann 2023). Dabei stellt die Plattform sogenannte Badges aus, die für andere Nutzende gemeinsam mit dem Namen des Avatars sichtbar sind und ihnen ermöglichen, nur mit Volljährigen zu interagieren. Dennoch mangelt es solchen Räumen bislang an geeigneten technischen Verifikationsmaßnahmen, um das Alter der Nutzenden sicher festzustellen, bzw. um altersgerechte Zugänge zu schaffen. Entsprechende Maßnahmen wären dann „nicht nur zu begreifen als Durchsetzungsinstrument für harte Alters- und Zugangssperren, sondern [...] eher als Schlüssel, um Voreinstellungen zu steuern, die [...] altersangemessene [sowie anonyme] Nutzungserfahrungen ermöglichen“ (I13).

In sogenannten „Walled Garden“-Formaten sind Austausch, Experimentieren und Befähigung von Kindern und Jugendlichen unter geschützten Bedingungen möglich. Zugänge können dann altersdifferenziert geöffnet werden – bei jüngeren Kindern unter elterlicher Begleitung, während ältere Jugendliche zunehmend eigenverantwortlich handeln können. Ein solcher gestufter Ansatz führt „schrittweise an eine gesunde, eigenverantwortliche und gemeinschaftsfördernde Mediennutzung“ heran (I13), womit Teilhaberechte und informationelle Selbstbestimmung gewahrt bleiben.

Pauschale Zugangsverbote für Personen unter einer bestimmten Altersgrenze zu sämtlichen Metaversen – analog zu dem kürzlich in Australien in Kraft getretenen (und in Deutschland aktuell diskutierten) Social-Media-Verbot für unter 16-Jährige (Stapf/Hecktor 2026) – werden der oben definierten Zielsetzung von Schutz, Befähigung und Teilhabe indes nicht gerecht. Kindern und Jugendlichen blieben hierdurch möglicherweise positive und für ihre Entwicklung wichtige Erfahrungen verwehrt. Zudem besteht die Gefahr eines hemmenden Einflusses auf die kindliche Befähigung zum kritischen Umgang mit der Technologie, die sonst nur noch aus theoretischer Distanz erworben werden könnte (I13). Weiterhin werden durch

1 Als weiterer Ansatz wird die virtuelle Begleitung von Kindern und Jugendlichen durch Erwachsene in Form eines Avatars vorgeschlagen (I14).

solche pauschalen Ausschlüsse die „Erziehungsrechte und -möglichkeiten von Eltern und Erziehungsberechtigten“ potenziell unterlaufen (Dreyer 2025).

Aus ethischer Sicht dienen Systeme zur Altersverifikation somit nicht nur dem Schutz von Kindern und Jugendlichen, sondern ermöglichen auch ihre Teilhabe an altersgerechten Angeboten und ihre Befähigung zu einem kritischen Umgang mit Metaversen.

4.2 Normative Kriterien für die Ausgestaltung der Altersverifikation

In der digitalen Kommunikation ist die Ausgestaltung von Infrastrukturen und Diensten ein zentraler Faktor für individuelle und gesellschaftliche Handlungsoptionen. Designlösungen bestimmen, welche Entscheidungen Nutzende von Metaversen überhaupt treffen können. Systeme zur Altersverifikation müssen daher ausreichend begründete Kriterien erfüllen, um individuelle Rechte und ethisch akzeptable Wertvorstellungen zu wahren.

Zentral ist zunächst, dass ihre Einführung nicht zur Etablierung einer *Überwachungsinfrastruktur* führen darf. Ein flächendeckender, verpflichtender Einsatz von Altersverifikationssystemen wäre unverhältnismäßig, da er „Eingriffe in die Informations- und Kommunikationsfreiheiten aller Nutzenden“ bedeuten würde (Dreyer 2025). Denn je intensiver Plattformen die Identität oder das Alter ihrer Nutzenden überprüfen müssen, desto stärker werden anonyme oder pseudonyme Teilhabeformen eingeschränkt – Formen, die für demokratische Partizipation, freie Meinungsäußerung und insbesondere für vulnerable Gruppen von erheblicher Bedeutung sind. Damit ergeben sich drei wesentliche Voraussetzungen für die Etablierung von (Alters-)Verifikationssystemen:

Erstens braucht es einen *risikobasierten Ansatz*, sodass Altersverifikationen nur dann eingesetzt werden dürfen, wenn entsprechende Risiken für die Entwicklung von Kindern und Jugendlichen bestehen. Sie fungieren dann weniger als Schranken, die Nutzende von bestimmten Inhalten und Funktionen aussperren, sondern vielmehr als Weichen, die sie in passende, altersadäquate und sichere Umgebungen führen (Meineck 2025) – und auf diese Weise auch Befähigung und Teilhabe ermöglichen.

Zweitens müssen Verifikationssysteme als solche mit *informationeller Selbstbestimmung, digitaler Souveränität und Privatheit* vereinbar sein. Die Befragten heben in diesem Zusammenhang den Wert und die Bedeutung von Anonymität und Pseudonymität hervor. Ein Vertreter eines deutschen

Digitalverbands betont die „Chance [...], sich einfach als der auszugeben, der man gerne sein möchte“ (I16). Ähnlich äußern sich auch ein Digitalpolitiker (I17) sowie ein Experte für Kinder- und Jugendmedienschutz (I13). Zentral ist diese Möglichkeit sowohl für die persönliche Entfaltung in der öffentlichen Kommunikation als auch für die freie Meinungsäußerung; andernfalls könnte das Gefühl des Überwachtwerdens zur *Selbstzensur* der Nutzenden führen (I19).

Verifikationssysteme sollten darum nur solche personenbezogenen Daten erheben, die für die gegenständliche Verifikation – hier die Überprüfung des Alters – zwingend notwendig sind. Wenn Plattform-Betreiber lediglich die Altersgruppe der Nutzenden feststellen können und sonst keine weiteren Informationen erhalten, dann sei dies im Hinblick auf Privatheit und informationelle Selbstbestimmung „sogar besser [...] als] im Laden, wo ich meinen Ausweis der Verkäuferin oder dem Verkäufer unter die Nase halte und die im Prinzip alle meine Daten auf dem Ausweis lesen können“ (I18). Neben Art und Umfang der erhobenen Daten ist auch relevant, wo und wie sie verarbeitet werden und wer darauf Zugriff hat. Maßgeblich ist hierbei die Gewährleistung einer sicheren, verschlüsselten Datenübermittlung (vgl. auch Europäische Kommission 2025, S. 15; zu den Sicherheitsrisiken durch Altersverifikationsfirmen s. Booth 2025). Ferner sollte die Datenerhebung zur Altersverifikation nicht mit anderen Datenerhebungen auf Online-Plattformen verknüpft und darum auch nicht durch den Betreiber des jeweiligen Metaversums selbst durchgeführt werden, sondern durch einen unabhängigen Dritten (vgl. Europäische Kommission 2025, S. 11). Erforderlich ist zudem eine wirksame Kontrolle der datenerhebenden Institutionen, um möglichem Missbrauch entgegenzuwirken.

Drittens sind aus ethischer Perspektive auch *Nicht-Diskriminierung* und *Zugangsgerechtigkeit* relevant, wozu besonders eine *Nutzungsfreundlichkeit* von Altersverifikationssystemen beitragen kann, die auch für deren Akzeptanz und praktische Umsetzung von Bedeutung ist (Schellenberg/Mikát 2022). Entsprechende Systeme sollten unabhängig von „Behinderung, Sprache, ethnischer Zugehörigkeit, Geschlecht, Religion und Zugehörigkeit zu einer Minderheit“ nutzbar sein (Europäische Kommission 2025, S. 15), um eine gleichberechtigte Partizipation an digitalen Anwendungen wie Metaversen zu ermöglichen. Faktische Ausschlüsse – etwa durch die Beschränkung auf bestimmte nationale Identitätsnachweise, die vorausgesetzte Mitwirkung Erwachsener oder komplexe technische bzw. sprachliche Anforderungen – sind daher zu vermeiden (Meineck 2025). Der Nicht-Diskriminierung dienen niedrigschwellige, altersgerechte und transparente *Korrektur- und Unterstüt-*

zungsoptionen bei technischen Fehlern oder Bedienungsproblemen (I18). Um unberechtigten Ausschlüssen entgegenzuwirken, müsste es Nutzenden demnach möglich sein, die Ergebnisse einer Verifikation überprüfen und revidieren zu lassen oder eine alternative Form der Verifikation zu beantragen, wenn das Verifikationssystem den Zugang zu einem bestimmten Dienstangebot etwa irrtümlich verweigern sollte. In diesem Zusammenhang sind kindgerechte Hilfsfunktionen (z. B. durch Erklärungen in einfacher Sprache), technische Assistenz (z. B. via Chat) sowie eine klar verständliche Offenlegung der Gründe für etwaige Zutrittsverweigerungen integrale Bestandteile einer verantwortungsvollen Systemgestaltung. Dies würde nicht nur die Selbstwirksamkeit junger Nutzenden fördern, sondern zugleich ihr Vertrauen in digitale Schutzmechanismen stärken.

5. Rechtliche Verpflichtungen und Anforderungen an Altersverifikationen

Ob geltendes Recht den Einsatz von Verifikationsmaßnahmen in Metaver- sen bereits reguliert und dabei auch die drei zuvor aufgeworfenen, zen- tralen Anforderungen berücksichtigt, ergibt sich unionsweit primär aus dem Gesetz über digitale Dienste (Digital Services Act, DSA) als Teil des Plattformrechts sowie der Datenschutz-Grundverordnung (DSGVO) als Fundament des Datenschutzrechts.

5.1 Risikobasierte Jugendschutzmaßnahmen nach dem DSA

Der DSA soll als unmittelbar geltende Verordnung EU-weit einheitlich die allgemeine Sicherheit digitaler Dienste bzw. Dienste der Informations- gesellschaft² erhöhen, indem die Anbieter digitaler Dienste zu verschiede- nen Sicherheitsmaßnahmen verpflichtet werden, darunter auch zu Jugend- schutz-Maßnahmen. So haben „Anbieter sehr großer Online-Plattformen“ mit bestimmten jährlichen Nutzendenzahlen und Umsätzen nach Art. 35 DSA Risikominderungsmaßnahmen zu ergreifen, z. B. „Werkzeuge zur Al- tersüberprüfung“ (Art. 35 Abs. 1 S. 2 lit. j DSA). Nach dem Wortlaut der Norm gehören „Werkzeuge zur Altersüberprüfung“ aber nur „unter Um-

2 Art. 3 lit. a DSA verweist als Definition für digitale Dienste auf Art. 1 Abs. 1 lit. b Richtlinie (EU) 2015/153, wonach es sich um „Dienstleistung(en) der Informationsge- sellschaft, d. h. jede in der Regel gegen Entgelt elektronisch im Fernabsatz und auf individuellen Abruf eines Empfängers erbrachte Dienstleistung“ handelt.

ständen“ zu solchen Risikominderungsmaßnahmen dazu, sodass es sich bei Art. 35 Abs. 1 S. 2 DSA um eine nicht-abschließende Liste an Maßnahmen handelt, deren konkrete Auswahl sich an den spezifischen Risiken eines digitalen Dienstes orientiert. Diese Risiken sind wiederum im Rahmen einer von Art. 34 DSA vorgesehenen Risikobewertung festzustellen.

Demgegenüber fordert Art. 28 DSA sämtliche (und nicht nur sehr große) Anbieter von Online-Plattformen auf, „geeignete und verhältnismäßige Maßnahmen“ zu ergreifen, um den Zugriff auf entwicklungsschädigende Inhalte durch Minderjährige einzuschränken und „für ein hohes Maß an Privatsphäre, Sicherheit und Schutz von Minderjährigen“ zu sorgen. Nach ErwG 71 DSA trifft die Pflicht alle Online-Plattformen, die Risiken für Minderjährige bergen und für solche zugänglich sind, etwa wenn die Allgemeinen Geschäftsbedingungen (AGB) die Nutzung erlauben oder Daten der Nutzenden verarbeitet werden, aus denen ihr Alter hervorgeht.

Zusätzlich geltende Leitlinien zu Art. 28, die als Bewertungsrahmen für die Angemessenheit der Maßnahmen nach Art. 28 DSA dienen (Europäische Kommission 2025, Rn. 3), und besonders die Ausgestaltung von Altersverifikationen (nach den Leitlinien sog. „Altersfeststellungen“) in den Fokus nehmen, beziehen sich dabei auf die Risiken der 5C-Typologie³ (Europäische Kommission 2025, Anhang) nach der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD 2021). Darunter fallen u. a. Inhalts- und Kontaktrisiken, die – wie eingangs beleuchtet – bei vielen Metaversen in hohem Maße vorhanden sein können (Steenbergen 2026).

Der ersten zentralen Anforderung, Altersverifikationssysteme nach einem *risikobasierten Ansatz* einzurichten, um u. a. die *Teilhabe* von Kindern und Jugendlichen nicht einzuschränken, wird durch den DSA somit bereits Rechnung getragen: Die Erforderlichkeit und die Ausgestaltung von Altersverifikationsmaßnahmen als mögliche Maßnahme nach Art. 28 Abs. 1 u. Art. 35 DSA ist davon abhängig, ob eine Online-Plattform wie ein Metaversum Risiken für Minderjährige birgt und wie hoch diese Risiken jeweils sind.

5.2 Wahrung von Nicht-Diskriminierung und Zugangsgerechtigkeit

Ferner greifen die Art. 28 DSA-Leitlinien auch die aufgeworfene Anforderung an *Nicht-Diskriminierung* und *Zugangsgerechtigkeit* auf, indem sie

3 Die 5C umfassen Inhaltsrisiken (content), Verhaltensrisiken (conduct), Kontaktrisiken (contact), Verbraucherrisiken (consumer) und Querschnittsrisiken (cross cutting risks).

fünf Merkmale vorschreiben, die Altersverifikationsmaßnahmen aufzuweisen haben: Genauigkeit, Zuverlässigkeit, Robustheit, Nicht-Intrusivität und die Nichtdiskriminierung selbst (Europäische Kommission 2025, Rn. 49). Dabei hat die Notwendigkeit der Genauigkeit zur Folge, dass bisher oftmals eingesetzte Eigenerklärungen wie die selbstständige Eingabe des Geburtsdatums fortan als ungeeignet gelten (Europäische Kommission 2025, Rn. 52, Steenbergen 2026, S. 154) – was im Hinblick auf die zuvor aufgeworfenen Risiken für Minderjährige zu befürworten ist. Stattdessen müssen Anbieter digitaler Dienste als präzisere Methode sog. „Altersüberprüfungen“ durchführen, wenn ihre Plattform hohe Risiken für Minderjährige aufweist. Eine solche Altersüberprüfung setzt eine verifizierte Identifizierung der Nutzenden voraus (Europäische Kommission 2025, Rn. 29).

Der *Nicht-Diskriminierung*, *Zugangsgerechtigkeit* und auch *Nutzungsfreundlichkeit* entgegenstehen könnte jedoch, dass nach den Leitlinien zumindest bei digitalen Diensten mit mittleren Risiken auch Altersschätzungen mittels Gesichts- oder Verhaltensprofil-Analysen ausreichend sind – dies, obwohl Altersschätzungen von der Europäischen Kommission selbst als weniger genaue Methode der Altersfeststellung eingestuft werden (Europäische Kommission 2025, Rn. 30 u. 33). Prinzipiell sollen Anbieter von Online-Plattformen durch die Zulässigkeit der Altersschätzung bei ihrer Pflicht unterstützt werden, mehr als eine Methode der Altersfeststellung anzubieten, damit Nutzende zwischen mehreren Methoden wählen können und nicht vereinzelt ausgeschlossen werden (Europäische Kommission 2025, Rn. 48). Vor dem Hintergrund, dass Altersschätzungen aber anfälliger für technische Fehler oder Bedienungsprobleme sein können (Jandt 2025, S. 72), ist fraglich, wie *nutzungsfreundlich* ihre alleinige Bereitstellung in der Praxis tatsächlich ist, und ob sie nicht auch negative Auswirkungen auf die *Teilhabe* von Kindern und Jugendlichen haben können, wenn fehlerhafte Altersschätzungen sie „[...] am Zugang zu Online-Plattformen hindern, auf die sie andernfalls zugreifen könnten“ (Europäische Kommission 2025, Rn. 33).

5.3 Datenschutzrechtliche Anforderungen zur Wahrung der Informationellen Selbstbestimmung und Privatheit

Altersschätzungen mittels Gesichts- oder Verhaltensprofil-Analysen werfen abschließend Fragen zur Wahrung der *informationellen Selbstbestimmung*, *digitalen Souveränität und Privatheit* als weitere Anforderung auf. Insbesondere stellt sich die Frage, ob geltendes Recht diese Anforderung bei

der Umsetzung von Altersfeststellungsmethoden bereits angemessen berücksichtigt.

Das Recht auf informationelle Selbstbestimmung als Ausprägung des Grundrechts auf Datenschutz (Art. 8 GRCh) wird maßgeblich durch die DSGVO geschützt, sodass entscheidend ist, ob die DSGVO beim Einsatz von Altersfeststellungen überhaupt Anwendung findet, und ob Anwender von Altersfeststellungen überhaupt entsprechende Maßnahmen nach der DSGVO umsetzen müssen. Immerhin wird dem Anschein nach nur eine abstrakte Information wie das Alter eines Nutzenden erfasst.

Als EU-weit unmittelbar anzuwendende Verordnung reguliert die DSGVO jede Verarbeitung personenbezogener Daten. Personenbezogene Daten sind nach Art. 4 Nr. 1 DSGVO alle Informationen, „die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen“, z. B. der Realname eines Nutzenden. Durch das sehr weite Begriffsverständnis fallen aber auch Informationen darunter, die erst gemeinsam mit anderen Informationen die Identität des Nutzenden preisgeben (z. B. Identifikationsnummern, IP-Adressen oder physiologische Merkmale wie das Alter). Nach aktueller Rechtsprechung und einem relativen Verständnis ist der Personenbezug jedoch dynamisch zu bestimmen; entscheidend ist, ob eine Stelle, der Informationen vorliegen, über (zusätzliche) Informationen verfügt, die eine Zuordnung zur betroffenen Person ermöglichen.⁴

Im ersten Schritt einer (Alters-)Verifikation muss das gegenständliche Datum (z. B. das Alter) einer Person (z. B. eines Metaverse-Nutzenden) von einer Stelle (z. B. einer Behörde) mittels einer vertrauenswürdigen Quelle (z. B. einem Ausweisdokument) festgestellt werden. Im zweiten Schritt muss das Ergebnis von der Stelle, die das Ergebnis benötigt (z. B. dem Metaverse-Betreiber), abgerufen oder eingesehen werden. Zudem ist eine überprüfbare Zuordenbarkeit zwischen dem Datum und dem den Nachweis präsentierenden Nutzenden erforderlich, damit festgestellt werden kann, ob das Ergebnis bzw. abgefragte Datum dem Nutzenden auch tatsächlich zugehörig ist, und nicht etwa ein Kind den Ausweis der Sorgeberechtigten nutzt. Während es für die Frage des Personenbezugs beim zweiten Schritt darauf ankommt, inwiefern die prüfende Stelle wie der Metaverse-Betreiber durch den Nachweis und die Zuordnung den Nutzen-

4 EuGH, Urteil vom 4.9.2025, Rechtssache C-413/23P, Europäischer Datenschutzbeauftragter (EDSB) ggf. einheitlicher Abwicklungsausschuss (SRB), NJW 2025, S. 3415–3423, Rn. 76 f.

den identifizieren kann, müssen Nutzende im ersten Schritt allenfalls ihre Identität preisgeben.

Aufgrund der weitreichenden Definition der Verarbeitung nach DSGVO genügt bereits eine Abfrage personenbezogener Daten, um eine Verarbeitung zu bejahen. Eine (dauerhafte) Speicherung des Ergebnisses ist keine Voraussetzung. Demnach werden Verifikationsmaßnahmen insgesamt ungeachtet ihrer Ausgestaltung als Altersprüfung, Altersschätzung oder zu sonstigen Verifikationszwecken der DSGVO unterliegen.

Da Betreiber eines Metaversums eine oder mehrere Methoden der Altersverifikation in den Nutzungsprozess ihrer Plattform einbinden müssen, entscheiden sie zugleich über die Zwecke und Mittel der mit der Verifikation verbundenen Verarbeitung und gelten damit als Verantwortliche nach Art. 4 Nr. 7 DSGVO (Geissler/Schlindwein 2025).

Damit haben sie wiederum die Datenschutz-Grundsätze nach Art. 5 DSGVO einzuhalten, wozu insbesondere der *Datenminimierungsgrundsatz* (Art. 5 Abs. 1 lit. c DSGVO) gehört, der voraussetzt, dass sämtliche an den Verifikationsmaßnahmen Beteiligte nur solche personenbezogenen Daten erheben und erhalten, die sie für den Zweck der Verifikation auch tatsächlich brauchen, ebenso wie den *Zweckbindungsgrundsatz* (Art. 5 Abs. 1 lit. b DSGVO), nach dem Verifikationsmaßnahmen u. a. – wie zuvor gefordert – so betrieben und implementiert werden müssen, dass keine Verknüpfung zwischen den zur Altersverifikation erhobenen Daten und den bei der Nutzung einer Plattform erhobenen Daten erfolgt.

Zur Einhaltung der Grundsätze haben Verantwortliche technische und organisatorische Maßnahmen nach Art. 25, 32 DSGVO zu ergreifen, die von der DSGVO – ähnlich wie Art. 28 und Art. 35 DSA – aufgrund des risikobasierten Ansatzes und der Technologieneutralität grundsätzlich offengehalten werden. Es besteht somit zwar keine zwingend einzuhaltende Liste mit Anforderungen, bestimmte Maßnahmen wie eine eingangs geforderte verschlüsselte Übermittlung der Daten werden in der DSGVO aber explizit als Maßnahme genannt und zählen mittlerweile oftmals zur gängigen Praxis. Zusätzlich ergeben sich erforderliche Maßnahmen teilweise bereits aus den Art. 28-DSA-Leitlinien: Dazu gehört z. B. die Einbindung eines Dritten⁵ zur Durchführung der Identitätsabfrage statt durch den Be-

5 Für solche Datenverarbeitungen sieht die DSGVO verschiedene Konstellationsmöglichkeiten vor, z. B. die Ausgestaltung als gemeinsame Verantwortlichkeit nach Art. 26 DSGVO oder als Auftragsverarbeitung nach Art. 28 DSGVO, wobei die Ausgestaltung vom konkreten Sachverhalt abhängt.

treiber eines Metaversums selbst (Europäische Kommission 2025, Rn. 41 u. 53). Indem Nutzende dadurch gegenüber dem Betreiber eines Metaverse weiterhin pseudonym auftreten können, trägt eine solche Einbindung und Trennung der Prozesse wiederum zur Datenminimierung bei. Als weitere Maßnahme könnten eingebundene Dritte das Ergebnis der Verifikation (z. B. die Zuordnung zu einer bestimmten Altersgruppe) dem Nutzenden oder dem Betreiber des Metaversums – wenn immer möglich – als einfache Ja- oder Nein-Antwort statt als genaues Geburtsdatum oder genaues Alter bereitstellen. Die Art. 28-DSA-Leitlinien empfehlen für diesen Prozess die Verwendung der geplanten europäischen Brieftasche für die Digitale Identität (European Digital Identity-Wallet, EUDI-Wallet)⁶ nach der eIDAS-2-Verordnung, die europäischen Bürgerinnen und Bürgern als europaweit einheitliche Lösung für verifizierbare Identitätsnachweise (Europäische Kommission 2025, Rn. 1 ff.) voraussichtlich ab 2027 zur freiwilligen Nutzung zur Verfügung stehen wird.

Allerdings ist zu berücksichtigen, dass technische und organisatorische Maßnahmen nicht erst beim Betrieb von Verifikationsmechanismen ansetzen, sondern bereits bei der Implementierung von datenverarbeitenden Technologien. Vor dem Hintergrund, dass von der Europäischen Kommission gestattete Altersschätzungen regelmäßig die Verarbeitung großer Mengen (verhaltensbezogener) personenbezogener Daten oder Daten besonderer Kategorien nach Art. 9 DSGVO wie biometrische Daten für Gesichtsanalysen erfordern (Kunz/Waldmann 2025b), ist daher fraglich, ob Altersschätzungen überhaupt die Grundsätze an Datensparsamkeit und -minimierung einhalten können.

Neben den Datenschutzgrundsätzen haben Betreiber von Metaversen als Verantwortliche zudem die Umsetzung der datenschutzrechtlichen Betroffenenrechte nach Art. 12 ff. DSGVO zu gewährleisten. Im Zuge dessen müssen Nutzende nicht nur über den Vorgang der Altersverifikation transparent informiert werden, im Rahmen des Rechts auf Löschung sowie Berichtigung muss es ihnen auch möglich sein, ihre personenbezogenen Daten – wozu auch der Altersnachweis gehört – beim Metaverse-Betreiber und eingebundenen Dritten löschen bzw. berichtigen zu lassen. Dies setzt insbesondere eingangs hervorgehobene *Korrektur- und Unterstützungsoptionen* während der Altersverifikation voraus, aber auch im Anschluss. Um die Richtigkeit der Daten im Sinne des Richtigkeitsgrundsatzes nach Art. 5

6 Siehe dazu auch eine entsprechende App zur Altersverifikation, das sogenannte Mini-Wallet, welches von der EU-Kommission im April 2026 vorgestellt wurde.

Abs.1 lit. d DSGVO von vornherein zu gewährleisten, sollten Nutzende rechtzeitig oder regelmäßig gebeten werden, den Nachweis zu erneuern, damit Betreiber von Metaversen die entsprechende Altersgruppe aktualisieren können, wenn ihnen das genaue Geburtsdatum nicht bekannt ist und eine selbstständige Aktualisierung somit ausscheidet.

Insgesamt besteht durch ein Zusammenspiel zwischen DSA, den zugehörigen Leitlinien sowie den Grundsätzen der DSGVO und dem dazugehörigen Aufsichtsregime durch Aufsichtsbehörden somit bereits ein rechtliches Fundament, um eine ethische Ausgestaltung von Altersverifikationssystemen anzustreben und sowohl die Privatheit als auch informationelle Selbstbestimmung als eine der drei wesentlichen Anforderungen wirksam durchzusetzen. Allein die von der Europäischen Kommission eingeräumte Möglichkeit zum Einsatz von Altersschätzungen bei Plattformen mit mittleren Risiken wahrt die Rechte der betroffenen Nutzenden nicht vollständig. Gerade aus Sicht der Datensparsamkeit bleiben Altersüberprüfungen die bevorzugte und rechtssichere Methode, wenn je nach Anlass nur die Altersgruppe der Nutzenden mitgeteilt wird.

6. Technische Umsetzung einer datenschutzfreundlichen Altersverifikation

Aktuelle Praktiken der Altersverifikation nutzen unzuverlässige und für Metaversen ungeeignete Verfahren. Wirksame und datenschutzfreundliche Verfahren sind dagegen auf Basis von Self-Sovereign Identity (SSI) möglich und auch für die Nutzung in Metaversen geeignet, da sie die flexible, kryptografisch gesicherte Ausstellung von Identitätsnachweisen ermöglichen und deren Nutzung unter der Kontrolle der betroffenen Nutzenden stellen.

6.1 Aktuelle Praxis der Altersverifikation

Altersverifikationsverfahren werden in der Praxis derzeit meist nur unzureichend umgesetzt. Entweder findet keine zuverlässige Altersverifikation, sondern lediglich eine Eigenerklärung statt, oder die Verfahren sind nicht datenschutzfreundlich, da mehr personenbezogene Daten als notwendig preisgegeben werden müssen (Kunz/Waldmann 2025b). Zu den Verfahren, die beispielsweise bei Social-Media-Diensten zum Einsatz kommen, zählen neben der Eingabe des Alters das Hochladen von Ausweiskopien (teilweise in Kombination mit Video-Selfies zum Abgleich), Video-Selfies mit KI-ge-

stützter Altersschätzung, Altersschätzung anhand der Online-Aktivitäten der Nutzenden sowie Social-Vouching, bei dem den Nutzenden bekannte Personen aufgefordert werden, das angegebene Alter zu bestätigen.

Unter aktuellen Metaversen nutzt *VRChat* einen Drittanbieter für die Altersverifikation, der Nutzende auffordert, über ihre Kamera ein offizielles Ausweisdokument zu präsentieren, anhand dessen das Alter festgestellt wird. Darüber hinaus haben Nutzende über 18 Jahre bei *VRChat* die Möglichkeit, nach erfolgreicher Altersverifikation ihrem Profil ein „18+“-Badge hinzuzufügen, das sie für andere Nutzende sichtbar machen können (VRChat 2025; s. Abschn. 4.1). Auch hier müssen Nutzende folglich mehr personenbezogene Daten offenlegen als zwingend erforderlich, wenn auch nicht gegenüber dem Betreiber von *VRChat* selbst.

Eine datenschutzfreundlichere Möglichkeit der Altersverifikation bietet die Online-AusweisFunction (eID) des Personalausweises und des elektronischen Aufenthaltstitels. Diese stehen jedoch nur Personen zur Verfügung, die 16 Jahre und älter sind.

Aus unserer Sicht stellt eine Altersverifikation auf Basis von Self-Sovereign Identity (SSI) eine zuverlässige, flexible, datenschutzfreundliche und diskriminierungsarme Alternative zu den genannten Verfahren dar, indem sie datenminimierte Identitätsnachweise beliebiger (auch jüngerer) Nutzer ermöglicht. SSI wird bisher für die Altersverifikation noch nicht eingesetzt, bietet aber entscheidende Vorteile. Denn die Ausstellung und Überprüfung SSI-basierter Nachweise kann auf verschiedene Beteiligte verteilt werden, sodass der Bedarf an zentralen Behörden abnimmt. Zugleich werden proprietäre Unternehmenslösungen und Monopolstellungen im Identitätsmanagement vermieden.

6.2 Lösungsentwurf SSI-basierter Identitätsnachweise für Metaversen

Die mit SSI-basierten Identitätsnachweisen verbundenen Prozesse lassen sich nutzerzentriert und anwendungsnah gestalten. Im Zuge der *Teilhabe* und *informationellen Selbstbestimmung* können Nutzende und damit auch Kinder und Jugendliche durch SSI ihre digitalen Identitäten selbst verwalten und über die Weitergabe von Informationen selbst entscheiden. In der Regel wird eine öffentliche, gegenüber den verschiedenen Dienstleistern weitgehend neutrale Blockchain als zentrale Registrierungskomponente eingesetzt. Dadurch wird sichergestellt, dass die betroffenen Nutzenden zur Kontrolle ihrer Informationen nicht auf private Architekturen der datenverarbeitenden

Dienstleister angewiesen sind. Zugleich kann auch das Tracking von Nutzen- den und die Verkettung von Aktivitäten im Sinne der datenschutzrechtlichen *Zweckbindung* erschwert werden (Kunz/Waldmann 2025a; Schumm u. a. 2025). Die Identitätsdaten befinden sich im Besitz des betroffenen Nutzen- den, um nicht nur die einzelnen Identitätsattribute, sondern auch die damit verbundenen Aktionen zu kontrollieren. Damit entzieht sich SSI weitgehend der Kontrolle einer einzelnen Organisation, eines einzelnen Softwareanbie- ters oder einzelnen Dienstes (Schumm u. a. 2025).

Um unseren SSI-basierten Lösungsansatz verständlich zu machen, wer- den zunächst die Grundlagen von SSI kurz erläutert. SSI basiert auf zwei Grundbausteinen, den sogenannten dezentralen IDs (*Decentralized Identi- fier*, DID) und den verifizierbaren Identitätsdaten (*Verifiable Credentials*, VC). Beide werden durch das World Wide Web Consortium (W3C) spezi- fiziert (Sporny u. a. 2022; 2025). Ein Aussteller (*Issuer*) stellt einer Person (*Holder*) digitale Identitätsnachweise (VC) aus, die eine Behauptung über eine Person (*Subject*) enthalten und vom *Holder* in ihrer digitalen Briefta- sche (Wallet) gespeichert werden. Der Aussteller hat die Identitätsdaten zu- vor geprüft, etwa durch Abgleich mit dem Personalausweis der betroffenen Person (*Subject*), und bestätigt durch eine digitale Signatur die Korrektheit der Daten. In der Regel werden beide Rollen, *Holder* und *Subject*, von der- selben Person eingenommen. Abb. 1 zeigt die grundlegenden Beziehungen der SSI-Rollen am Beispiel der Ausstellung eines Altersnachweises durch eine Meldebehörde an Nutzende als *Subject* und *Holder*.

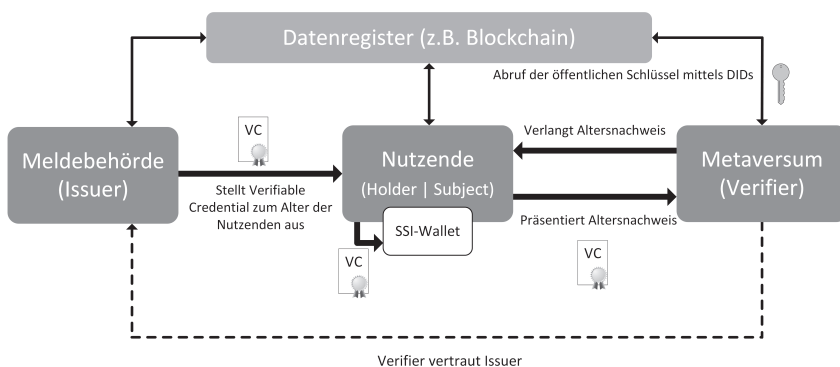


Abbildung 1: Die Beziehungen zwischen Issuer, Holder, Subject und Verifier am Beispiel einer Altersverifikation

Der Betreiber eines Metaversums kann nun als Dienstleister in der Rolle des *Verifier* von der nutzenden Person (*Holder* und *Subject* des VC) den Altersnachweis verlangen. Die nutzende Person als *Holder* des VCs kann entscheiden, ob sie der Anfrage nachkommen und das entsprechende VC präsentieren möchte. In diesem Fall signiert sie das ausgewählte VC und sendet es an den *Verifier*. Dieser kann nun die Gültigkeit und Korrektheit des Nachweises verifizieren.

Dazu sind die an einem VC beteiligten Parteien über ihre Kennungen (DIDs) eindeutig referenziert. Diese DIDs sind in öffentlich-zugänglichen, meist blockchainbasierten Datenregistern hinterlegt und dort mit weitergehenden Informationen (z. B. Links auf Einträge in der europäischen „Trusted List“ mit den zugelassenen VC-Ausstellern) verknüpft. Jede DID ist mit einem öffentlichen Schlüssel verbunden, wodurch die Identifizierung eines *Subjects* mit der Authentifizierung eines *Holder*s auch kryptografisch gesichert ist. Ein *Verifier* kann die öffentlichen Schlüssel über das Datenregister abrufen und damit das vom *Holder* präsentierte VC verifizieren.

Auf Basis von SSI können beliebige Beziehungen zwischen beliebigen *Holdern*, *Subjects*, *Issuern* und *Verifiern* abgebildet werden. So können sich die vom *Holder* in seinem Wallet verwalteten Identitätsnachweise auch auf ein anderes *Subject* beziehen, beispielsweise eine andere Person, eine Organisation, ein physisches oder digitales Objekt (z. B. einen Avatar). Ein *Holder* kann beliebig viele DIDs besitzen. Im Gegensatz zu herkömmlichen Registrierungsstellen werden die eigentlichen Identitätsdaten aber nicht in einem Datenregister verwaltet, sondern unterliegen der Kontrolle des jeweiligen *Holder*s in dessen Wallet. Voraussetzung ist, dass der *Verifier* dem betreffenden *Issuer* vertraut, die Identitätsdaten des *Subjects* ausreichend geprüft zu haben. Umgekehrt müssen die *Issuer* die *Verifier* gar nicht kennen und nicht mit ihnen kommunizieren. Die *Issuer* erfahren somit nicht, wann und für welche Dienste die VCs zum Einsatz kommen, sodass die DIDs dem Schutz der Kommunikation und der Privatheit dienen.

6.3 Nutzung SSI-basierter Identitätsnachweise in Metaversen

In Metaversen lässt sich SSI auf vielfältige Weise nutzen. Ein Anwendungsfall betrifft das erstmalige Betreten eines Metaversums und die in den meisten Fällen damit verbundene Registrierung durch dessen Betreiber (Kunz/Waldmann 2025a). Der Betreiber kann einen Altersnachweis auf Basis eines VC von den Nutzenden verlangen, um so etwa rechtlichen

Verpflichtungen des Jugendschutz- bzw. Plattformrechts nach Art. 28 DSA nachzukommen. Gleichzeitig können damit aus ethischer Sicht altersangemessene Nutzungserfahrungen (s. Abschn. 4.1) ermöglicht werden. Die Altersnachweise sollten von einer offiziellen Meldebehörde (*Issuer*) nach einer Identifizierung (z. B. Vorlage einer Geburtsurkunde) ausgestellt werden. So wird eine vom DSA vorgesehene und zur Wahrung der Privatheit erforderliche *Verifikation durch unabhängige Dritte* gewährleistet; zugleich kann sich der Betreiber (*Verifier*) auf die Korrektheit der enthaltenen Daten verlassen. Betreiber von Metaversen übernehmen hingegen ausschließlich die Altersangabe aus dem Zertifikat und verknüpfen sie mit dem Konto des Nutzenden, ohne das Alter erneut zu bestimmen. Eine *Datenverknüpfung* von Aktivitäten der Nutzenden in der realen Welt mit Aktivitäten im Metaversum wird durch die getrennte Nutzung von Zertifikaten in der realen Welt und Zertifikaten der virtuellen Identitäten erschwert. Durch ihre insgesamt *datensparsame* Ausgestaltung dient der Lösungsansatz somit dem Schutz der Kommunikation und der Privatheit und beugt einer *Überwachungsinfrastruktur* vor.

Ein zweiter Anwendungsfall betrifft die Kommunikation zwischen zwei beliebigen Avataren in öffentlichen Räumen eines Metaversums. Da Avatare frei gestaltet werden können, ist es kaum möglich, vom Aussehen eines Avatars auf das Alter des dazugehörenden Nutzenden zu schließen. Auch hier sollen VCs als Identitätsnachweise eingesetzt werden, sodass Nutzende untereinander ihr Alter nachweisen können. Dies dient insbesondere dem Schutz von Kindern und Jugendlichen. Dabei ist nicht erforderlich, dass die Altersnachweise das exakte Lebensalter enthalten. Stattdessen ist es ausreichend, wenn sich der Nachweis lediglich auf eine Altersgruppe bezieht (z. B. „zwischen 13 und 16 Jahren“ oder „über 18 Jahre“) und diese nur dann für andere Nutzende sichtbar wird, wenn beide Seiten dem zugestimmt haben (s. Abb. 2). Die Nachweise sollen vom Betreiber des Metaversums ausgestellt werden, nachdem dieser im Rahmen der Registrierung das Alter der Nutzenden geprüft hat (vgl. erster Anwendungsfall in Abb. 1). Dies hat u. a. den Vorteil, dass keine Angaben über eine Meldebehörde in den Nachweisen enthalten sind und Nutzende beim Prüfen somit keine weiteren Rückschlüsse (z. B. auf den Wohnort ihres Gegenübers) ziehen können.

Abb. 2 zeigt schematisch den Ablauf des Austauschs von Identitätsnachweisen zwischen zwei Nutzenden im Metaversum. Dies wird durch die explizite Anfrage eines der beiden Nutzenden initiiert. Nutzer N1 möchte in Gestalt des Avatars A1 mit Nutzer N2 kommunizieren. Bevor N2 dem zustimmt, möchte er zunächst das Alter seines Gegenübers wissen (Schritt

1). N1 hat den entsprechenden Nachweis in seinem Wallet gespeichert und kann ihn N2 präsentieren (Schritt 2). Nachdem N2 die Nachweise gesehen und erfolgreich verifiziert hat (Schritt 3), kann N2 informiert entscheiden, ob er mit N1 in Kontakt treten möchte. Auf die gleiche Weise kann N1 auch entsprechende Nachweise von N2 verlangen.

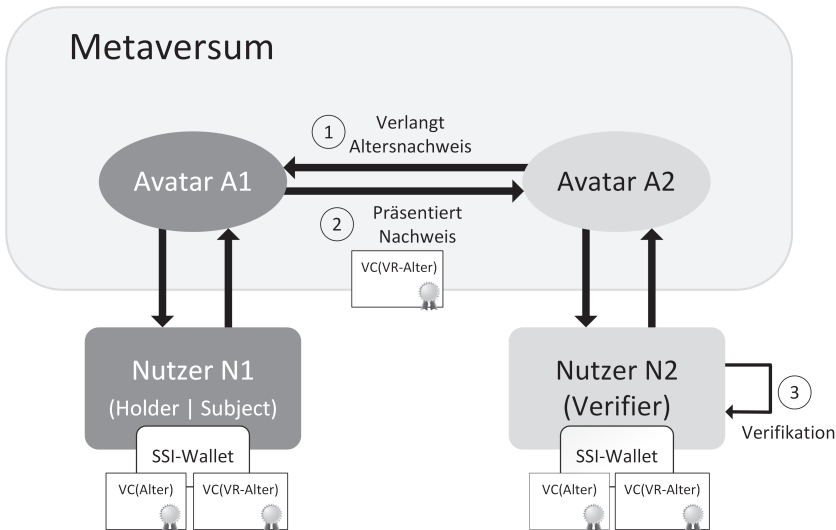


Abbildung 2: Nutzung von SSI in Metaversen

Aufgrund der Flexibilität von SSI sind weitere Nutzungsmöglichkeiten in Metaversen denkbar. So können beispielsweise VCs verwendet werden, um das Besitzverhältnis zwischen Nutzenden und ihren Avataren digital abzubilden. Solche VCs können Nutzenden im Metaversum helfen, zu erkennen, ob sie mit einem Avatar kommunizieren, der von einem anderen Nutzenden gesteuert wird, oder mit einem autonom agierenden, KI-gesteuerten Avatar (Kunz/Waldmann 2025a). Dies wirkt u. a. dem Risiko entgegen, dass insbesondere Kinder und Jugendliche intensive parasoziale Beziehungen zu KI-Agenten aufbauen und dann von diesen etwa in Bezug auf Konsum oder Weltanschauungen manipuliert werden könnten (s. Abschn. 3).

SSI-basierte Nachweise der Altersgruppe lassen sich zudem mit der Anwendung von Abstandsregelungen in Metaversen kombinieren, um Nutzende vor *virtuellen Belästigungen* durch andere Nutzende zu schützen.

Beispielsweise werden in *Horizon Worlds* und *VRChat* bereits „Safe Zones“ oder „Personal Space Options“ eingesetzt, die einen unsichtbaren Schutzkreis um den Avatar des Nutzenden herum bilden und dafür sorgen, dass sich andere Avatare nicht ohne gegenseitige Zustimmung nähern können (vgl. Krell/Wettmann 2023, S. 41). Denkbar wäre, dass solche Abstandsregelungen bzw. „Safe Zones“ für die Kommunikation und evtl. weitere Interaktionsmöglichkeiten für junge Nutzende erst geöffnet werden, nachdem sie gegenseitig einen Kontaktwunsch signalisiert und freiwillig ihre Altersgruppe ausgetauscht haben. Jeder Nutzende kann jederzeit die Kommunikation beenden und seine eigene „Safe Zone“ wieder gegenüber anderen verschließen. Ein solcher Ansatz ist besonders geeignet, um Kindern und Jugendlichen zu ermöglichen, mit Erwachsenen in Metaversen zu interagieren, sie aber auch vor entsprechenden unerwünschten Interaktionen zu schützen (Befähigung und Schutz; s. Abschn. 4.1). Wir schlagen folgende drei Aktionen vor, die zusammen ausgeführt werden müssen, um anderen Nutzenden einen Kontaktwunsch mitzuteilen:

- Unterschreitung eines bestimmten Abstands zum anderen Avatar für wenige Sekunden;
- Blickkontakt zum anderen Avatar;
- Ausführen einer bestimmten Geste (z. B. Winken) in Richtung des anderen Avatars.

Erwidert der andere Nutzer mit seinem Avatar den Blickkontakt und die Geste, wird im Hintergrund eine Bitte um Mitteilung der Altersgruppe (und ggf. weiterer freiwillig gegebener Informationen) an ihn gesendet. Ergänzend empfiehlt sich die Kennzeichnung von Avataren, die einem Nutzenden bereits aus einer früheren Begegnung bekannt sind. Sinnvoll ist die Angabe des Avatar-Namens (Benutzername), der Altersgruppe und der letzten Begegnung (ob diese zustande kam oder abgelehnt wurde).

Um zu vermeiden, dass Nutzende über verschiedene Angebote hinweg wiedererkannt, ihre dortigen Aktivitäten miteinander verknüpft und somit in ihrem Recht auf *Privatheit* und *informationelle Selbstbestimmung* eingeschränkt werden, ist aber unabdingbar, dass Nutzende verschiedene, voneinander getrennte Identitäten schaffen und nutzen können. Dafür ist nicht nur erforderlich, dass Nutzende verschiedene Nachweise (mit unterschiedlichen Schlüsseln und von unterschiedlichen Ausstellern) in staatlichen als auch privatwirtschaftlichen Wallets verwalten können, die SSI-Infrastruktur muss auch offen gestaltet werden und für viele unterschiedliche Aussteller von Nachweisen und Diensteanbietern nutzbar sein. Damit kann eine die

Privatheit gefährdende Verknüpfung und Profilbildung personenbezogener Daten zumindest erschwert werden.

Dennoch werden Dienstanbieter in der Rolle der *Verifier* den betroffenen Nutzenden zunächst überlegen bleiben, solange Dienstanbieter bestimmen können, welche persönlichen Attribute die Nutzenden präsentieren müssen. Denn Dienstanbieter könnten mehr Daten anfordern, als sie für den angegebenen Verarbeitungszweck benötigen, und Nutzende würden vom Dienstangebot ausgeschlossen werden, wenn sie der Präsentation aller angeforderten Daten nicht nachkämen (Schumm u. a. 2025). Um ausufernde Identifizierungspflichten zu verhindern, müssen daher SSI-Anwendungsprofile definiert werden, die den Betreibern von Metaversen vorschreiben, welche Daten sie von Nutzenden für welche Zwecke erheben dürfen. Das ursprüngliche SSI-Paradigma, welches ausgestellte Zertifikate ausschließlich in den Wallets des betroffenen Nutzenden speichert und die darin vom Nutzenden auch einfach gelöscht werden können, sollte dabei nicht aufgegeben werden.

7. Grenzen der Altersverifikation

Eine flächendeckende Einführung von Altersverifikationsmaßnahmen wird grundsätzlich kontrovers diskutiert. Insbesondere wird sie abgelehnt, wenn sie durch proprietäre, intransparente Verfahren großer Internetkonzerne realisiert wird (Open Letter 2026). Eine pseudonyme Altersverifikation, wie sie in diesem Beitrag beschrieben wurde, kann dagegen als transparent und datenschutzfreundlich gelten, bei einem Einsatz auf Basis des EUDI-Wallets interoperabel eingesetzt werden, und damit einen wichtigen Beitrag zum Schutz von Kindern und Jugendlichen in digitalen Räumen wie Metaversen leisten.

Dennoch kann eine SSI-basierte Altersverifikation nicht alle Probleme lösen – die spezifischen Herausforderungen von Metaversen sowie allgemeine Einschränkungen von (Alters-)Verifikationssystemen müssen auch für die vorgeschlagenen SSI-Prozesse Beachtung finden:

Erstens verbleibt trotz der im Vergleich zu Eigenerklärungen geringen Umgehungsmöglichkeiten ein Restrisiko, dass Kinder und Jugendliche dennoch Wege finden, Zugangssperren erfolgreich zu umgehen (Meineck 2025). Unverantwortliches Handeln kann damit nicht unterbunden werden. So müssten SSI-Wallets und die Präsentation der Zertifikate allein unter der *Kontrolle der Nutzenden* stehen, wozu eine herstellerunabhängige

Unterstützung von Wallets sowie eine Auswahl geeigneter, barrierefreier Authentifizierungsverfahren notwendig ist. Trotz Erfüllung dieser Voraussetzungen bleibt dennoch die Möglichkeit bestehen, dass eine volljährige Person ihr Smartphone mit Wallet und Altersnachweis mitsamt der zugehörigen PIN einer jüngeren Person überlässt oder nach erfolgreich erbrachtem Altersnachweis das Headset weiterreicht. Umgekehrt können auch Erwachsene die Verifikation von Minderjährigen nutzen, um sich als Kind auszugeben und sich auf diese Weise Kindern in missbräuchlicher Weise zu nähern (*Cybergrooming*). Kontinuierliche biometrische Authentifizierungsverfahren sind in dieser Hinsicht schwieriger zu umgehen, da sie einer Person sicherer zugeordnet werden können als eine einmalige PIN-Eingabe. Gerade im Hinblick auf Diskussionen und Kritik an Verifikationsmaßnahmen (Open Letter 2026) sowie auf die Sorge vor einer Überwachungsinfrastruktur bleibt die geringere Datenschutzfreundlichkeit solcher Verfahren aber hervorzuheben.

Zweitens gilt bezüglich der *Nutzungsfreundlichkeit und Unterstützung der digitalen Souveränität* zu beachten, dass in der Regel ein VR-Headset getragen wird, um ein Metaversum für immersive Erlebnisse zu nutzen, obwohl jede Form der Verifikation regelmäßig ein weiteres Medium erfordern wird – nach dem SSI-basierten Vorschlag ein entsprechendes Wallet, das meistens auf den Smartphones der Nutzenden gespeichert wird. Bei gleichzeitigem Tragen eines VR-Headsets wird die Nutzung des Smartphones eventuell eingeschränkt sein und die immersiven Erfahrungen gestört werden. Zudem müssen sich Nutzende (*Holder*) für gewöhnlich gegenüber ihrem Wallet authentifizieren (z. B. durch die Eingabe einer PIN), um die gespeicherten VCs anwenden und ihre privaten kryptografischen Schlüssel nutzen zu können. Aus diesem Grund erscheint eine Kopplung von Smartphone und VR-Headset etwa über Bluetooth sinnvoll, sodass Nutzende über das Sichtfeld ihres VR-Headsets mit dem Smartphone und dem darauf installierten Wallet interagieren können. Alternativ ist auch ein im VR-Headset installiertes Wallet oder die Nutzung eines cloudbasierten Wallets denkbar. Letzteres kann nutzungsfreundlich sein, stellt aber die alleinige Kontrolle des Wallets durch den Nutzenden infrage. Damit Kinder und Jugendliche ihre Altersnachweise zudem effektiv kontrollieren können, sind hier kindgerechte *Unterstützungs- und Hilfefunktionen* wie Erklärungen in einfacher Sprache oder direkte technische Assistenz entscheidend.

Drittens lässt sich selbst bei wirksamer Altersverifikation das Risiko schädlicher Online-Erfahrungen nicht vollständig eliminieren. Inhalts- und Interaktionsrisiken wie Mobbing, Stalking, sexuelle Belästigung, Desinfor-

mation, Hassrede oder die negative Beeinflussung von Selbstbildern können hinter der Altersschranke ungehindert stattfinden (Dreyer 2025).

8. Geteilte Verantwortung für sichere Metaversen

Altersverifikationen können den Zugang zu digitalen Räumen demnach zwar steuern, der Schutz von Kindern und Jugendlichen im digitalen Raum bleibt aber eine geteilte Verantwortung.

Kinder und Jugendliche benötigen zunächst die entsprechenden Kompetenzen, um Online-Risiken zu erkennen und verantwortungsvoll mit digitalen Angeboten umzugehen. Nur dann können sie sich selbstbestimmt und gefahrenbewusst in digitalen Räumen bewegen. Von *Bildungseinrichtungen und Sorgeberechtigten* braucht es daher Sensibilisierung, medienpädagogische Konzepte, kontinuierliche Begleitung und einen dialogischen Austausch mit Kindern und Jugendlichen.

Aufseiten der *Betreiber von Metaversen* muss die gesetzlich festgelegte und von der Europäischen Kommission vorgesehene Einbindung von Altersverifikationssystemen umgesetzt und das Ergebnis der Altersverifikation durchgesetzt werden. Dies ist Teil eines Bündels von Anforderungen an die Betreiber für eine altersgerechte, inklusive und sichere Ausgestaltung von Metaversen, zu der neben der Umsetzung der (datenschutz-)rechtlichen Anforderungen auch die Gewährleistung einer effektiven Inhaltsmoderation gehört.

Somit stehen *Politik und Gesetzgeber* nicht nur in der Verantwortung, Medienkompetenzerziehung finanziell zu fördern, ausreichende Weiterbildungsangebote für Erwachsene zu schaffen bzw. zu unterstützen und Beratungsangebote für Kinder und Jugendliche bei negativen Nutzungserfahrungen bereitzustellen, sondern vor allem bereits bestehende regulatorische Maßnahmen gegenüber den Betreibern von Metaversen wirksam durchzusetzen – dies über die EU hinweg und grenzüberschreitend. Hierzu sollten Betreiber von Metaversen möglichst verpflichtet werden, verbindliche, unabhängige und wahrnehmbare Klassifizierungen ihrer Angebote vorzunehmen, damit sich solche ihrer Verpflichtung zum Jugendschutz nicht schlichtweg entziehen, indem die Angebote einfach mit „Nur für Erwachsene“ gekennzeichnet werden.

Hinzu kommt, dass die Betreiber von Metaversen selbst zwar die Altersverifikation einbinden und Nutzungsschranken durchsetzen müssen, allerdings nicht maßgeblich an der Entwicklung von Altersverifikationsmecha-

nismen beteiligt sind, insbesondere wenn unabhängige Dritte eingebunden werden müssen. Es wird somit standardisierte Lösungen brauchen, deren Entwicklung durch *Politik und Gesetzgeber* vorangetrieben werden müssen. Dadurch treffen die datenschutzrechtlichen Anforderungen aber nicht ausschließlich die Betreiber von Metaversen, sondern besonders die *an der Entwicklung* von (SSI-basierten) Altersverifikationen *Beteiligten*. Insofern könnte die allgemeine Einführung von Altersverifikationssystemen auch ein Impuls für aktuelle Diskussionen im Rahmen der DSGVO-Reform sein, Anbieter bzw. Hersteller von Hard- und Software stärker „in das etablierte System datenschutzrechtlicher Pflichten“ einzubinden (DSK 2026, S. 4 f.).

9. Fazit

Kinder und Jugendliche sind in Metaversen besonderen Risiken ausgesetzt. Aus ethischer Perspektive können Altersverifikationsmaßnahmen dazu beitragen, sie vor diesen Risiken zu schützen, gleichzeitig aber auch ihre Teilhabe und Befähigung zum kritischen Umgang mit den entsprechenden Anwendungen fördern. Neben „Walled Garden“-Formaten, in denen Kinder und Jugendliche unter sich sind und sich im geschützten Raum bewegen und austauschen können, erlauben Altersverifikationssysteme, Zugänge zu Metaversen altersdifferenziert und teils auch eigenverantwortlich zu öffnen.

Dazu müssen entsprechende Altersverifikationssysteme aus ethischer Sicht u. a. Kriterien wie die Vermeidung von Diskriminierung sowie Datensparsamkeit erfüllen. Letzteres fordert der aktuelle Rechtsrahmen bereits anhand des weiten Anwendungsbereichs der DSGVO, der sich auch auf Verifikationsmaßnahmen erstreckt, sowie anhand des DSA, der die Anbieter von Online-Plattformen zum Einsatz von Altersverifikationsverfahren verpflichtet, wenn Plattformen mittlere bis hohe Risiken für die Entwicklung von Minderjährigen bergen. Durch die Immersion und Interaktion wird dies auf Metaversen regelmäßig zutreffen, sodass entsprechende Pflichten lediglich eine ausreichend effektive Rechtsdurchsetzung erfordern.

Doch selbst Altersverifikationssysteme, die den hier aufgeführten ethischen und rechtlichen Kriterien genügen, können nicht alle Gefahren für Kinder und Jugendliche in Metaversen beseitigen. Weitere Maßnahmen wie Inholdemoderation durch Betreiber sowie eine Sensibilisierung von Kindern und Jugendlichen können sie nicht ersetzen. Für einen wirksamen und nachhaltigen Schutz bei gleichzeitiger kritischer Befähigung braucht

es daher ein Zusammenspiel von technischen Maßnahmen, rechtlichen Rahmenbedingungen und pädagogischer Unterstützung.

Literatur

- Al-kfairy, Mousa; Alrabae, Saed und Alfandi, Omar (2025): Navigating the ethical landscape of metaverse. Challenges, solutions and governance initiatives. *IEEE Access*, Vol. 13, S. 79996–80018, 2025, <https://doi.org/10.1109/ACCESS.2025.3564498>.
- Booth, Robert (2025): Hack of age verification firm may have exposed 70,000 Discord users' ID photos. *The Guardian* vom 09.10.2025 URL: <https://www.theguardian.com/media/2025/oct/09/hack-age-verification-firm-discord-users-id-photos> (besucht am 29.01.2026).
- Brailovskaia, Julia; Buchmann, Johannes; Hertwig, Ralph; Metzinger, Thomas; Montag, Christian; Sadeghi, Ahmad-Reza; Schneider, Silvia; Spiecker gen. Döhm, Indra und Waldherr, Annie (2025): Soziale Medien und die psychische Gesundheit von Kindern und Jugendlichen. Diskussion Nr. 40, Halle (Saale): Nationale Akademie der Wissenschaften Leopoldina, URL: https://levana.leopoldina.org/servlets/MCRFileNodeServlet/leopoldina_derivate_01077/2025_Leopoldina_Diskussion_40.pdf (besucht am 29.01.2026).
- Council of Europe and IEEE (2024): The Metaverse and its impact on human rights, the rule of law and democracy. URL: <https://rm.coe.int/the-metaverse-and-its-impact-on-human-rights-the-rule-of-law-and-democ/1680b178b0> (besucht am 29.01.2026).
- Croll, Jutta und Krause, Thorsten (2024): Das Metaversum schlägt zurück?! Eine kinderrechtliche Perspektive. *merz. Zeitschrift für Medienpädagogik*, 68(3), S. 50–56. <https://doi.org/10.21240/merz/2024.3.16>.
- Datenschutzkonferenz (2026): Entschließung vom 10. März 2026. Stellungnahme zum Digital Fitness Check. Vorschläge zu Modifikationen der DSGVO. URL: https://www.datenschutzkonferenz-online.de/media/en/2026-03-10_DSK_Entschliessung-Digital_Fitness_Check.pdf.
- Dreyer, Stephan (21.03.2025): Virtuelle Schranken oder offene Scheunentore. Fünf Thesen zu verpflichtenden Altersüberprüfungen als Instrument des Jugendmedienschutzes. *Mediendiskurs*. URL: <https://mediendiskurs.online/beitrag/virtuelle-schranken-oder-offene-scheunentore-beitrag-772/> (besucht am 29.01.2026).
- Europäische Kommission (2025): Mitteilung der Kommission. Leitlinien (C/2025/5519) für Maßnahmen zur Gewährleistung eines hohen Maßes an Privatsphäre, Sicherheit und Schutz von Minderjährigen im Internet gemäß Artikel 28 Absatz 4 der Verordnung (EU) 2022/2065. (C/2025/5519). URL: https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:C_202505519 (besucht am 29.01.2026).
- Fiani, Cristina; Saeghe, Pejman; McGill, Mark und Khamis, Mohamed (2024): Exploring the perspectives of social VR-aware non-parent adults and parents on children's use of social virtual reality. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 54, S. 1–25 <https://doi.org/10.1145/3652867>

- Fu, Yue; Schwamm, Samuel; Baughan, Amanda; Powell, Nicole M.; Kronberg, Zoe; Owens, Alicia; Izenman, Emily R.; Alsabeh, Dania; Hunt, Elizabeth; Rich, Michael; Bickham, David; Radesky, Jenny und Hiniker, Alexis (2025): Understanding children's avatar making in social online games. <https://doi.org/10.48550/arXiv.2502.18705>.
- Geissler, Ines und Schlindwein, Lisa (2025): Datenschutz in Metaversen. In: Selzer, Annika (Hrsg.), *Aktuelle Entwicklungen des Rechtsrahmens der Cybersicherheit und Privatheit*, Sammelband. Wiesbaden: Springer-Vieweg, S. 1–10. https://doi.org/10.1007/978-3-658-49640-1_1.
- Jandt, Silke (2025): Gesetzliches Mindestalter für soziale Netzwerke in Deutschland? *Zeitschrift für Rechtspolitik ZRP*, 3/2025, S. 69–72.
- Kim, Soyeon (2023): Emergence of the metaverse and psychiatric concerns in children and adolescents. *Journal of the Korean Academy of Child and Adolescent Psychiatry*, 34(4), S. 215–222. <https://doi.org/10.5765/jkacap.230047>
- Krell, Felix und Wettmann, Nico (2023): Corporeal interactions in VRChat. Situational intensity and body synchronization. *Symbolic Interaction*, 46(2), S. 1–23. <https://doi.org/10.1002/symb.629>.
- Kunz, Thomas und Waldmann, Ulrich (2025a): SSI-basierte Altersverifikation und Herkunftsnachweise für Avatare zum Schutz von Minderjährigen in Metaversen. In: Lucke, Ulrike; Stieglitz, Stefan; Uebernickel, Falk; Lamprecht, Anna-Lena und Klein, Maik (Hrsg.): *INFORMATIK 2025*. Bonn: Gesellschaft für Informatik, S. 173–187. https://doi.org/10.18420/inf2025_10.
- Kunz, Thomas und Waldmann, Ulrich (2025b): Online-Altersverifikation zur Gewährleistung des Minderjährigenschutzes. *Datenschutz und Datensicherheit*, 49(6), S. 384–389. <https://doi.org/10.1007/s11623-025-2108-1>.
- Livingstone, Sonja; Stoilova, Mariya und Nandagiri, Rishita (2019): *Children's data and privacy online. Growing up in a digital age*. London: London School of Economics and Political Science.
- Meineck, Sebastian (28.04.2025): Die trügerische Sicherheit von Alterskontrollen im Netz. Mediendiskurs. URL: <https://mediendiskurs.online/beitrag/die-truegerische-sicherheit-von-alterskontrollen-im-netz-beitrag-772/> (besucht am 29.01.2026).
- Meta (2025): Content-Richtlinien für Meta Horizon Worlds. URL: <https://www.meta.com/de-de/help/quest/481214418021533/> (besucht am 29.01.2026).
- Niestadt, Maria (2024): Protecting children in virtual worlds (the metaverse). URL: [https://www.europarl.europa.eu/RegData/etudes/ATAG/2024/762294/EPRS_ATA\(2024\)762294_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2024/762294/EPRS_ATA(2024)762294_EN.pdf) (besucht am 29.01.2026).
- OECD (2021): Children in the digital environment. Revised typology of risks. URL: [https://one.oecd.org/document/DSTI/CDEP/DGP\(2020\)3/FINAL/en/pdf](https://one.oecd.org/document/DSTI/CDEP/DGP(2020)3/FINAL/en/pdf) (besucht am 15.04.2026).
- Open Letter (2026): Joint statement of security and privacy scientists and researchers on Age Assurance. URL: <https://csa-scientist-open-letter.org/ageverif-Feb2026> (besucht am 16.03.2026).

- Schellenberg, Stefan und Mikat, Claudia (2022): Die harten Inhalte finden sich auf Webseiten, nicht in Apps. Mediendiskurs, veröffentlicht im Februar 2022 URL: <https://mediendiskurs.online/beitrag/die-harten-inhalte-finden-sich-auf-webseiten-nicht-in-apps-beitrag-1024/> (besucht am 29.01.2026).
- Schnell, Rainer; Hill, Paul B. und Esser, Elke (2008): Methoden der empirischen Sozialforschung. München: Oldenbourg.
- Schumm, Daria; Müller, Katharina O. und Stiller, Burkhard (2025): Are we there yet? A study of decentralized identity applications. <https://doi.org/10.48550/arXiv.2503.15964>
- Sporny, Manu; Longley, Dave; Sabadello, Markus; Reed, Drummond; Steele, Orie und Allen, Christopher (2022): Decentralized identifiers (DIDs) v1.0. W3C recommendation. URL: <https://www.w3.org/TR/did-1.0/> (besucht am 29.01.2026).
- Sporny, Manu; Longley, Dave; Chadwick, David und Herman, Ivan (2025): Verifiable credentials data model v2.0. W3C recommendation. URL: <https://www.w3.org/TR/vc-data-model-2.0/> (besucht am 29.01.2026).
- Standards Australia (2024): Virtual fight for kids' safety. Standards Australia releases landmark draft standard to protect children in the metaverse. URL: <https://www.standards.org.au/news/virtual-fight-for-kids-safety-standards-australia-releases-landmark-draft-standard-to-protect-children-in-the-metaverse> (besucht am 29.01.2026).
- Stapf, Ingrid; Meinert, Judith; Heesen, Jessica; Krämer, Nicole; Ammicht Quinn, Regina; Bieker, Felix; Friedewald, Michael; Geminn, Christian; Martin, Nicholas; Nebel, Maxi; Ochs, Carsten; Watzinger, Lea und Bauer, Andreas (2023): Privacy and children's rights. Plattform Privatheit. <https://doi.org/10.24406/publica-793>.
- Stapf, Ingrid und Hecktor, Jana (2026): Digital Wellbeing von Kindern. Ethische Perspektiven auf die Social Media-Debatte angesichts aktueller Entwicklungen der KI. *Communicatio Socialis*, 59(1), S. 127–139.
- Stapf, Ingrid und Heesen, Jessica (2026): Digitale Kommunikation. Medienethik, Medienkompetenz. In: Bobbert, Monika und Sautermeister, Jochen (Hrsg.): *Handbuch Ethik und Psychologie*. Wiesbaden: Springer VS, S. 611–628.
- Steenbergen, Marina (2026): Die Altersfeststellung nach den Art. 28 DSA-Leitlinien im Metaverse. Das Ende der „Über-18-Jahre“-Buttons? *Datenschutz und Datensicherheit (DuD)*, 50(3), S. 152–157.
- UNICEF (2023): The metaverse, extended reality and children. URL: <https://www.unicef.org/innocenti/media/756/file/UNICEF-Innocenti-Rapid-Analysis-Metaverse-XR-and-children-2023.pdf> (besucht am 29.01.2026).
- VRChat (2025): Age Verification. URL: https://wiki.vrchat.com/wiki/Age_Verification (besucht am 29.01.2026).
- Wienrich, Karolin (15.11.2023): Psychologische Folgen. Vortrag im Rahmen der Veranstaltung „Lost in ‚Metaverse‘? Zur Verschränkung realer und digitaler Welten“. Deutscher Ethikrat, URL: <https://www.ethikrat.org/fileadmin/PDF-Dateien/Veranstaltungen/herbsttagung-2023-11-15-transkript.pdf> (besucht am 29.01.2026).
- Witzleb, Normann (2025): Children's privacy in XR applications. A rights-based approach. *Santa Clara High Technology Law Journal*, 41(1), S. 67–133. <https://doi.org/10.2139/ssrn.5287183>

Akzeptanz staatlicher Überwachung privater Telekommunikation in Deutschland: Ein faktorielles Survey-Experiment

Bruno Albert, Joanna Klauser, Andreas Hammer, Christian Lindenmeier, Sabine Pfeiffer, Dirk Heckmann und Felix Freiling

Zusammenfassung

Der Beitrag untersucht die gesellschaftliche Akzeptanz staatlicher Überwachungsmaßnahmen in Abhängigkeit von deren technischer und rechtlicher Ausgestaltung sowie dem Kontext ihrer Umsetzung. Zu diesem Zweck wurde ein faktorielles Survey-Experiment mit 1.254 Teilnehmenden in Deutschland durchgeführt. Dabei wurden der Kontext sowie die technische und rechtliche Ausgestaltung einer vorgeschlagenen Überwachungsmaßnahme für private Smartphones variiert. Die Ergebnisse zeigen, dass die Unterstützung für Überwachungsmaßnahmen im Kontext von Terrorismus höher ist als im Kontext von Wohnungseinbruchsdiebstahl oder Gewaltverbrechen. Darüber hinaus führt eine Beschränkung der Massenüberwachungstauglichkeit sowie der durch die Überwachungsmaßnahme erhobenen Daten zu einer höheren Unterstützung für die Maßnahme. Ebenfalls zu einer höheren Unterstützung führt eine verpflichtende gerichtliche Anordnung der Maßnahme. Die Ergebnisse deuten somit auf eine hohe Relevanz der technischen und rechtlichen Ausgestaltung der Maßnahmen für die gesellschaftliche Akzeptanz hin.

1. Einleitung

Unter dem Eindruck verschiedener sich gegenseitig überlagernder und gegenseitig bedingender Krisen und Krisenerzählungen erleben wir aktuell wieder verstärkte Diskussionen über die öffentliche Sicherheit (Perkowski/Epik 2024). Der allgemeine Trend zur Digitalisierung macht dabei auch vor der Kommunikationsüberwachung nicht halt. Auf der einen Seite steht der Wunsch der Sicherheitsbehörden nach mehr Möglichkeiten der Überwachung digitaler Kommunikation, um auf die zunehmende Verschiebung

von Kommunikation in den digitalen Raum zu reagieren. Auf der anderen Seite besteht die Sorge vor zu großen staatlichen Eingriffen in die Privatsphäre der Bürger:innen, die in der Angst vor einem „Überwachungsstaat“ (Lüdemann/Schlepper 2013) Ausdruck findet.

In einem demokratischen Staat ist es jedoch auch wichtig, dass staatliche Maßnahmen von den Bürger:innen angenommen, mitgetragen und akzeptiert werden (Lüdemann/Schlepper 2011, S. 119). Der vorliegende Beitrag untersucht daher die gesellschaftlichen Einstellungen gegenüber staatlichen Überwachungsmaßnahmen. Da Einstellungen gegenüber der Privatsphäre und staatlichen Maßnahmen, die diese potenziell einschränken, stark kontextabhängig sind (Wright u. a. 2009, S. 69), wurden die Einstellungen in der Befragung, die in diesem Beitrag vorgestellt wird, mithilfe eines faktoriellen Survey-Experiments erhoben. Das Experiment erlaubte es, den Kontext der Maßnahme sowie die technische und rechtliche Ausgestaltung der Maßnahme kontrolliert zu variieren und so den Einfluss auf die Einstellung der Befragten zu untersuchen.

Als gesellschaftliche Akzeptanz bezeichnen wir in diesem Beitrag die positive bzw. negative Einstellung der Bevölkerung gegenüber bestimmten Überwachungsmaßnahmen (Petermann/Scherz 2005, S. 46f.). Dabei ist es uns wichtig festzuhalten, dass grundgesetzliche Wertungen nicht durch die Akzeptanzforschung ersetzt werden können und sollen. Die Erkenntnisse zur gesellschaftlichen Akzeptanz sollten vielmehr als Orientierung bei der Ausfüllung des verfassungsrechtlich vorgezeichneten Spielraums dienen. Ziel des Beitrags ist deshalb auch explizit nicht, die „Grenzen des (gerade noch) Akzeptablen“ (Kühne 2019, S. 38) auszuloten, um die Einführung gerade noch akzeptabler Überwachungsmaßnahmen zu unterstützen. Vielmehr geht es uns darum, ein differenziertes Bild der Akzeptanz in der deutschen Gesellschaft zu zeichnen und dabei den Blick auf die konkrete rechtliche und technische Ausgestaltung der Maßnahmen zu richten. Dies ist nur durch eine interdisziplinäre Betrachtungsweise möglich. Aus diesem Grund wurden Perspektiven aus den Bereichen Informatik, Jura und Soziologie in die Gestaltung des Experiments einbezogen.

Die Ergebnisse des Survey-Experiments zeigen: die Massenüberwachungstauglichkeit einer Maßnahme und die Intensität des Eingriffs in die Privatsphäre haben einen signifikanten Einfluss auf die Akzeptanz der Maßnahme. Ebenfalls einen signifikanten Einfluss auf die Akzeptanz von Überwachungsmaßnahmen haben die richterliche Kontrolle bei der Anwendung der Maßnahme und der zugrundeliegende Straftatbestand. Wie die Überwachung und damit der Zugang zu einem Gerät jedoch technisch

realisiert wird und ob dabei künstliche Intelligenz zum Einsatz kommt, zeigt im Experiment keine Auswirkung auf die Akzeptanz der Maßnahme.

2. Überwachung und Privatheit – konzeptuelle Fundierung

Überwachung kann viele Formen annehmen und es existiert eine Vielzahl unterschiedlicher Definitionen dieses Phänomens. Für diesen Beitrag übernehmen wir die Definition von Überwachung als „einseitige systematische, routinemäßige Beobachtung von Einzelpersonen oder Gruppen zu einem bestimmten Zweck“ (Trüdinger/Steckermeier 2017, S. 421). Moderne Überwachungstechnologien können ohne Zustimmung oder Wissen der Betroffenen personenbezogene Daten von Einzelpersonen extrahieren (Marx 2015). Dies kann durch die Nutzung öffentlich zugänglicher Informationen, die Überwachung öffentlicher Räume, zum Beispiel durch Kameras, oder den technischen Zugriff auf die Daten oder Geräte der Überwachten erfolgen. Überwachung kann darüber hinaus von verschiedenen Akteur:innen ausgehen. Neben privaten Akteur:innen, geht Überwachung vor allem von staatlichen Akteur:innen aus, die Überwachungsmaßnahmen unter anderem als strategisches Mittel zur Informationsbeschaffung einsetzen (Trüdinger/Steckermeier 2017, S. 421). Dieser Beitrag fokussiert sich vor allem auf Telekommunikationsüberwachung, eine Form der Überwachung, die Maßnahmen zur Überwachung von privaten Nachrichtenübermittlungen mittels technischer Einrichtungen umfasst (Zdun 2006, S. 325).

Privatheit ist ebenfalls ein facettenreiches Phänomen, von dem in verschiedenen Disziplinen verschiedenste Definitionen existieren (Möller u. a. 2025). Das Grundgesetz definiert die Privatsphäre nicht ausdrücklich als Rechtsbegriff. Auch der hiervon zu unterscheidende Begriff der Privatheit verfügt rechtlich über keine einheitliche Definition. Aspekte der Privatheit sind z. B. in Art. 10 GG, Art. 13 GG und in beinahe jedem Grundrecht enthalten (Heckmann, in: Stern u. a. 2022; Geminn/Roßnagel 2015, S. 703). Ausdrücklich findet sich die Achtung der Privatheit im Unionsrecht in Art. 7 GRCh.

Finn u. a. (2012, S. 7ff.) identifizieren sieben Typen von Privatheit. Von diesen sieben Typen ist für diesen Artikel vor allem der Typ „Privatheit von Kommunikation“ relevant. Privatheit von Kommunikation ist für eine demokratische Gesellschaft essenziell, da sie die Grundlage für eine freie Diskussionskultur bildet (Greenwald 2014). Gerade deshalb sind Einstellungen von Bürger:innen gegenüber Überwachungsmaßnahmen auch in Demo-

kratien von besonderer Relevanz und bereits Gegenstand verschiedener Studien (z. B. Davis/Silver 2004; Hallinan u. a. 2012; Lüdemann/Schlepper 2013). Allerdings behandeln viele Studien Überwachung als ein abstraktes Konzept und reduzieren die Einstellung dazu auf einen Wertekonflikt zwischen Freiheit und Sicherheit (z. B. Davis/Silver 2004; Jäger 2023). Die Studien nehmen dabei an, dass Bürger:innen rational die erwarteten Sicherheitsgewinne gegen die erwarteten Freiheits- und Privatheitseinschränkungen abwägen. Diese Annahme spiegelt sich oft auch im Studiendesign wider, in dem Befragte gebeten werden, ihre Präferenz zwischen Freiheit und Sicherheit anzugeben (Davis/Silver 2004).

Dabei wird davon ausgegangen, dass Bürger:innen den Zusammenhang zwischen Freiheit und Sicherheit als Nullsummenspiel verstehen. Studien zeigen allerdings, dass dies nicht der Fall ist. Die meisten Bürger:innen erwarten von ihren Regierungen, dass Überwachungsmaßnahmen so gestaltet sind, dass sie sowohl die Sicherheit erhöhen als auch die Privatsphäre schützen (Friedewald u. a. 2014; Pavone/Degli Esposti 2012). Daher ist es wichtig zu verstehen, wie staatliche Überwachungsmaßnahmen ausgestaltet und kontrolliert werden können, um potenzielle Risiken für die Privatsphäre möglichst zu minimieren, und wie sich diese Ausgestaltung auf die gesellschaftliche Einstellung gegenüber den Maßnahmen auswirkt.

3. Überwachungsmaßnahmen und Einstellung: Forschungsstand und Hypothesen

Ein weiterer Kritikpunkt an der Einstellungsforschung zu Überwachung als abstraktes Konzept ist, dass der Kontext, in dem die Überwachungsmaßnahme stattfindet, häufig ignoriert wird (Degli Esposti u. a. 2021; Wright u. a. 2009). Dabei wird vernachlässigt, dass Einstellungen gegenüber Überwachungsmaßnahmen nicht einfach existent sind, sondern Ausdruck der Interaktion zwischen Bürger:innen und sozialen und institutionellen Kontexten (Dourish/Anderson 2006).

Es ist daher unwahrscheinlich, dass alle Bürger:innen eine generelle Einstellung gegenüber allen staatlichen Überwachungsmaßnahmen entwickeln. Sie können eine bestimmte Überwachungsmaßnahme in einem bestimmten Kontext für gerechtfertigt halten, während sie eine andere Überwachungsmaßnahme im selben Kontext oder dieselbe Überwachungsmaßnahme in einem anderen Kontext ablehnen würden. Eine Möglichkeit diesen Umstand zu beforschen, ist der Vergleich von Einstellungen gegen-

über unterschiedlichen Überwachungstechnologien wie zum Beispiel Videoüberwachung und Online-Überwachung (Antoine 2023; Van den Broek u. a. 2017). Dabei wird allerdings außer Acht gelassen, dass diese Technologien unterschiedlich ausgestaltet sein können. Diese Ausgestaltungen könnten einen großen Einfluss auf die potenzielle Bedrohung der Privatheit und damit auch auf die Akzeptanz der Bevölkerung haben. Ein Video-Überwachungssystem kann beispielsweise Aufnahmen nur live übertragen oder sie auch speichern, zudem könnte automatisierte Gesichtserkennung zum Einsatz kommen (Li 2024).

Welchen Einfluss der jeweilige Kontext und die Ausgestaltung einer Überwachungsmaßnahme auf die Einstellung der Bürger:innen hat, soll daher in diesem Beitrag anhand der staatlichen Telekommunikationsüberwachung überprüft werden. Telekommunikation genießt sowohl nach dem Grundgesetz gemäß Art. 10 GG als auch nach der Grundrechtecharta der Union gemäß Art. 7 GRCh einen besonderen Schutz. Nach der jüngeren Rechtsprechung des Bundesverfassungsgerichts ist zudem, wenn zur Erfassung von Kommunikation ein Eingriff in das Endgerät erfolgt, auch das IT-System-Grundrecht gemäß Art. 2 Abs. 1 GG i. V. m. Art. 1 Abs. 1 GG einschlägig (BVerfG, Beschluss vom 24.06.2025 – 1 BvR 2466/19).

Einstellungen zu Privatheit und Maßnahmen, die diese potenziell einschränken, sind stark kontextabhängig (Wright u. a. 2009). Zum Kontext der Maßnahme gehört unter anderem der Grund, aus welchem die Maßnahme durchgeführt werden soll. Er stellt gewissermaßen den staatlichen Anlass für Eingriffe in die Privatsphäre dar (vgl. Lyon 2007). In der Debatte um staatliche Überwachungsmaßnahmen werden verschiedene Gründe angeführt. Während infolge der Terroranschläge des 11. September 2001 die Notwendigkeit von (weiteren) staatlichen Überwachungsmaßnahmen mit dem Schutz vor terroristischen Aktivitäten begründet wurde (Lyon/Haggerty 2012), spielt in der aktuellen Debatte auch der Schutz vor (organisierter) Kriminalität eine Rolle (Gruber u. a. 2022).

Diese unterschiedlichen Kontexte spiegeln sich auch in der Forschung wider. Die meisten Studien beforschen die Einstellungen zu staatlichen Überwachungsmaßnahmen allerdings nur in einem der beiden Kontexte, entweder im Kontext von Terrorismus (Garcia/Geva 2016; Trüdinger/Ziller 2022; Ziller/Helbling 2021) oder Kriminalität (Antoine 2023). Dabei wird der Effekt des Kontextes meist nicht direkt ermittelt oder adressiert. Eine Studie von Mondak und Hurwitz (2012) verglich die beiden Kontexte und untersuchte dabei nicht die Einstellung gegenüber Überwachung, sondern gegenüber anderen Maßnahmen, die die bürgerlichen Freiheiten

einschränken, wie beispielsweise Hausdurchsuchungen. Die Autoren stellen fest, dass der Kontext der Maßnahme einen messbaren Effekt auf die Zustimmung hat, die Maßnahmen im Zusammenhang mit Terrorismus erhielten eine höhere Zustimmung. Es ist plausibel anzunehmen, dass dieser Effekt bei Einstellungen zur Überwachung ebenfalls auftritt. Es ist allerdings auch möglich, dass Bürger:innen Überwachungsmaßnahmen im Kontext der Kriminalitätsbekämpfung eher akzeptieren, da die Wahrscheinlichkeit, selbst von Kriminalität betroffen zu sein und somit auch der potenzielle persönliche Nutzen für die eigene Sicherheit deutlich größer ist. Dies führt zu zwei möglichen Hypothesen:

H1a: Die Nutzung staatlicher Überwachung zur Bekämpfung von Terrorismus führt zu einer höheren Akzeptanz.

H1b: Die Nutzung staatlicher Überwachung zur Bekämpfung von Kriminalität führt zu einer höheren Akzeptanz.

Neben dem Kontext bestimmt die Ausgestaltung von Überwachungsmaßnahmen maßgeblich die potenzielle Gefahr, die von den Maßnahmen auf die Privatheit ausgeht. Frühere Studien zur Einstellung gegenüber staatlicher Überwachung identifizieren zwei Aspekte der Ausgestaltung, die die Einstellung der Öffentlichkeit gegenüber diesen Technologien stark beeinflussen: die Reichweite der Maßnahme und die Tiefe des Eingriffs in die Privatsphäre. Auch im Verfassungsrecht werden die Art und der Umfang der Datenerhebung als vertiefende Aspekte eines Grundrechtseingriffs berücksichtigt (BVerfG, Beschluss vom 24.06.2025 – 1 BvR 180/23, Rn. 187). Gleiches gilt für eine große Streubreite von Maßnahmen (BVerfGE 154, 152).

Je mehr Menschen potenziell von einer Überwachungsmaßnahme betroffen sind, desto größer ist die Gefahr anlassloser Massenüberwachung. Die Reichweite einer Überwachungsmaßnahme – gemessen als Anteil der davon betroffenen Bevölkerung – hat nachweislich einen erheblichen Einfluss auf die Einstellung gegenüber dieser Maßnahme: Während gezielte Maßnahmen, die sich gegen bestimmte Verdächtige richten, positiver bewertet werden, zeigt sich für Überwachungsmaßnahmen, die große Teile der Bevölkerung betreffen, weniger Akzeptanz (Antoine 2023; Jäger 2023; Ziller/Helbling 2021). Dies deckt sich mit den hohen verfassungsrechtlichen Hürden, welche von der Rechtsprechung an anlasslose Maßnahmen von hoher Streubreite gestellt werden (BVerfGE 154, 152, Rn. 168ff.; BVerfGE 125, 260, 320f.; differenzierend EuGH, Urteil vom 30.04.2024 – C-470/21). Unsere zweite Hypothese lautet daher:

H2: Eine Beschränkung der Reichweite der Maßnahme führt zu einer größeren Akzeptanz.

Abhängig davon, welche Daten bei einer Überwachungsmaßnahme erfasst werden, kann aus verfassungsrechtlicher Perspektive die Tiefe des mit der Maßnahme verbundenen Grundrechtseingriffs variieren. Der Eingriff in die Privatsphäre ist bei staatlicher Überwachung notwendig, da Überwachungsmaßnahmen explizit zum Ziel haben, Informationen über die Absichten und Pläne von Kriminellen oder Terrorist:innen zu ermitteln. Trotzdem obliegt es den staatlichen Stellen, den Eingriff in die Privatsphäre auf ein notwendiges Mindestmaß zu begrenzen. Dies ergibt sich aus dem Verhältnismäßigkeitsgrundsatz des Verfassungsrechts bei Grundrechtseingriffen. Vorangegangene Studien zeigten einen Zusammenhang zwischen der Sensibilität der bei der Überwachung erhobenen Daten und der öffentlichen Einstellung. So ist die öffentliche Akzeptanz für die Erfassung von Adressdaten höher als für die Überwachung von Internetnutzungsdaten (Antoine 2023). Unsere dritte Hypothese lautet daher:

H3: Eine Begrenzung des Ausmaßes der Eingriffe in die Privatsphäre führt zu einer höheren Akzeptanz.

Bürger:innen bilden ihre Einstellungen gegenüber staatlichen Überwachungsmaßnahmen in einem Kontext der Unsicherheit. Sie können den Einsatz der Maßnahme und die daraus resultierenden Auswirkungen nicht direkt kontrollieren. Dies gilt im Prinzip für jede Form des staatlichen Handelns, besonders aber für Überwachungsmaßnahmen, da diese oft verdeckt erfolgen und sich der Kontrolle damit aktiv entziehen (Trüdinger/Steckermeier 2017). Gleichzeitig sind die möglichen Konsequenzen, die sich aus einer Gefährdung der Privatheit bei einer missbräuchlichen Verwendung durch staatliche Stellen ergeben, besonders schwerwiegend, da diese meist über erhebliche Ressourcen verfügen. Folglich sind die Bürger:innen auf Dritte angewiesen, die die Anwendung von Überwachungsmaßnahmen in ihrem Namen kontrollieren. In einem demokratischen System sind die anderen Staatsgewalten für die Kontrolle der Exekutive zuständig. In diesem Zusammenhang nimmt die Justiz in Form der richterlichen Kontrolle eine Sonderstellung ein (Finn u. a. 2012, S. 8).

Pavone und Degli Esposti (2012) zeigten in einer qualitativen Studie mit spanischen Teilnehmenden, dass Bürger:innen die richterliche Kontrolle von Überwachungsmaßnahmen erwarten und in die Bewertung der Maßnahmen mit einbeziehen (Pavone/Degli Esposti 2012). Reddick und andere

(2015) zeigten außerdem in einer quantitativen Befragung in den USA, dass Bürger:innen, die davon ausgehen, dass Gerichte angemessene Grenzen für staatliche Überwachung setzen, eher dazu neigten, die staatlichen Überwachungsmaßnahmen zu billigen. Dies führt zu unserer vierten Hypothese:

H4: Die Notwendigkeit einer richterlichen Anordnung zur Durchführung der Überwachungsmaßnahme führt zu einer höheren Akzeptanz.

Die zunehmende Nutzung von Ende-zu-Ende-verschlüsselter digitaler Kommunikation stellt staatliche Behörden heutzutage vor große Probleme, denn klassische Telekommunikationsüberwachungsmaßnahmen, die die Kommunikation „von der Leitung“ abgreifen, können aus dem verschlüsselten Datenstrom kaum nützliche Informationen ableiten. Neuere Telekommunikationsüberwachungsmaßnahmen, wie etwa die Quellen-Telekommunikationsüberwachung oder die Online-Durchsuchung, funktionieren jedoch nur dann, wenn es den Behörden gelingt, eine (oft als Staatstrojaner bezeichnete) Software auf den verwendeten Endgeräten der Benutzer:innen auszuführen. Die Einbringung einer solchen Software gestaltet sich in der Praxis, beispielsweise aufgrund des hohen Sicherheitsniveaus moderner Smartphones, als schwierig. In der Praxis werden hierfür derzeit häufig nicht behobene Software-Sicherheitslücken genutzt, um Staatstrojaner zu installieren (Lawall/Scharping 2024, S. 606). Diese Praxis stellt ein erhebliches Risiko für andere Nutzende dar, da Kriminelle die nicht behobenen Software-Sicherheitslücken ebenfalls ausnutzen könnten (Vettermann u. a. 2023, S. 7). Behörden stehen daher vor dem Dilemma, die Öffentlichkeit durch die Schließung solcher Sicherheitslücken zu schützen und gleichzeitig den Fernzugriff für strafrechtliche Ermittlungen zu ermöglichen. Ein weiteres Problem betrifft den Erwerb von Kenntnissen über Schwachstellen. Dies führt häufig dazu, dass Regierungsbehörden an einem kriminellen Markt teilnehmen, auf dem diese Schwachstellen gehandelt werden (Lindenmeier u. a. 2024a, S. 27:2).

Zuletzt wurden jedoch neue technische Ansätze für den Zugriff auf Endgeräte vorgeschlagen, die eine an die klassische Telekommunikationsüberwachung angelehnte Mitwirkung der Softwareherstellenden vorsehen (Lindenmeier u. a. 2024a; 2024b). Da die Nutzung nicht behobener Software-Schwachstellen Risiken für die Öffentlichkeit mit sich bringt, ist zu erwarten, dass diese Praxis die öffentliche Akzeptanz verringern wird. Dies führt zur fünften Hypothese:

H5: Die Nutzung von nicht behobenen Software-Sicherheitslücken führt zu einer niedrigeren Akzeptanz.

Die rasanten Fortschritte in den Bereichen künstliche Intelligenz (KI) und maschinelles Lernen wirken sich auf fast alle Bereiche der Technologie aus. Überwachungstechnologien bilden dabei keine Ausnahme. KI-Systeme sind dabei vor allem im Bereich der automatisierten Datenauswertung relevant, da digitale Überwachungssysteme große Datenmengen produzieren. KI-Systeme könnten theoretisch aber auch genutzt werden, um die Privatsphäre der Betroffenen zu schützen, indem beispielsweise besonders schützenswerte Inhalte, die nicht mit dem Straftatbestand in Verbindung stehen, durch eine KI ausgefiltert werden, bevor ein Mensch die Daten zu Gesicht bekommt (Gleizer 2020).

Da die großen Fortschritte im Bereich von KI erst verhältnismäßig kurz zurückliegen, gibt es noch wenige Untersuchungen zu den möglichen Auswirkungen auf die Einstellung gegenüber Überwachungsmaßnahmen. Die wenigen vorliegenden Befunde sind dabei uneinheitlich. Nur eine Minderheit der deutschen Befragten steht dem Einsatz von KI durch Sicherheitsbehörden positiv gegenüber, die Tendenz ist jedoch steigend (Marcinkowski u. a. 2026). Eine Studie zur Akzeptanz von Videoüberwachung im Kontext von öffentlichem Nahverkehr stellte fest, dass die Befragten fortschrittlichere Kamertechnologien mit Echtzeit-Gesichtserkennung bevorzugen (Patil u. a. 2014). Andere Studien haben jedoch gezeigt, dass der Kontext und der Ort der Überwachung einen starken Einfluss auf die Einstellungen haben. Die Akzeptanz von Überwachung ist im öffentlichen Raum größer (Van Heek u. a. 2017). Es ist daher unklar, ob diese Ergebnisse auf den Kontext der Überwachung privater Telekommunikation übertragbar sind.

Ein weiteres Argument, das vor allem von Mitarbeiter:innen staatlicher Behörden vorgebracht wird, lautet, dass Betroffene die Überwachung als weniger invasiv empfinden, wenn personenbezogene Daten nicht von einem Menschen eingesehen, sondern automatisch von einer Maschine analysiert werden (Cayford/Pieters 2018, S. 97; Moll/Schneider 2021, S. 102).

Diese Befunde führen zu drei alternativen Hypothesen:

H6a: Der Einsatz künstlicher Intelligenz in Überwachungsmaßnahmen führt generell zu einer geringeren Akzeptanz.

H6b: Der Einsatz künstlicher Intelligenz in Überwachungsmaßnahmen führt generell zu einer höheren Akzeptanz.

H6c: Der Einsatz künstlicher Intelligenz in Überwachungsmaßnahmen führt nur dann zu einer größeren öffentlichen Akzeptanz, wenn durch den Einsatz von künstlicher Intelligenz die Privatsphäre der Betroffenen geschützt werden soll.

4. Datenbasis und methodisches Vorgehen

Um die aufgestellten Hypothesen empirisch zu überprüfen, wurde im Februar und März 2025 in Deutschland ein faktorielles Survey-Experiment durchgeführt. Bei diesem wurden den Befragten verschiedene Versionen einer möglichen Überwachungsmaßnahme präsentiert. Dabei wurden bestimmte Abschnitte des Textes, welche die rechtliche und technische Ausgestaltung der Maßnahme beschreiben, experimentell variiert. Die einzelnen Versionen des Beschreibungstexts werden als „Vignetten“ bezeichnet. Anschließend gaben die Befragten ihre Zustimmung zu der Maßnahme an. Dies erlaubt uns, den Einfluss der rechtlichen und technischen Ausgestaltung auf die Zustimmung zu messen (Auspurg/Hinz 2015). Die Befragung wurde mit Hilfe eines Online-Access Panel Anbieters (Norstat) realisiert. Bei der Kontaktierung der Befragten wurden Quoten für Alter, Geschlecht und Bildung eingesetzt. Die Stichprobe entspricht hinsichtlich der Verteilung dieser Merkmale der deutschen Gesamtbevölkerung. Insgesamt haben 1.254 Personen an der Umfrage teilgenommen.

Im vorherigen Abschnitt haben wir argumentiert, dass die Einstellung gegenüber staatlichen Überwachungsmaßnahmen in der Bewertung einer komplexen Situation entsteht, in der soziale und institutionelle Kontexte eine Rolle spielen (Wright u. a. 2009). Wenn sich Bürger:innen eine Meinung zu einer bestimmten Überwachungsmaßnahme bilden, müssen sie verschiedene Faktoren abwägen. Dazu gehören die Art, Leistungsfähigkeit und Eingriffstiefe der eingesetzten Technologie und die daraus resultierenden Gefahren für die Privatsphäre sowie die Gründe für die Überwachungsmaßnahme.

Um diese Abwägung für die Befragten greifbar zu machen, erfassen wir ihre Einstellungen anhand des konkreten Beispiels der polizeilichen Überwachung privater Smartphones. Die Überwachung privater Endgeräte kann dabei als „least-likely case“ für die Akzeptanz der Maßnahme angesehen werden. Vorangegangene Studien zeigen zum einen, dass heimliche Überwachung im privaten Raum deutlich stärker abgelehnt wird als die sichtbare Überwachung öffentlicher Räume (Pavone/Degli Esposti 2012;

Van den Broek u. a. 2017; Ziller/Helbling 2021). Konsequenterweise stellte Trüdinger (2018) in einer repräsentativen Umfrage fest, dass die Überwachung privater Telekommunikation die am wenigsten unterstützte staatliche Überwachungsmaßnahme in der deutschen Bevölkerung darstellt. Wenn Bürger:innen also bereit sind, staatliche Maßnahmen zur Überwachung privater Telekommunikation unter bestimmten Bedingungen zu unterstützen, ist es wahrscheinlich, dass diese Bedingungen auch bei weniger privatsphäreintensiven Eingriffen zu einer höheren Akzeptanz führen könnten.

Bei der Befragung kam ein vollfaktorielles Survey-Experiment ($3 \times 2 \times 2 \times 2 \times 3 \times 2$) (Auspurg/Hinz 2015) zum Einsatz. Aus den einzelnen Dimensionen des Experiments wurden 144 einzigartige Vignettentexte erstellt. In diesen Vignettentexten wurde ein hypothetisches Gesetz beschrieben, das den Einsatz einer Überwachungsmaßnahme vorsieht. Im Anschluss konnten die Befragten auf einer siebenstufigen Likert-Skala angeben, ob sie das Gesetz eher unterstützen oder ablehnen. Die Befragten erhielten zufällig drei der 144 Vignetten zur Bewertung, was zu einer Anzahl von 3.762 bewerteten Vignetten führte. Aufgrund fehlender Werte können nur 3.760 der bewerteten Vignetten in der folgenden Analyse verwendet werden.

Tabelle 1: Dimensionen und Level des Experiments

Dimension	Level 1	Level 2	Level 3
Kontext	Zur Bekämpfung von Terrorismus	Zur Bekämpfung von Gewaltverbrechen	Zur Bekämpfung von Wohnungseinbruchdiebstahl
Reichweite	Technisch wäre es möglich mit der Überwachungsmaßnahme große Teile der Bevölkerung zu überwachen	Die Überwachungsmaßnahme wird technisch so beschränkt, dass eine massenhafte Überwachung von großen Teilen der Bevölkerung nicht möglich ist	
Ausmaß Eingriff Privatsphäre	Nachrichten von Messengerdiensten (z.B. WhatsApp, Telegram, Signal) einzusehen	auf alle Inhalte eines Smartphones (z.B. Fotos, Videos & Textnachrichten) zuzugreifen	
Richterliche Anordnung	Für den Einsatz der Software ist ein richterlicher Beschluss notwendig	Für den Einsatz der Software ist kein richterlicher Beschluss notwendig	

Dimension	Level 1	Level 2	Level 3
Ausnutzung Software Schwachstellen	Dabei werden technische Schwachstellen genutzt, die auch von Kriminellen ausgenutzt werden könnten	Dabei werden die Hersteller der Smartphones zur Zusammenarbeit mit der Polizei verpflichtet.	
KI-Einsatz	Um die Ergebnisse der Auswertung zu verbessern, soll künstliche Intelligenz eingesetzt werden	Mithilfe von künstlicher Intelligenz sollen private Inhalte aussortiert werden, bevor ein Mensch die Daten betrachtet	Keine Angabe

Die Details der Überwachungsmaßnahme werden experimentell über sechs Dimensionen hinweg variiert (s. Tab. 1). Die Dimensionen umfassen den strafrechtlichen Kontext der Maßnahme (Kriminalität vs. Terrorismus), die Reichweite in Form der Massenüberwachungstauglichkeit der Maßnahme, das Ausmaß des Eingriffes in die Privatsphäre, die Notwendigkeit einer richterlichen Anordnung der Maßnahme, die Ausnutzung von Schwachstellen für den Zugriff sowie den Einsatz von künstlicher Intelligenz. Der folgende Absatz enthält ein Beispiel für eine Vignette, wobei die manipulierten Dimensionen kursiv hervorgehoben sind.

Zur Bekämpfung von *Gewaltverbrechen* soll ein Gesetz verabschiedet werden, das der Polizei ermöglicht *Nachrichten von Messengerdiensten (z. B. WhatsApp, Telegram, Signal) einzusehen*. Dabei werden *die Hersteller der Smartphones zur Zusammenarbeit mit der Polizei verpflichtet*. Technisch wäre es möglich, mit der Überwachungsmaßnahme *große Teile der Bevölkerung zu überwachen*. *Mithilfe von künstlicher Intelligenz sollen private Inhalte aussortiert werden, bevor ein Mensch die Daten betrachtet*. Für den Einsatz der Software ist *ein richterlicher Beschluss* notwendig.

Im Folgenden gehen wir näher auf die einzelnen Dimensionen ein, die im Experiment manipuliert werden. Jede Dimension kann bis zu drei Ausprägungen annehmen. Bei der ersten Dimension handelt es sich um den Kontext der Maßnahme. Neben dem Unterschied zwischen Terrorismus und Alltagskriminalität interessieren wir uns außerdem dafür, welchen Einfluss die Verletzung der körperlichen Unversehrtheit auf die Zustimmung zur Maßnahme hat. Daher wurden im Experiment die Bekämpfung von Terrorismus, die Bekämpfung von Gewaltverbrechen und die Bekämpfung von Wohnungseinbruchsdiebstahl als Gründe für die Maßnahme genannt.

Die zweite und dritte Dimension beschreiben die technischen und rechtlichen Beschränkungen hinsichtlich der Reichweite und der Eingriffstiefe

der Überwachungsmaßnahme. Die Reichweite der Maßnahme wird durch die Eignung der Maßnahme zur Massenüberwachungstauglichkeit operationalisiert. Die Eingriffstiefe in die Privatsphäre wird anhand der Menge der von der Maßnahme betroffenen privaten Daten operationalisiert. Die erste Ausprägung beschreibt eine Situation, in der der Zugriff auf Kommunikation beschränkt ist. Dies entspricht den erhobenen Daten bei der Telekommunikationsüberwachung, zum Beispiel nach § 100a StPO. Die zweite Ausprägung beschreibt eine Situation, in der die Polizei Zugriff auf alle Informationen auf dem Gerät hat, einschließlich privater Daten, die nicht Teil eines Kommunikationsprozesses waren. Die erhobenen Daten in der zweiten Stufe entsprechen der Online-Durchsuchung, zum Beispiel nach § 100b StPO.

Die Ausnutzung von unbekannten Software-Schwachstellen zur Ermöglichung von staatlichen Maßnahmen wird vor allem deswegen kritisiert, weil dies in der Konsequenz dazu führt, dass diese Schwachstellen nicht sofort geschlossen werden, was möglicherweise durch Kriminelle ausgenutzt werden könnte (Klauser u. a. 2026). Da dieser Diskurs vermutlich nicht allen Befragten bekannt ist, wurde die mögliche Konsequenz dieser Praxis im Vignettentext erwähnt. Als Alternative zur Nutzung von Schwachstellen wird in der zweiten Ausprägung die Verpflichtung der Herstellenden zur Kooperation beim Zugang zum Gerät (Lindenmeier u. a. 2024a, S. 27:6) genannt.

Der Einsatz von KI wird über drei Ebenen variiert. Die erste Ausprägung beschreibt den Einsatz von KI mit dem Ziel, die Privatsphäre besser zu schützen. Dies soll dadurch erreicht werden, dass eine KI besonders sensible Daten, die nichts mit der polizeilichen Ermittlung zu tun haben, ausfiltert, bevor ein Mensch sie zu Gesicht bekommt. Die zweite Ausprägung beschreibt den Einsatz von KI zur Verbesserung der Datenauswertung. Zum Vergleich wird die Nutzung von KI in der dritten Ausprägung nicht erwähnt. Die letzte Dimension beschreibt, ob für die Durchführung der Überwachungsmaßnahme eine richterliche Anordnung erforderlich ist. Zur Überprüfung der aufgestellten Hypothesen wurde eine OLS-Regression mit clusterrobusten Standardfehlern auf Befragtenebene durchgeführt.

5. Ergebnisse

Von allen experimentellen Varianten der vorgeschlagenen Überwachungsmaßnahmen wurden 47,4% der Vignetten von den Befragten zustimmend

bewertet (Antworten über der mittleren Kategorie der siebenstufigen Antwortskala), während 32,5% ablehnend bewertet wurden. 20,1% der Vignetten erhielten unentschiedene Bewertungen in der mittleren Kategorie. Diese Zustimmungsrate entspricht den Ergebnissen anderer faktorieller Experimente zur Akzeptanz von Überwachungsmaßnahmen in Deutschland (Antoine 2023; Ziller/Helbling 2021).

Abb. 1 zeigt die Ergebnisse des faktoriellen Survey-Experiments. Auf der vertikalen Achse sind die Dimensionen des Experiments aufgeführt, darunter eingeschränkte Massenüberwachungstauglichkeit, eingeschränkter Datenzugriff, die Notwendigkeit eines richterlichen Beschlusses, Nutzung von Schwachstellen, Wohnungseinbruchdiebstahl, Gewaltverbrechen sowie KI zur Auswertung und KI zum Schutz der Privatsphäre. Auf der horizontalen Achse sind geschätzte Effekte relativ zu einer vertikalen gestrichelten Nulllinie dargestellt. Punkte mit Linien zeigen Effekte und Konfidenzintervalle. Effekte, bei denen das Konfidenzintervall nicht die Null-Linie schneidet, können als signifikant betrachtet werden. Das bedeutet, dass es unwahrscheinlich ist, dass die Effekte zufällig in der Stichprobe aufgetreten sind. Eingeschränkter Datenzugriff, eingeschränkte Massenüberwachungstauglichkeit und notwendiger richterlicher Beschluss weisen signifikant positive Effekte auf. Diese Dimensionen haben die Zustimmung zur Maßnahme signifikant erhöht. Signifikant negative Effekte zeigen hingegen Wohnungseinbruchdiebstahl und Gewaltverbrechen im Vergleich zur Referenzkategorie Terrorismus. Daraus lässt sich ableiten, dass die Zustimmung zur Maßnahme bei der Präsentation der Maßnahme im Kontext der beiden genannten Verbrechen signifikant niedriger war als im Kontext von Terrorismus. Die Effekte für die Nutzung von Schwachstellen und den Einsatz von KI sind eher klein und nicht signifikant, wir können uns daher nicht sicher sein, ob die Effekte durch Zufall in unserer Stichprobe entstanden sind.

Wenn der Kontext der vorgeschlagenen Überwachungsmaßnahme Kriminalität statt Terrorismus ist, ist die Akzeptanz der Maßnahme deutlich geringer (s. Abb. 1). Dies gilt für beide vorgestellten Straftaten: Einbruch und Gewaltverbrechen. Hypothese 1b muss daher verworfen werden (s. Tab. 2). Die Akzeptanz der Maßnahme ist zudem deutlich höher, wenn ihre Reichweite beschränkt ist. Dieses Ergebnis steht im Einklang mit Hypothese 2. Der gleiche Effekt tritt auch bei einer Begrenzung des Ausmaßes des Eingriffs in die Privatsphäre auf, was mit der dritten Hypothese übereinstimmt. Wie in Hypothese vier postuliert, erhöht die Notwendigkeit einer richterlichen Anordnung für den Einsatz der Überwachungsmaßnahme die

Akzeptanz deutlich. Die Nutzung von Software-Schwachstellen hat hingegen keinen signifikanten Einfluss auf die Akzeptanz. Die fünfte Hypothese konnte somit nicht bestätigt werden. Ebenfalls keinen signifikanten Einfluss hat der Einsatz von künstlicher Intelligenz. Weder zur Verbesserung des Schutzes der Privatsphäre noch zur Verbesserung der Auswertung. Die Hypothesen zur Nutzung von künstlicher Intelligenz konnten somit im Experiment ebenfalls nicht bestätigt werden.

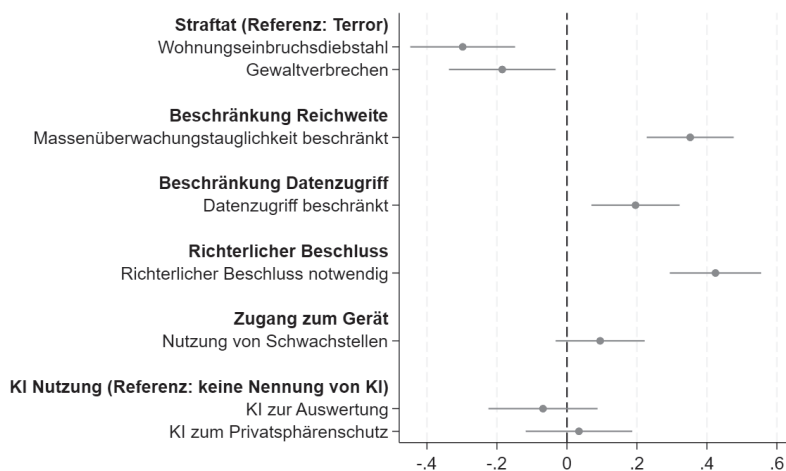


Abbildung 1: Regressionskoeffizienten mit 95%-Konfidenzintervallen

Zur Überprüfung der Robustheit der Ergebnisse wurde ein Regressionsmodell mit zusätzlichen Kontrollvariablen berechnet. In diesem Modell wurden Geschlecht, Alter, Haushaltseinkommen, Bildung, Region (Ost- oder Westdeutschland), Migrationshintergrund (ob die Person selbst oder ihre Eltern migriert sind) sowie die Zugehörigkeit zu einer religiösen Minderheit berücksichtigt. Alter hat in diesem Modell, unabhängig von der Ausgestaltung der Maßnahme, einen signifikanten positiven Einfluss auf die Akzeptanz. Dieses Ergebnis deckt sich mit anderen Studien zur Akzeptanz von Überwachungsmaßnahmen in Deutschland (Antoine 2023, S. 430; Bug/Bukow 2017, S. 203; Friedewald u. a. 2015, S. 70; Kostka u. a. 2023, S. 8). Das Haushaltseinkommen und der Bildungsstand haben, unabhängig von der Ausgestaltung der Maßnahme, einen signifikanten negativen Einfluss auf die Akzeptanz. Auch dieses Ergebnis stimmt mit der Literatur

überein (Friedewald u. a. 2015, S. 70; Trüdinger/Steckermeier 2017, S. 426). Trotz der signifikanten Effekte bei einigen der Kontrollvariablen weichen die Effekte der Vignetten-Dimensionen im Kontrollmodell nicht signifikant vom unkontrollierten Modell ab.

Tabelle 2: Ergebnisse der Hypothesenprüfung

Hypothesen	Unterstützung der Hypothese
H1a: Die Nutzung staatlicher Überwachung zur Bekämpfung von Terrorismus führt zu einer höheren Akzeptanz.	✓
H1b: Die Nutzung staatlicher Überwachung zur Bekämpfung von Kriminalität führt zu einer höheren Akzeptanz.	✗
H2: Eine Beschränkung der Reichweite der Maßnahme führt zu einer höheren Akzeptanz.	✓
H3: Eine Begrenzung des Ausmaßes der Eingriffe in die Privatsphäre führt zu einer höheren Akzeptanz.	✓
H4: Die Notwendigkeit einer richterlichen Anordnung zur Durchführung der Überwachungsmaßnahme führt zu einer höheren Akzeptanz.	✓
H5: Die Nutzung von nicht behobenen Software-Sicherheitslücken führt zu einer niedrigeren Akzeptanz.	✗
H6a: Der Einsatz künstlicher Intelligenz in Überwachungsmaßnahmen führt generell zu einer geringeren Akzeptanz.	✗
H6b: Der Einsatz künstlicher Intelligenz in Überwachungsmaßnahmen führt generell zu einer höheren Akzeptanz.	✗
H6c: Der Einsatz künstlicher Intelligenz in Überwachungsmaßnahmen führt nur dann zu einer größeren öffentlichen Akzeptanz, wenn durch den Einsatz von künstlicher Intelligenz die Privatsphäre der Betroffenen geschützt werden soll.	✗

6. Diskussion und Fazit

Das Hauptziel dieser Studie bestand darin, mithilfe eines faktoriellen Survey-Experiments festzustellen, wie der Kontext sowie die rechtliche und technische Ausgestaltung einer staatlichen Überwachungsmaßnahme die gesellschaftliche Akzeptanz der Maßnahme beeinflussen.

Im Kontext von Wohnungseinbruchsdiebstahl und Gewaltverbrechen ist die Zustimmung zu den vorgeschlagenen Überwachungsmaßnahmen deutlich geringer als im Kontext von Terrorismus. Da die Wahrscheinlichkeit, Opfer eines Einbruchs oder Gewaltverbrechens zu werden, deutlich höher ist als die Wahrscheinlichkeit, Opfer eines Terroranschlags zu werden, lässt

diese Erkenntnis den Schluss zu, dass die Befragten den Einsatz staatlicher Überwachung zur Bekämpfung gesellschaftlicher Bedrohungen für legitimer halten als zur Bekämpfung von Individualkriminalität. Dieses Ergebnis steht im Einklang mit den Ergebnissen von Mondak und Hurwitz (2012), die ebenfalls eine höhere Akzeptanz für Maßnahmen zur Einschränkung bürgerlicher Freiheiten im Zusammenhang mit Terrorismus feststellten.

Die Akzeptanz der Überwachungsmaßnahmen ist im Experiment deutlich höher, wenn die Massenüberwachungstauglichkeit begrenzt ist und der Eingriff in die Privatsphäre der Betroffenen gering ausfällt. Diese Ergebnisse stehen ebenfalls im Einklang mit vorhergegangenen Studien, die eine stärkere Akzeptanz gegenüber Überwachungsmaßnahmen feststellten, die sich gegen bestimmte Personen richten und nicht die gesamte Bevölkerung betreffen (Antoine 2023; Ziller/Helbling 2021). Die Ergebnisse bestätigen vorangegangene Studien, die eine geringere Akzeptanz von Überwachungsmaßnahmen mit einem größeren Potenzial für Eingriffe in die Privatsphäre feststellten (Antoine 2023).

Das Gleiche gilt für die institutionelle Kontrolle in Form richterlicher Anordnungen. Wenn diese Anordnungen für die Durchführung der Überwachungsmaßnahme erforderlich sind, steigt die Akzeptanz der Maßnahme deutlich an. Der Effekt der richterlichen Kontrolle ist der stärkste unter den manipulierten Variablen im Experiment. Dieses Ergebnis deckt sich mit den Forschungsergebnissen von Pavone und Degli Esposti (2012, S. 563), die die Wichtigkeit richterlicher Kontrolle für die Einstellungsbildung gegenüber staatlicher Überwachung in qualitativen Fokusgruppeninterviews in Spanien feststellten. Unsere Ergebnisse zeigen, dass der Effekt richterlicher Kontrolle vom spanischen auf den deutschen Kontext übertragbar ist und deuten deshalb darauf hin, dass die Aufsicht durch Gerichte und Richter eine wichtige Voraussetzung für die Akzeptanz von Überwachungsmaßnahmen in verschiedenen nationalen Kontexten ist. Die qualitativen Befunde der Fokusgruppeninterviews aus der Studie von Pavone und Degli Esposti (2012) konnten wir darüber hinaus in unserer Studie in einem quantitativen, bevölkerungsrepräsentativen Sample replizieren.

In unserem Experiment beschreiben wir eine Vorabkontrolle der Überwachungsmaßnahme durch einen Richter. In zeitkritischen Situationen kann auch eine Abweichung davon zulässig sein (§100e Abs. 1 Satz 2, 3 StPO). Für zukünftige Forschungsvorhaben könnte es von Interesse sein, zu ermitteln, inwiefern es bezüglich der Akzeptanz einen Unterschied macht, ob die Kontrolle vorab oder erst nachträglich erfolgt.

Die Frage, ob kritische Software-Schwachstellen zur Durchführung staatlicher Überwachungsmaßnahmen auf Smartphones genutzt werden sollten, ist ein zentrales Thema der deutschen Überwachungsdebatte (Freiling u. a. 2017; Rehak 2013; Vettermann u. a. 2023).

Überraschenderweise hat die Nennung der Nutzung von Schwachstellen im Experiment keine signifikante Wirkung im Vergleich zur Verpflichtung der Smartphone-Herstellenden zur Zusammenarbeit. Dafür gibt es mehrere mögliche Erklärungen. Die Debatte betrifft möglicherweise nur Expert:innen und findet in der Öffentlichkeit keinen Widerhall. Da das Problem der Software-Schwachstellen sehr technisch ist, ist es zudem möglich, dass einige Befragte den Inhalt dieser Dimension nicht verstanden haben. Eine alternative Erklärung ist, dass die Verpflichtung der Herstellenden zur Kooperation für die Befragten keine signifikant bessere Alternative darstellt, da auch dieses Vorgehen eine Reihe ungeklärter Probleme, wie die regelmäßige Involvierung einer Vielzahl von Herstellenden aus unterschiedlichen nationalen Kontexten zur Folge hätte (vgl. Klauser u. a. 2026). Vorangegangene Studien zeigen zudem, dass die Skepsis gegenüber privatwirtschaftlichen Akteur:innen im Kontext der Überwachung deutlich höher ist als gegenüber staatlichen Akteur:innen (Van den Broek u. a. 2017, S. 21f.). Dieser Umstand könnte ebenfalls dazu beitragen, dass die Befragten die Involvierung privater Herstellenden nicht stärker unterstützen als das Ausnutzen von Schwachstellen.

Trotz der gesellschaftlichen Debatte über den Einsatz künstlicher Intelligenz und die damit einhergehenden Gefahren (Ittefaq u. a. 2025) konnten wir in unserem Experiment keinen signifikanten Effekt durch die Nutzung von künstlicher Intelligenz feststellen. Dies gilt auch, wenn der Einsatz der künstlichen Intelligenz explizit zur Verbesserung des Schutzes der Privatsphäre verwendet werden soll. Das Argument, dass Befragte Überwachungsmaßnahmen positiver gegenüberstehen, wenn diese weitestgehend automatisiert und ohne menschliches Zutun ablaufen (Cayford/Pieters 2018, S. 97), konnte somit im Experiment nicht gestützt werden.

Insgesamt zeigen die Ergebnisse des Experiments, dass es sinnvoll ist, die gesellschaftlichen Einstellungen zu staatlichen Überwachungsmaßnahmen nicht abstrakt zu erfragen, sondern die konkrete Ausgestaltung der Maßnahmen mit zu berücksichtigen.

Grundsätzlich lässt sich feststellen, dass neben der Begrenzung der Maßnahme auf schwerste Straftaten wie Terrorismus vor allem die Dimensionen, die mit der Beschränkung und Kontrolle der Maßnahme zusammenhängen, einen Einfluss auf die Akzeptanz hatten. Dazu gehören die

Beschränkung der Massenüberwachungstauglichkeit und des Eingriffs in die Privatsphäre sowie die richterliche Kontrolle. Gleichzeitig haben die Dimensionen, die eher mit der technischen Umsetzung der Maßnahme zusammenhängen, wie die Nutzung von Schwachstellen für den Zugang zum Gerät und der Einsatz von künstlicher Intelligenz, keinen signifikanten Einfluss auf die Zustimmung zur Maßnahme. Für die Politikgestaltung lässt sich daraus ableiten, dass die Sicherstellung der Beschränkung und Kontrolle der Maßnahme für die gesellschaftliche Akzeptanz wichtiger ist als die technischen Details der Ausgestaltung.

Je nach gewähltem Ansatz weisen empirische Studien Stärken, aber auch Limitationen auf, die hinsichtlich der Interpretation und Generalisierbarkeit der Ergebnisse berücksichtigt werden müssen. Ein Vorteil des in dieser Studie gewählten Ansatzes ist, dass die detaillierte Beschreibung der technischen und rechtlichen Aspekte der vorgestellten Überwachungsmaßnahme eher dem alltäglichen politischen Diskurs entspricht, in dem bestimmte Maßnahmen öffentlich diskutiert werden.

Gleichzeitig besteht die Gefahr, dass genau dieser Detailreichtum bei den beschriebenen Maßnahmen für einige Befragte überwältigend sein könnte. Dies gilt insbesondere, wenn die Befragten ihre Einstellungen zur staatlichen Überwachung auf einer allgemeineren Ebene bilden. Daher sind weitere Untersuchungen erforderlich, um festzustellen, wie Überwachungsszenarien auf verschiedenen Abstraktionsebenen dargestellt werden sollten.

In dieser Studie wurden Einstellungen zur staatlichen Überwachung am Beispiel der Überwachung privater Kommunikationsgeräte untersucht. Die Frage, inwiefern sich die Ergebnisse auf andere Formen staatlicher Überwachung in anderen Kontexten, wie beispielsweise Videoüberwachung im öffentlichen Raum oder die Überwachung des privaten Surfverhaltens im Internet, übertragen lassen, bleibt offen. Hierzu ist weitere Forschung notwendig, die die Einstellungen gegenüber staatlichen Überwachungsmaßnahmen in Abhängigkeit von deren konkreter rechtlicher und technischer Ausgestaltung untersucht.

Danksagung

Dieser Beitrag entstand im Rahmen des SEpES-Projekts am bayerischen Forschungsinstitut für Digitale Transformation (bidt). Wir bedanken uns bei den Teilnehmenden der Jahreskonferenz der Plattform Privatheit 2025 für die wertvollen Anmerkungen. Für die Unterstützung bei der Formulierung und der Recherche bedanken wir uns bei unserer studentischen Hilfskraft Maj Britt Busch.

Literatur

- Antoine, Lukas (2023): Costs, Inconvenience, or Civil Rights? Investigating Determinants of Public Support for Surveillance. *Surveillance & Society*, 21(4), S. 409–431. <https://doi.org/10.24908/ss.v21i4.15796>
- Auspurg, Katrin und Hinz, Thomas (2015): *Factorial Survey Experiments*. Thousand Oaks: SAGE Publications. <https://doi.org/10.4135/9781483398075>
- Bug, Mathias und Bukow, Sebastian U. (2017): Civil Liberties vs. Security: Why Citizens Accept or Reject Digital Security Measures. *German Politics*, 26(2), S. 292–313. <https://doi.org/10.1080/09644008.2016.1226811>
- Cayford, Michelle und Pieters, Wolter (2018): The effectiveness of surveillance technology: What intelligence officials are saying. *The Information Society*, 34(2), S. 88–103. <https://doi.org/10.1080/01972243.2017.1414721>
- Davis, Darren W. und Silver, Brian D. (2004): Civil Liberties vs. Security: Public Opinion in the Context of the Terrorist Attacks on America. *American Journal of Political Science*, 48(1), S. 28–46. <https://doi.org/10.1111/j.0092-5853.2004.00054.x>
- Degli Esposti, Sara; Ball, Kirstie und Dibb, Sally (2021): What's In It For Us? Benevolence, National Security, and Digital Surveillance. *Public Administration Review*, 81(5), S. 862–873. <https://doi.org/10.1111/puar.13362>
- Dourish, Paul und Anderson, Ken (2006): Collective Information Practice: Exploring Privacy and Security as Social and Cultural Phenomena. *Human-Computer Interaction*, 21(3), S. 319–342. https://doi.org/10.1207/s15327051hci2103_2
- Finn, Rachel L.; Wright, David und Friedewald, Michael (2012): Seven types of privacy. In: Gutwirth, Serge; Leenes, Ronald; de Hert, Paul und Poulet, Yves (Hrsg.): *European data protection: Coming of age*. Dordrecht: Springer, S. 3–32.
- Freiling, Felix; Safferling, Christoph und Rückert, Christian (2017): Quellen-TKÜ und Online-Durchsuchung als neue Maßnahmen für die Strafverfolgung: Rechtliche und technische Herausforderungen. *Juristische Rundschau*, 1/2018, S. 9–22. <https://doi.org/10.1515/juru-2017-0104>
- Friedewald, Michael; van Lieshout, Marc; Rung, Sven; Ooms, Merel und Ypma, Jelmer (2014): Privacy and security perceptions of European citizens: A test of the trade-off model. In: Camenisch, Jan; Fischer-Hübner, Simone und Hansen, Marit (Hrsg.): *Privacy and Identity Management for the Future Internet in the Age of Globalisation. Privacy and Identity 2014. IFIP Advances in Information and Communication Technology*, vol. 457. Cham: Springer, S. 39–53.
- Friedewald, Michael; Rung, Sven; van Lieshout, Marc; Ooms, Merel; van der Broek, Tijs und Ypma, Jelmer (2015): Deliverable 10.1: Report on statistical analysis of the PRISMS survey. Brüssel: PRISMS Project (EU FP7, GA 285399). URL: <https://public.a.fraunhofer.de/handle/publica/297523> (besucht am 09.04.2026).
- Garcia, Blake E. und Geva, Nehemia (2016): Security Versus Liberty in the Context of Counterterrorism: An Experimental Approach. *Terrorism and Political Violence*, 28(1), S. 30–48. <https://doi.org/10.1080/09546553.2013.878704>
- Geminn, Christian und Roßnagel, Alexander (2015): „Privatheit“ und „Privatsphäre“ aus der Perspektive des Rechts–ein Überblick. *JuristenZeitung*, 14/2015, S. 703–708.

- Gleizer, Orlandino (2020): Algorithmischer Kernbereichsschutz. Zum Einsatz von künstlicher Intelligenz in der Online-Durchsuchung als Staatspflicht. In: Kusche, Carsten; Beck, Susanne und Valerius, Brian (Hrsg.): *Digitalisierung, Automatisierung, KI Und Recht*. Baden-Baden: Nomos, S. 535–556.
- Greenwald, Glenn (2014): *No place to hide: Edward Snowden, the NSA, and the US surveillance state*. New York: Metropolitan Books.
- Gruber, Jan; Brodowski, Dominik und Freiling, Felix (2022): Die polizeiliche Aufgabe und Pflicht zur digitalen Gefahrenabwehr. *Zeitschrift Für Das Gesamte Sicherheitsrecht*, 4/2022, S. 171–176.
- Hallinan, Dara; Friedewald, Michael und McCarthy, Paul (2012): Citizens' perceptions of data protection and privacy in Europe. *Computer Law & Security Review*, 28(3), S. 263–272. <https://doi.org/10.1016/j.clsr.2012.03.005>
- Heckmann, Dirk (2022): Schutz von Persönlichkeit und Privatsphäre. In: Stern, Klaus; Sodan, Helge und Möstl, Markus (Hrsg.): *Das Staatsrecht der Bundesrepublik Deutschland im europäischen Staatenverbund*. München: C. H. Beck, S. 108–122.
- Ittefaq, Muhammad; Zain, Ali; Arif, Rauf; Ala-Uddin, Mohammad; Ahmad, Taufiq und Iqbal, Azhar (2025): Global news media coverage of artificial intelligence (AI): A comparative analysis of frames, sentiments, and trends across 12 countries. *Telematics and Informatics*, 96, 102223. <https://doi.org/10.1016/j.tele.2024.102223>
- Jäger, Felix (2023): Security vs. civil liberties: How citizens cope with threat, restriction, and ideology. *Frontiers in Political Science*, 4. 1006711. <https://doi.org/10.3389/fpos.2022.1006711>
- Klauser, Joanna; Albert, Bruno; Lindenmeier, Christian; Hammer, Andreas; Freiling, Felix; Heckmann, Dirk und Pfeiffer, Sabine (2026): Telekommunikationsüberwachung am Scheideweg: Zur Regulierbarkeit des Zugriffs auf verschlüsselte Kommunikation. *Zeitschrift für Cyberstrafrecht – CyberStR* 1/2026, S. 17–23.
- Kostka, Genia; Steinacker, Léa und Meckel, Miriam (2023): Under big brother's watchful eye: Cross-country attitudes toward facial recognition technology. *Government Information Quarterly*, 40(1), 101761. <https://doi.org/10.1016/j.giq.2022.101761>
- Kühne, Sylvia (2019): Sanfte Überwachung? Eine Untersuchung zur Akzeptanz von digitalen Fingerabdrucktechnologien im Alltag. Dissertation, Universität Hamburg. URL: <https://ediss.sub.uni-hamburg.de/bitstream/ediss/8356/1/Dissertation.pdf> (besucht am 22.01.2026).
- Lawall, Alexander und Scharping, Davorin (2024): Staatstrojaner: Recht auf Privatsphäre: Analyse der Wirksamkeit der außerparlamentarischen Kontrollmechanismen. *Datenschutz und Datensicherheit - DuD*, 48(9), S. 603–607.
- Li, Robin G. (2024): Institutional trustworthiness on public attitudes toward facial recognition technology: Evidence from U.S. policing. *Government Information Quarterly*, 41(3), 101941. <https://doi.org/10.1016/j.giq.2024.101941>
- Lindenmeier, Christian; Gruber, Jan und Freiling, Felix (2024a): InveSTEE: A TEE-supported Framework for Lawful Remote Forensic Investigations. *Digital Threats: Research and Practice*, 5(3), S. 27:1–27:20. <https://doi.org/10.1145/3680294>

- Lindenmeier, Christian; Hammer, Andreas; Gruber, Jan; Röckl, Jonas und Freiling, Felix (2024b): Key extraction-based lawful access to encrypted data: Taxonomy and survey. *Forensic Science International: Digital Investigation*, 50, 301796. <https://doi.org/10.1016/j.fsidi.2024.301796>
- Lüdemann, Christian und Schlepper, Christina (2011): Der überwachte Bürger zwischen Apathie und Protest. Eine empirische Studie zum Widerstand gegen staatliche Kontrolle. In: Zurawski, Nils (Hrsg.): *Überwachungspraxen. Praktiken Der Überwachung Und Kontrolle*. Leverkusen: Budrich UniPress, S. 119–138.
- Lüdemann, Christian und Schlepper, Christina (2013): Angst im Überwachungsstaat: Eine empirische Studie zur Akzeptanz neuer staatlicher Kontrolltechnologien. In: Gaycken, Sandro (Hrsg.): *Jenseits von 1984 – Datenschutz und Überwachung in der fortgeschrittenen Informationsgesellschaft. Eine Versachlichung*. Bielefeld: transcript Verlag, S. 147–162. <https://doi.org/10.14361/transcript.9783839420034.147>
- Lyon, David (2007): *Surveillance Studies: An Overview*. Cambridge: Polity Press.
- Lyon, David und Haggerty, Kevin D. (2012): The surveillance legacies of 9/11: Recalling, reflecting on, and rethinking surveillance in the security era. *Canadian Journal of Law and Society/La Revue Canadienne Droit et Société*, 27(3), S. 291–300.
- Marcinkowski, Frank; Lünich, Marco; Anicker, Fabian; Flaßhoff, Florian G. und Keller, Birte (2026): Meinungsmonitor Künstliche Intelligenz 3.0. (Interaktives Dashboard). Düsseldorf: Heinrich-Heine-Universität. URL: <https://www.sozwiss.hhu.de/meinungsmmonitor-kuenstliche-intelligenz> (besucht am 25.01.2026)
- Marx, Gary T. (2015): Surveillance studies. *International Encyclopedia of the Social & Behavioral Sciences*, 23(2), S. 733–741.
- Moll, Ricarda und Schneider, Franziska (2021): Ausbau der Datenerhebungsbefugnisse von Sicherheitsbehörden – Lässt die wissenschaftliche Empirie Chilling-Effekte in der Bevölkerung erwarten? *Monatsschrift Für Kriminologie Und Strafrechtsreform*, 104(2), S. 92–106. <https://doi.org/10.1515/mks-2021-0113>
- Möller, Johanna; Schmitz, Lukas und Rehms, Sebastian (2025): Das Privacy Fabric Model: Ein Vorschlag für interdisziplinäre Verständigung in der Privatheitsforschung. In: Friedewald, Michael; Roßnagel, Alexander; Geminn, Christian L. und Karaboga, Murat (Hrsg.): *Freiheit in digitalen Infrastrukturen*. Baden-Baden: Nomos, S. 167–186.
- Mondak, Jeffery J. und Hurwitz, Jon (2012): Examining the Terror Exception: Terrorism and Commitments to Civil Liberties. *Public Opinion Quarterly*, 76(2), S. 193–213. <https://doi.org/10.1093/poq/nfr068>
- Patil, Sunil; Potoglou, Dimitris; Lu, Hui; Robinson, Neil und Burge, Peter (2014): Trade-off Across Privacy, Security and Surveillance in the Case of Metro Travel in Europe. *Transportation Research Procedia*, 1(1), S. 121–132. <https://doi.org/10.1016/j.trpro.2014.07.013>
- Pavone, Vincenzo und Degli Esposti, Sara (2012): Public assessment of new surveillance-oriented security technologies: Beyond the trade-off between privacy and security. *Public Understanding of Science*, 21(5), S. 556–572. <https://doi.org/10.1177/0963662510376886>
- Perkowski, Nina und Epik, Aziz (2024): Alltägliche (Un-) Sicherheit in Hamburg: Erkenntnisse Einer Partizipativen Studie. *Wissen Schafft Demokratie*, 16, S. 142–153.

- Petermann, Thomas und Scherz, Constanze (2005): TA und (Technik-) Akzeptanz (-forschung). *TATuP-Zeitschrift Für Technikfolgenabschätzung in Theorie Und Praxis*, 14(3), S. 45–53.
- Reddick, Christopher G.; Chatfield, Akemi T. und Jaramillo, Patricia A. (2015): Public opinion on National Security Agency surveillance programs: A multi-method approach. *Government Information Quarterly*, 32(2), S. 129–141. <https://doi.org/10.1016/j.giq.2015.01.003>
- Rehak, Rainer (2013): *Angezapft – Technische Möglichkeiten einer heimlichen Online-Durchsuchung und der Versuch ihrer rechtlichen Bändigung*. Münster: MV-Verlag Edition Wissenschaft.
- Trüdinger, Eva-Maria (2018): Sicherheit aus Vertrauen? Der Einfluss politischen und sozialen Vertrauens auf Präferenzen für staatliche Antiterrormaßnahmen. In: Steinbrecher, Markus; Biehl, Heiko; Bytzek, Evelyn und Rosar, Ulrich (Hrsg.): *Freiheit oder Sicherheit? Ein Spannungsverhältnis aus Sicht der Bürgerinnen und Bürger*. Wiesbaden: Springer Fachmedien, S. 77–100. https://doi.org/10.1007/978-3-658-23611-3_4
- Trüdinger, Eva-Maria und Steckermeier, Leonie C. (2017): Trusting and controlling? Political trust, information and acceptance of surveillance policies: The case of Germany. *Government Information Quarterly*, 34(3), S. 421–433. <https://doi.org/10.1016/j.giq.2017.07.003>
- Trüdinger, Eva-Maria und Ziller, Conrad (2022): Considered effective? How policy evaluations and threat perceptions affect support for surveillance in the context of terrorism. *Politics & Policy*, 50(5), S. 894–912. <https://doi.org/10.1111/polp.12498>
- Van den Broek, Tijs; Ooms, Merel; Friedewald, Michael; Van Lieshout, Marc und Rung, Sven (2017): Privacy and security: Citizens' desires for an equal footing. In: Friedewald, Michael; Burgess, Peter; Čas, Johann; Bellanova, Rocco und Walter, Peisl (Hrsg.): *Surveillance, privacy and security*. London: Routledge, S. 15–35.
- Van Heek, Julia; Arning, Katrin und Ziefle, Martina (2017): The Surveillance Society: Which Factors Form Public Acceptance of Surveillance Technologies? In: Helfert, Markus; Klein, Cornel; Donnellan, Brian und Gusikhin, Oleg (Hrsg.): *Smart Cities, Green Technologies, and Intelligent Transport Systems*. Cham: Springer International Publishing, S. 170–191. https://doi.org/10.1007/978-3-319-63712-9_10
- Vettermann, Oliver; Wagner, Manuela; Leicht, Maximilian und Freiling, Felix (2023): Lücken schließen: Der verantwortungsbewusste Umgang mit IT-Sicherheitslücken. München: bidt. <https://doi.org/10.35067/B0BJ-IM05>
- Wright, David; Gutwirth, Serge; Friedewald, Michael; De Hert, Paul; Langheinrich, Marc und Mosciro, Anna (2009): Privacy, trust and policy-making: Challenges and responses. *Computer Law & Security Review*, 25(1), S. 69–83.
- Zdun, Steffen (2006): Telekommunikationsüberwachung. In: Lange, Hans-Jürgen und Gasch, Matthias (Hrsg.): *Wörterbuch zur Inneren Sicherheit*. Wiesbaden: Springer, S.325-329. https://doi.org/10.1007/978-3-531-90596-9_79
- Ziller, Conrad und Helbling, Marc (2021): Public support for state surveillance. *European Journal of Political Research*, 60(4), S. 994–1006. <https://doi.org/10.1111/1475-6765.12424>

Unkontrollierte Nutzung Künstlicher Intelligenz durch Polizeibehörden? Fragmentierte Accountability- und Compliance-Vorkehrungen

Hartmut Aden und Steven Kleemann

Zusammenfassung

Dieser Beitrag betrachtet die *Accountability*- und *Compliance*-Vorkehrungen für die Kontrolle der Nutzung Künstlicher Intelligenz (KI) durch Polizeibehörden, mit denen sichergestellt werden soll, dass auftretende Risiken möglichst frühzeitig erkannt und minimiert werden. Die Kontrollstrukturen der 2024 verabschiedeten KI-Verordnung der Europäischen Union sind grundsätzlich auch auf Strafverfolgungsbehörden anwendbar. Allerdings relativieren zahlreiche „Hintertüren“ in Form von Sonderregelungen und Ausnahmen diesen integrierten Regelungsansatz. Dieser Beitrag zeigt, dass die EU-KI-VO eine Reihe von *Accountability*- und *Compliance*-Strukturen auf EU-Ebene und in den Mitgliedstaaten vorsieht. Im Mittelpunkt stehen prozedurale *Accountability*-Vorkehrungen, u. a. Pflichten zur Einführung eines Risikomanagementsystems und bei staatlichen Stellen zur Durchführung einer Grundrechte-Folgenabschätzung. Allerdings sind auch diese Vorgaben aufgrund zahlreicher Sonderregelungen für die Strafverfolgung nur eingeschränkt anwendbar.

1. Einleitung¹

Die KI-Verordnung der Europäischen Union (englisch *Artificial Intelligence Act*, im Folgenden: EU-KI-VO) wurde 2021 von der Europäischen Kommission vorgeschlagen und 2024 nach kontroversen Diskussionen als Ver-

1 Dieser Beitrag basiert zu Teilen auf Erkenntnissen des Verbundforschungsprojekt *VIKING (Vertrauenswürdige Künstliche Intelligenz für polizeiliche Anwendungen)*, 2022 bis 2025, gefördert vom Bundesministerium für Bildung und Forschung (seit 2025: Bundesministerium für Forschung, Technologie und Raumfahrt), Förder-Kennzeichen 13N16241.

ordnung (EU) 2024/1689 verabschiedet.² Die Vorschriften treten in einem mehrstufigen, gestreckten Zeitrahmen in Kraft (vgl. Art. 113 EU-KI-VO in der durch Verordnung (EU) 2026/1744 geänderten Fassung). Die EU-KI-VO schafft einen allgemeinen Rahmen für die KI-Nutzung und die Bewertung der von ihr ausgehenden Risiken.

Anders als die EU-Datenschutz-Grundverordnung (DSGVO) gilt die EU-KI-VO auch für die strafrechtliche Ermittlung – nach ihrem Art. 3 Abs. 3 nicht aber für Nachrichtendienste, auf die sich die Gesetzgebungszuständigkeit der EU bislang nicht erstreckt. Eingriffsbefugnisse für die tatsächliche Einführung und Nutzung von KI-Systemen durch staatliche Behörden wie die Polizei müssen indes die Mitgliedstaaten im Rahmen ihrer Gesetzgebung etablieren. Vor dem Hintergrund der Intensität möglicher Folgen für die Grundrechte der Betroffenen und der Rechtsprechung des Europäischen Gerichtshofs, des Europäischen Gerichtshofs für Menschenrechte sowie in Deutschland auch des Bundesverfassungsgerichts sind spezialgesetzliche Regelungen erforderlich, soweit die staatliche Nutzung von KI-Systemen zu Grundrechtseingriffen führt. Bund und Länder sind daher bestrebt, mehr spezialgesetzliche Eingriffsbefugnisse für die KI-Nutzung durch Strafverfolgungsbehörden zu schaffen. Gesetzgebungsvorschläge für das Ermöglichen polizeilicher KI-Nutzung sind daher bereits Teil der rechtspolitischen Agenda, u. a. für die Polizeigesetze (vgl. Martini/Botta 2025, S. 1038ff.).

Vor Verabschiedung der EU-KI-VO unterlagen KI-Systeme nur den Anforderungen des Datenschutzrechts und weiteren, nicht KI-spezifischen rechtlichen Regelungen. Auch wenn es sich bei der EU-KI-VO um eine nach Art. 288 AEUV unmittelbar geltende Verordnung handelt, bedürfen die *Accountability*-Anforderungen der EU-KI-VO in vielerlei Hinsicht noch der Konkretisierung, auch durch die Mitgliedstaaten. Der Aufbau der institutionellen Strukturen steht erst am Anfang. Die EU-KI-VO nutzt in mancherlei Hinsicht den Ansatz „regulierter Selbstregulierung“ – die kontrollierten Akteure haben daher erhebliche Spielräume bei der Ausgestaltung der Kontrollstrukturen.

Bereits im Hinblick auf die Privatwirtschaft erscheint dies im Ausgangspunkt problematisch, da die Kontrolle unabhängig und der Einfluss der

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

Kontrollierten auf die Kontrolle begrenzt sein sollte. Im Folgenden wird gezeigt, wie diese Problematik bei Strafverfolgungsbehörden durch ihre weitreichenden Befugnisse zur KI-Nutzung und die zahlreichen, speziell für sie geschaffenen Ausnahmetatbestände oder „Hintertüren“ in der EU-KI-VO verschärft wird. Ausnahmen für die Strafverfolgung gibt es z. B. für die biometrische Fernidentifizierung (Art. 5 Abs. 1 lit. h EU-KI-VO für die eigentlich verbotenen Systeme und Anhang III Nr. 1 lit. a, Art. 26 Abs. 10 EU-KI-VO für Hochrisiko-Anwendungen), für Tests unter Realbedingungen (Art. 60 Abs. 4 lit. i EU-KI-VO) oder durch Verweise auf die neu eingeführte und dem Datenschutzrecht unbekannte Kategorie „sensible operative Daten“ (Art. 3 Nr. 38 EU-KI-VO), deren Kontrolle Einschränkungen und Sonderregelungen unterliegt (vgl. Kleemann/Aden 2025). Hinzu kommt die unübersichtliche Vielzahl von *Compliance*- und *Accountability*-Vorkehrungen der EU-KI-VO, die im Endeffekt zu einer „*Accountability*-Falle“ werden können, wenn sie unkoordiniert und intransparent bleiben – und somit ein ganzheitlicher *Accountability*-Ansatz fehlt. Daher erscheint es fraglich, ob die Regelungen der EU-KI-VO ausreichen, um die polizeiliche KI-Nutzung im Hinblick auf die Grundrechte Betroffener wirksam zu kontrollieren und zu begrenzen. Die mitgliedstaatliche Gesetzgebung könnte zwar trotz der unzulänglichen Regelungen der EU-KI-VO diese Aufgabe übernehmen. Angesichts des Trends zu sicherheitsorientierter Gesetzgebung (vgl. Aden/Fährmann 2021) ist dies allerdings kaum zu erwarten.

Dieser Beitrag untersucht Kriterien für die Beurteilung der Effektivität von *Accountability*- und *Compliance*-Vorkehrungen für die KI-Nutzung im Allgemeinen sowie deren Möglichkeiten und Grenzen im Kontext der polizeilichen KI-Nutzung auf der Basis der EU-KI-VO. Zudem nimmt er die (begrenzten) Einflussmöglichkeiten etablierter *Accountability*-Foren (Europäischer Datenschutzbeauftragter, Europäischer Datenschutzausschuss, mitgliedstaatliche Datenschutzaufsicht, mitgliedstaatliche Menschenrechtseinstitutionen, Europäische Bürgerbeauftragte usw.) auf die Kontrolle der KI-Nutzung durch Strafverfolgungsbehörden in den Blick. Der Beitrag verfolgt das Ziel, die *Accountability*- und *Compliance*-Vorkehrungen für die KI-Entwicklung und -Nutzung aus einer ganzheitlichen Perspektive zu untersuchen, die technische Eigenschaften, den rechtlichen Rahmen und den sozio-technischen Kontext von KI sowie deren Wechselwirkungen einbezieht. Zur exemplarischen Illustration werden die KI-gestützte

Objektdetektion und die Sprechendenerkennung verwendet.³ Während bei der Sprechendenerkennung der Eingriff in die Telekommunikationsfreiheit im Fokus steht, ist die Objektdetektion ein relevantes Beispiel für die polizeiliche Auswertung großer Datenmengen. Der Einsatz von KI-Systemen zur automatisierten Detektion von Objekten in großen Video- und Bilddatenbeständen (z. B. im Rahmen einer Rasterfahndung oder der Auswertung sichergestellter Speichermedien) und zur KI-gestützten Sprechendenerkennung bei der Telekommunikationsüberwachung erfordert eigenständige spezialgesetzliche Grundlagen, da sie unmittelbar in Grundrechte – insbesondere das Recht auf informationelle Selbstbestimmung und den Schutz des Fernmeldegeheimnisses – eingreifen.

2. Die Relevanz von Accountability- und Compliance-Vorkehrungen für polizeiliche KI-Anwendungen

Das Ziel, umfängliche KI-*Accountability* zu gewährleisten, umfasst die Komponenten *Compliance*, Berichterstattung, unabhängige Aufsicht und Durchsetzung (Novelli u. a. 2024). Eine effektive Kontrolle der polizeilichen KI-Nutzung ist vor dem Hintergrund potenziell weitreichender Auswirkungen auf die Grundrechte Betroffener von hoher Relevanz. Die EU-KI-VO etabliert hierfür eine Reihe von Instrumenten, die als *Accountability*-Foren fungieren. Ein *Accountability-Forum* verbindet nach Bovens (2007, S. 450) einen Akteur, der für ein Handeln Verantwortung trägt, und eine Kontrollinstanz, z. B. ein Aufsichtsgremium. Der rechenschaftspflichtige Akteur, hier die Polizei, ist verpflichtet, Verhalten und Handlungen gegenüber dem Forum zu rechtfertigen. Das Forum kann nachfragen, das Verhalten beurteilen und in einigen Fällen auch Sanktionen verhängen. In Rechtsstaaten sind staatliche Behörden wie die Polizei typischerweise gegenüber mehreren *Accountability*-Foren rechenschaftspflichtig (Bovens u. a. 2014), insbesondere gegenüber vorgesetzten Behörden, speziellen Aufsichtsbehörden wie den Datenschutzbeauftragten und den Rechnungshöfen, den Parlamenten, den Gerichten und einer breiteren Öffentlichkeit. *Accountability* nimmt somit eine umfassendere Perspektive ein als *Compliance*, die in erster Linie Regelkonformität in den Blick nimmt. Die *Compliance* kann zumeist mit Checklisten überprüft werden, wohingegen *Accountability* die Kontrollstrukturen und ihren Kontext betrachtet.

3 Anknüpfend an das Forschungsprojekt VIKING, näher hierzu Fischer u. a. 2025.

In der EU-KI-VO werden beide Konzepte unterschiedlich aufgegriffen. Je nach Risikokategorie, in die ein KI-System einzuordnen ist, müssen unterschiedliche *Compliance*- und Informationspflichten erfüllt werden. Die *Accountability*-Ebene geht darüber hinaus und enthält beispielsweise Anforderungen an ein Qualitätsmanagementsystem (Art. 17 EU-KI-VO), menschliche Aufsicht (Art. 14 EU-KI-VO), technische Dokumentations- und Protokollierungspflichten (z. B. Art. 11, 12, 18, 19 EU-KI-VO), aber auch erst noch im Detail auszugestaltende Instrumente wie die Grundrechte-Folgenabschätzung (Art. 27 EU-KI-VO) oder die Etablierung der verschiedenen Aufsichtsgremien auf europäischer und nationaler Ebene (z. B. Art. 64, 65, 70 EU-KI-VO).

Künstliche Intelligenz wird von vielen Autor:innen als *Black Box* eingestuft, weil die genutzten KI-Systeme und ihre Ergebnisse nur begrenzt transparent und erklärbar sind. Bei der polizeilichen KI-Nutzung kommt ein institutioneller *Black Box*-Effekt hinzu, da die organisatorischen Prozesse und die konkrete Nutzung von KI-Systemen aufgrund von Geheimhaltungsbedürfnissen in Sicherheitsbehörden tendenziell in weiten Teilen intransparent sind (vgl. Fährmann u. a. 2022; Kleemann/Aden 2026; Martinez u. a. 2026).

Dieser doppelte *Black Box*-Effekt verstärkt Risiken für die Grundrechte derjenigen, deren Daten in polizeilichen KI-Systemen verarbeitet werden. Fehler, die polizeiliche KI-Systeme erzeugen, können unmittelbare und im Vergleich zu privaten Nutzungen ungleich intensivere Auswirkungen auf die Grundrechte der Betroffenen haben – von diskriminierenden Entscheidungen bis zu irrtümlichen Festnahmen. Da Polizeibehörden zumeist KI-Systeme nutzen, die von externen Firmen entwickelt wurden, sind auch diese externen KI-Anbietenden in die Betrachtung einzubeziehen.

Bei der Objektdetektion und der Sprechendenerkennung in der Telekommunikationsüberwachung führt dieser doppelte *Black Box*-Effekt dazu, dass in der Regel weder die Betroffenen noch externe Kontrollinstanzen wissen, welche Algorithmen die erkannten Objekte oder Stimmen bestimmten Personen zuordnen. Auch das Ausmaß der Fehleranfälligkeit des Systems und die Integration von Treffern in polizeiliche Entscheidungsprozesse – etwa für eine Durchsuchung oder Festnahme – sind für Betroffene und Kontrollinstanzen kaum nachvollziehbar.

Wie lässt sich die Effektivität von *Accountability*-Vorkehrungen für die KI-Nutzung allgemein und speziell für die polizeiliche KI-Nutzung bewerten? Von Relevanz sind hier insbesondere die institutionelle Unabhängigkeit des *Accountability*-Forums gegenüber der Polizeibehörde, die KI-An-

wendungen (weiter-)entwickelt und nutzt, sowie die Reichweite der Kontrollbefugnisse. Die Unabhängigkeit von *Accountability*-Foren gegenüber den Akteur:innen, die KI-Anwendungen entwickeln oder nutzen, reduziert das Risiko, dass kontrollierende Akteure ihre eigenen Interessen verfolgen. Dies gilt in besonderem Maße für KI im polizeilichen Kontext, in dem die Rollen der KI-Entwickelnden und -Betreibenden mit der Macht des staatlichen Gewaltmonopols zusammenfallen können. Ein sich selbst regulierender Markt, der sich bei Versagen der Technologie in Teilen selbst regulieren könnte, fehlt hier, da der Staat der „Markt“ selbst ist und Sicherheitsbehörden dazu noch Geheimhaltungsinteressen haben. Unabhängige *Accountability*-Foren können dies durch ihre Kontrolltätigkeit jedenfalls ein Stück weit kompensieren.

3. Normativ-inhaltliche Anknüpfungspunkte der KI-Accountability und -Compliance

Vor dem Inkrafttreten der EU-KI-VO stammten die normativen Anforderungen an KI-Accountability vorwiegend aus der Ethik-Diskussion zu den Grenzen dessen, was mit KI-Unterstützung möglich sein sollte. Die Zielsetzungen der EU-KI-VO sind ambitioniert und zugleich so breit angelegt, dass Zielkonflikte unvermeidbar sind. Die EU-KI-VO soll dazu beitragen,

„das Funktionieren des Binnenmarkts zu verbessern und die Einführung einer auf den Menschen ausgerichteten und vertrauenswürdigen künstlichen Intelligenz (KI) zu fördern und gleichzeitig ein hohes Schutzniveau in Bezug auf Gesundheit, Sicherheit und die in der Charta verankerten Grundrechte, einschließlich Demokratie, Rechtsstaatlichkeit und Umweltschutz, vor schädlichen Auswirkungen von KI-Systemen in der Union zu gewährleisten und die Innovation zu unterstützen“ (Art. 1 Abs. 1).

Eine zentrale normative Frage war und ist eng mit *Accountability*-Fragen verknüpft: Wie kann KI so ausgestaltet werden, dass sie sich gegenüber menschlichen Entscheidungen nicht verselbständigt? Auch für die polizeiliche KI-Nutzung, deren Ergebnisse zu weitreichenden Grundrechtseingriffen führen können, ist dieses Thema von höchster Relevanz. Die Nutzung komplexer, nicht ohne Weiteres nachvollziehbarer KI-Systeme kann die Handlungsautonomie und den Ermessensgebrauch der nutzenden Polizeibeamt:innen faktisch einschränken, wenn sie sich auf die Ergebnisse des

Systems verlassen (müssen), was wiederum Auswirkungen auf die Grundrechte Betroffener haben kann. In der EU-KI-VO spiegelt sich dieses Anliegen insbesondere in den Regelungen zur menschlichen Aufsicht (Art. 14) sowie zur Genauigkeit, Robustheit und Cybersicherheit (Art. 15) wider, die für Hochrisiko-KI-Systeme und damit auch für weite Teile der polizeilichen KI-Nutzung gelten (vgl. Kleemann/Aden 2025). Was z. B. menschliche Aufsicht, Genauigkeit, Robustheit und Cybersicherheit bei einer KI-gestützten Objektdetektion über tausende Video-Frames bedeutet, ob es als menschliche Aufsicht ausreicht, dass Beamt:innen die vom System markierten Treffer nachträglich sichten oder ob bei der KI-Sprechendenerkennung die Einordnung eines Datensatzes durch Analyst:innen als hinreichende menschliche Kontrolle gilt, bleibt in der EU-KI-VO offen und wird weiter zu konkretisieren sein, u. a. durch Rechtsprechung. Die Kommission soll zudem die „Entwicklung von Benchmarks und Messmethoden“ für Genauigkeit, Robustheit und Cybersicherheit von KI-Systemen fördern (Art. 15 Abs. 2).

Die Pflicht zur Gewährleistung menschlicher Aufsicht „auch mit geeigneten Instrumenten einer Mensch-Maschine-Schnittstelle“ (Art. 14 Abs. 1 EU-KI-VO) schlägt zugleich den Bogen zu zentralen normativen Elementen des Schutzzwecks: „Risiken für Gesundheit, Sicherheit und Grundrechte“ sollen minimiert werden (Art. 14 Abs. 2). Allerdings bleiben die normativen Anforderungen der EU-KI-VO an den Schutz der Grundrechte allgemein. Die Grundrechtsanforderungen für die Minimierung von KI-Risiken werden ebenso wenig konkretisiert wie die Auflösung möglicher Zielkonflikte zwischen sicherer und grundrechtsschützender KI-Nutzung. Die Konkretisierung der Anforderungen und die Bewertung der Risiken werden damit weitgehend den Anbietenden und Betreibenden von KI-Systemen sowie den *Accountability*-Foren überantwortet. Weitere einzelfallübergreifende Konkretisierungen werden erforderlich sein, auch im Polizei- und Datenschutzrecht, um sicherzustellen, dass Risiken frühzeitig erkannt und wirksam bewältigt werden.

4. Institutionalisierte Aufsichtsstrukturen nach der EU-KI-Verordnung

Die EU-KI-VO sieht institutionalisierte Aufsichtsstrukturen sowohl auf EU-Ebene (4.1) als auch in den Mitgliedstaaten vor (4.2), die mit Einschränkungen auch für die polizeiliche KI-Nutzung zuständig sind bzw. sein werden. Allerdings handelt es sich nicht um *Accountability*-Foren mit um-

fassenden Kontrollaufgaben. Allenfalls die Marktüberwachungsbehörden, die auf mitgliedstaatlicher Ebene für Hochrisiko-KI-Systeme einzurichten sind (Art. 70 EU-KI-VO), kommen im Hinblick auf ihre institutionelle Positionierung, ihre Aufgaben und Befugnisse einem unabhängigen *Accountability*-Forum nahe. Die Relevanz dieser Kontrollstrukturen für polizeiliche KI-Systeme wurde indes im Laufe des Gesetzgebungsverfahrens signifikant dadurch reduziert, dass weniger KI-Anwendungsfelder der Hochrisiko-Kategorie zugeordnet wurden als durch den Verordnungsentwurf der Europäischen Kommission zunächst beabsichtigt (vgl. Kleemann/Aden 2025, S. 20f.).

4.1 Institutionelle Strukturen auf EU-Ebene

Die EU-KI-VO etabliert zentrale Governance-Strukturen für KI-Fragen auf EU-Ebene. Wichtige Funktionen übernimmt dabei das *Büro für Künstliche Intelligenz* (englisch *Artificial Intelligence Office*), das gemäß Art. 64 EU-KI-VO der Europäischen Kommission zugeordnet ist, die damit zentrale Aufgaben bei der KI-Entwicklung und -Nutzung wahrnimmt. Bereits vor Inkrafttreten der EU-KI-VO richtete die Kommission das Büro per Beschluss ein (C(2024) 390 final vom 24.1.2024). Die EU-KI-VO und der Einrichtungsbeschluss weisen dem Büro umfangreiche Aufgaben bei der Umsetzung der EU-KI-VO zu, u. a. bei der Durchführung von Überwachungsmaßnahmen (Art. 89). Die Kommission kann gemäß Art. 101 EU-KI-VO Geldbußen für Verstöße gegen die Verordnung durch Anbietende von KI-Modellen mit allgemeinem Verwendungszweck (*General Purpose AI*) verhängen. Das KI-Büro ist in diesen Prozess eingebunden und wurde daher bereits als *AI Police* bezeichnet (Krahn 2024, S. 122). Aufgrund der Höhe der möglichen Bußgelder – bis zu 3 % des Jahresumsatzes bzw. 15 Mio. € (Art. 99 Abs. 3 EU-KI-VO) – kann die Kompetenz des KI-Büros jedenfalls in Bezug auf Sanktionsmöglichkeiten als „scharfes Schwert“ angesehen werden. Somit übernimmt das KI-Büro gewisse *Accountability*-Aufgaben; ihm mangelt es aber aufgrund der Einbindung in die Verwaltungsstrukturen der Kommission an institutioneller Unabhängigkeit.

Im KI-Gremium (englisch *Artificial Intelligence Board*) ist – entsprechend der üblichen Logik des EU-Mehrebenensystems (vgl. Novelli u. a. 2025) – jeder Mitgliedstaat vertreten; der Europäische Datenschutzbeauftragte nimmt nur als Beobachter teil (Art. 65 Abs. 2 und 3 EU-KI-VO). Das Gremium „berät und unterstützt die Kommission und die Mitglied-

staaten, um die einheitliche und wirksame Anwendung dieser Verordnung zu erleichtern“ (Art. 66 EU-KI-VO). Eigene Kontrollbefugnisse hat das KI-Gremium allerdings nicht. Allenfalls im Rahmen von Empfehlungen und Stellungnahmen kann es zum Schutz der von der KI-Nutzung Betroffenen beitragen (so insbesondere gemäß Art. 66 lit. e, n und o EU-KI-VO). Vor dem Hintergrund, dass die mitgliedstaatlichen Regierungen und deren Interessen in dem Gremium repräsentiert sind, dürften hier kaum substanzielle Vorgaben für die polizeiliche KI-Nutzung zu erwarten sein.

Noch weniger Kontrollbefugnisse hat das Beratungsforum, das gemäß Art. 67 EU-KI-VO die Europäische Kommission berät und bei dessen Zusammensetzung „auf ein ausgewogenes Verhältnis zwischen wirtschaftlichen und nicht-wirtschaftlichen Interessen“ geachtet werden soll. Die Mitglieder sollen „Industrie, Start-up-Unternehmen, KMU, die Zivilgesellschaft und die Wissenschaft“ vertreten (Abs. 2). Die Kommission übt u. a. durch ihre Befugnis, die Mitglieder auszuwählen, erheblichen Einfluss auf das Forum aus (Abs. 3). Deutlich wurde dies auch im Rahmen der Besetzung, bei der große Tech-Firmen, wirtschaftliche Akteure und deren Interessenvertretungen überproportional vertreten sind, wohingegen zivilgesellschaftliche Akteure wenig und nationale Menschenrechtsinstitutionen gar nicht ausgewählt wurden (ECNL 2026).

Das wissenschaftliche Gremium unabhängiger Sachverständiger, das gemäß Art. 68 EU-KI-VO eingerichtet wird, unterstützt auch „Durchsetzungstätigkeiten“ (Abs. 1). Seine konkreten Aufgaben sind beratend und unterstützend (Abs. 3), so dass es sich ebenfalls nicht um ein voll ausgeprägtes *Accountability*-Forum mit effektiven Kontrollrechten handelt. Immerhin kann das Gremium aber Warnungen vor systemischen Risiken aussprechen, die von KI-Modellen mit allgemeinem Verwendungszweck ausgehen (Art. 90 EU-KI-VO).

Im Ergebnis wurde keine der nach der EU-KI-VO neu eingerichteten institutionellen Strukturen auf EU-Ebene als unabhängiges *Accountability*-Forum mit umfassenden Kontrollfunktionen ausgestattet. Für die verantwortliche KI-Nutzung im Allgemeinen und die speziellen Grundrechtsrisiken der polizeilichen KI-Nutzung bleiben die bestehenden *Accountability*-Foren daher von großer Bedeutung, insbesondere der Europäische Datenschutzbeauftragte, die EU-Bürgerbeauftragte, die EU-Grundrechteagentur und der Europäische Gerichtshof (EuGH) (s. auch Abschnitt 6).

4.2. Institutionelle Strukturen in den Mitgliedstaaten

Die EU-KI-VO schreibt die Einrichtung weiterer institutioneller Strukturen für die administrative Implementation der KI-Verordnung in den Mitgliedstaaten vor. Jeder Mitgliedstaat muss mindestens eine Marktüberwachungsbehörde und eine notifizierende Behörde einrichten, die ihre Befugnisse unabhängig ausüben sollen (Art. 70 Abs. 1 EU-KI-VO). Die konkretere Ausgestaltung ist den Mitgliedstaaten überlassen, insbesondere auch die Frage, ob diese Aufgaben bestehenden Aufsichtsbehörden übertragen werden und ggf. welchen. Für Hochrisiko-KI-Systeme, die für Strafverfolgung, Grenzmanagement sowie „Justiz und Demokratie“ eingesetzt werden, müssen die Mitgliedstaaten indes die nach der DSGVO oder der Richtlinie (EU) 2016/680 (JI-Richtlinie) bereits etablierten Datenschutzaufsichtsbehörden als Marktüberwachungsbehörden benennen oder vergleichbar unabhängige Aufsichtsbehörden für die Marktüberwachung schaffen. Marktüberwachungstätigkeiten dürfen hier „in keiner Weise die Unabhängigkeit von Justizbehörden beeinträchtigen [...]“ (Art. 74 Abs. 8 EU-KI-VO).

Die Marktüberwachungsbehörden nehmen einige typische *Accountability*-Aufgaben wahr. So üben sie Überwachungsaufgaben für Hochrisiko-KI-Systeme aus (Art. 74 EU-KI-VO) und bearbeiten Meldungen schwerwiegender Vorfälle bei der Nutzung solcher Systeme (Art. 73 EU-KI-VO). Sie haben im Rahmen ihrer Aufsichtstätigkeit weitreichende Rechte auf Zugang zu Informationen über Hochrisiko-Systeme (Art. 73 Abs. 12 und 13 EU-KI-VO).

Die Mitgliedstaaten haben zudem notifizierende Behörden als Teil eines Konformitätsbewertungssystems einzurichten. Hierzu zählt mindestens eine notifizierende Behörde pro Mitgliedstaat, die die Konformitätsbewertungsstellen (Art. 31 EU-KI-VO) notifiziert und überwacht. Mit den Anforderungen an „Objektivität und Unparteilichkeit“ (Art. 28 Abs. 3 EU-KI-VO) weisen diese Behörden ebenfalls Merkmale von *Accountability-Foren* auf. Die von ihnen überwachten Konformitätsbewertungsstellen bewerten die Konformität von Hochrisiko-KI-Systemen mit der EU-KI-VO (Art. 43). Sie sind ein weiterer Baustein in der *Accountability*-Kaskade der EU-KI-VO. Gegenüber den Anbietenden des jeweiligen Hochrisiko-KI-Systems und „allen anderen Akteuren, die ein wirtschaftliches Interesse an den zu bewertenden Hochrisiko-KI-Systemen haben“, müssen sie unabhängig sein. Bei Hochrisiko-Systemen für Strafverfolgungsbehörden gilt erneut eine abweichende Zuständigkeit, nämlich der Aufsichtsbehörden, die ge-

mäß Art. 74 Abs. 8 EU-KI-VO die Aufgaben der Marktüberwachung mit übernehmen (Art. 43 Abs. 1 Satz 3 EU-KI-VO).

Zudem ermöglicht Art. 77 EU-KI-VO, andere mitgliedstaatliche „Behörden oder öffentliche Stellen“ einzubinden, „die die Einhaltung des Unionsrechts zum Schutz der Grundrechte einschließlich des Rechts auf Nichtdiskriminierung [...] beaufsichtigen oder durchsetzen“ (Abs. 1 Satz 1), die sogenannten *Artikel 77-Stellen*. Allerdings beschränken sich die formalen Befugnisse dieser Stellen auf den Zugang zu Dokumentationen, die über Hochrisiko-KI-Systeme zu führen sind (Abs. 1). Mit der durch Verordnung (EU) 2026/1744 geänderten Fassung wurden diese Stellen dahingehend geschwächt, dass die Informationsanforderung nur noch über die Marktüberwachungsbehörden möglich ist und nicht mehr direkt bei Anbietenden und Betreibenden. Sind diese Informationen nicht ausreichend, können *Artikel 77-Stellen* bei der Marktüberwachungsbehörde die Durchführung eines technischen Tests des Hochrisiko-Systems beantragen (Abs. 3).

Diese im Ausgangspunkt schwachen *Accountability*-Befugnisse werden indes ergänzt durch Kontrollrechte der jeweiligen Stelle aufgrund anderer Rechtsvorschriften, z. B. der Datenschutz- und Antidiskriminierungs-Gesetzgebung. Für Deutschland benannte das Bundeswirtschaftsministerium die Datenschutzbeauftragten des Bundes und der Länder, die Antidiskriminierungsstelle des Bundes sowie die Ombudsstelle für Gleichbehandlung und gegen Diskriminierung nach dem Landes-Antidiskriminierungsgesetz Berlin.⁴ Nicht benannt wurden bisher die unabhängigen Polizeibeauftragten, die seit 2014 in einer Reihe von Bundesländern und 2024 auch auf Bundesebene etabliert worden sind (hierzu Töpfer u. a. 2023; Aden 2025), sowie das Deutsche Institut für Menschenrechte als unabhängige nationale Menschenrechtsinstitution Deutschlands.

Im Ergebnis schreibt die EU-KI-VO somit fragmentierte *Accountability*-Strukturen vor, deren Zuständigkeit sich im Wesentlichen auf Hochrisiko-KI-Systeme beschränkt. Polizeiliche KI-Systeme fallen in die Zuständigkeit dieser Strukturen, allerdings mit einigen abweichenden Regelungen für die Zuständigkeit und die Reichweite der Kontrollbefugnisse (vgl. Kleemann/Aden 2025).

4 Liste nach Art. 77 Abs. 2 EU-KI-VO (Stand 19. Dezember 2024). https://www.bundeswirtschaftsministerium.de/Redaktion/DE/Downloads/J-L/liste-nach-art-77-abs-2-ki-verordnung.pdf?__blob=publicationFile&v=8

5. Prozedurale Accountability- und Compliance-Vorkehrungen der EU-KI-Verordnung

Für Hochrisiko-KI-Systeme etabliert die EU-KI-VO eine Reihe von prozeduralen *Accountability*- und *Compliance*-Vorkehrungen. Spezielle Verfahren sollen sicherstellen, dass Risiken für die Grundrechte, die Sicherheit und die Gesundheit frühzeitig erkannt und minimiert werden. Die inhaltlichen Anforderungen werden dagegen in der EU-KI-VO nur mit allgemeinen Postulaten konkretisiert (s. Abschnitt 3). Den Anbietenden und Betreibenden von KI-Systemen verbleiben damit weitreichende Spielräume, die allerdings durch die Aufsichtsstrukturen (s. Abschnitt 4) und die Rechtsprechung konkretisiert werden dürften. Für Polizeibehörden gelten indes auch hier Ausnahmen.

Polizeibehörden, die KI-Software privater Firmen einkaufen und in eigenen bzw. öffentlichen Strukturen (wie etwa HessenDATA oder VeRA) mit speziell zugeschnittenen Funktionen weiterentwickeln und verwenden, könnten gem. Art. 25 Abs. 1 lit. a und b EU-KI-VO Quasi-Anbietende sein und müssen dann sowohl die Anforderungen an Betreibende als auch die an Anbietende einhalten (vgl. Martini/Botta 2025, S. 1037, a. A. Bosman 2025, Art. 25, Rn. 37ff.).

5.1 Risikomanagement

Die Etablierung von Risikomanagementsystemen gehört heute zum Standardinstrumentarium des EU-Rechts, wobei ausgehend vom Umweltrecht weitreichende rechtliche Instrumente zur Durchsetzung des Vorsorgeprinzips (*precautionary principle*) für den Umgang mit Risiken entwickelt wurden (Guerra 2019). Nach der EU-KI-VO haben die Anbietenden von Hochrisiko-KI-Systemen ein solches System einzurichten, das als „kontinuierlicher iterativer Prozess [...] während des gesamten Lebenszyklus“ des Systems angelegt sein soll (Art. 9 Abs. 2 EU-KI-VO). Wer in die Anbietenden-Kategorie fällt, ist in Art. 25 EU-KI-VO geregelt.

Normativer Anknüpfungspunkt für die Risikoermittlung sind hier nach Art. 9 Abs. 2 lit. a wiederum „Gesundheit, Sicherheit oder Grundrechte“ ohne nähere Spezifizierung. Trotz des zentralen Stellenwerts der Risikokategorisierung in der EU-KI-VO bleiben hier also weite Interpretationsspielräume, die der Konkretisierung bedürfen (Schuett 2024). Dies liegt zunächst in der Verantwortung der Anbietenden und Betreibenden und

kann durch die zuständigen Gremien, Behörden und Gerichte überprüft werden. Nicht vorgeschrieben ist, wer für die Durchführung zuständig ist, so dass Anbietende das Risikomanagementsystem „mit Bordmitteln“ betreiben könnten, was zu Interessenkonflikten führen kann. Gerade hierfür bedarf es folglich einer effektiven Aufsicht. Dies gilt auch und besonders für Anbietende von KI-Systemen für polizeiliche Nutzungszwecke.

So könnte etwa ein privates Unternehmen, das ein KI-System zur Objektdetektion oder Sprechendenerkennung für Polizeibehörden entwickelt und betreibt, das Risikomanagementsystem eigenverantwortlich gestalten. Die Folge könnte sein, dass Informationen über systemische Fehler (z. B. erhöhte Falsch-Positiv-Raten bei bestimmten Bevölkerungsgruppen) intern bleiben und von keiner unabhängigen Stelle für den Hochrisikoeinsatz ausreichend überprüft werden, bevor das System durch die Polizei genutzt wird.

5.2 Grundrechte-Folgenabschätzung

Für Einrichtungen, die öffentliche Dienste erbringen, schreibt die EU-KI-VO zudem eine Grundrechte-Folgenabschätzung (GRFA) vor – ein zusätzliches prozedurales Element zur Sicherung des Grundrechtsschutzes bei Hochrisiko-KI-Systemen, das auf Initiative des Europäischen Parlaments und zivilgesellschaftlicher Akteure erst in der Aushandlungsphase der EU-KI-VO zusätzlich eingefügt wurde. Die GRFA ist somit auch für polizeiliche KI-Systeme vorgeschrieben, die in die Hochrisiko-Kategorie fallen. Sie ist vor Inbetriebnahme des Systems durchzuführen und bei Änderungen zu aktualisieren (Abs. 2). Die Inhalte der GRFA werden durch die EU-KI-VO vorstrukturiert (Art. 27 Abs. 1), nicht aber die normativen Anforderungen an das zu erreichende Schutzniveau für die Grundrechte.

Die Inhalte der GRFA überschneiden sich mit denen des Risikomanagementsystems nach Art. 9 EU-KI-VO. Allerdings sind für die GRFA die Betreibenden, für das Risikomanagementsystem dagegen die Anbietenden eines Hochrisiko-KI-Systems zuständig (vgl. Kleemann/Aden 2026). Inhalte einer zuvor durchgeführten Datenschutz-Folgenabschätzung (DSFA) nach Art. 35 DSGVO oder Art. 27 JI-Richtlinie können bereits Bestandteile des Pflichtprogramms der GRFA erfüllen und entsprechend übernommen werden. Die GRFA ergänzt dann die DSFA (Art. 27 Abs. 4 EU-KI-VO; vgl. Thomaidou/Limniotis 2025; Garcia-Godinez 2025) – und umgekehrt,

falls für neue KI-Systeme zuerst eine GRFA und erst danach eine DSFA durchgeführt wird.

Die Ergebnisse der GRFA werden an die zuständige Marktüberwachungsbehörde übermittelt. Die Pflichten zur Veröffentlichung der GRFA-Ergebnisse gelten für die Strafverfolgung nur mit erheblichen Einschränkungen (vgl. Kleemann/Aden 2026). Denn Betreibende müssen Hochrisiko-Systeme gem. Art. 49 Abs. 1 EU-KI-VO in einer EU-Datenbank (Art. 71 EU-KI-VO) registrieren. Behörden haben gem. Art. 49 Abs. 3 i. V. m. Anhang VIII Abschnitt C Nr. 4 EU-KI-VO grundsätzlich eine Ergebniszusammenfassung der GRFA zu veröffentlichen. Für KI-Systeme der Strafverfolgung (Anhang III Nr. 6 EU-KI-VO) entfällt dies gem. Art. 49 Abs. 4 EU-KI-VO. Diese Informationen werden lediglich in einem nicht-öffentlichen Teil der EU-Datenbank gespeichert. Zugriff haben nur die Europäische Kommission und die nationalen Marktüberwachungsbehörden (Art. 49 Abs. 4 UAbs. 2, Art. 74 Abs. 8 EU-KI-VO), und dies auch nur beschränkt auf Informationen gem. Anhang VIII Abschnitt C Nr. 1–3 (Art. 49 Abs. 4 lit. c EU-KI-VO), wobei die GRFA und DSFA dort in Nr. 4 und 5 genannt und somit gar nicht umfasst sind. Die Transparenz- und Dokumentationspflichten in besonders sensiblen Bereichen (Anhang III Nr. 1, 6 und 7) werden somit erheblich untergraben.

Das kann mit Blick auf unser Beispiel bedeuten, dass eine GRFA, die etwa die Fehlerquoten eines Systems zur automatisierten Sprechendererkennung bei der Überwachung von Telekommunikationsverbindungen oder einer KI-gestützten Objektdetektion in sichergestellten Datensätzen (inklusive möglicher diskriminierender Wirkungen für bestimmte Sprach- oder Bevölkerungsgruppen) enthält, für Betroffene, Zivilgesellschaft und unabhängige Forschung nicht einsehbar ist. *Accountability* im Sinne von Rechenschaft, Verantwortlichkeit und Nachweisbarkeit ist damit erheblich eingeschränkt. In einem effektiven *Accountability*-Forum ist die regelkonforme Nutzung nicht einfach anzunehmen, sondern muss auch nachgewiesen werden. Diese Anforderung trägt dazu bei, dass Verantwortung und die Konsequenzen getragen werden müssen, wenn etwas schief läuft.

5.3 Beschwerdeverfahren und Recht auf Erläuterung der Entscheidungsfindung

Die EU-KI-VO sieht ausdrücklich ein Beschwerderecht jeder natürlichen oder juristischen Person bei der zuständigen Marktüberwachungsbehörde

vor, wenn jemand „Grund zu der Annahme hat, dass gegen die Bestimmungen dieser Verordnung verstoßen wurde“ (Art. 85 EU-KI-VO). Dies mag zunächst selbstverständlich erscheinen, stärkt aber die Funktion der Marktüberwachungsbehörden als *Accountability*-Foren und erfordert die Einrichtung einer institutionalisierten Beschwerdebearbeitung. Diese Regelung ähnelt zwar dem Recht auf Beschwerde bei einer Aufsichtsbehörde bei Datenschutzverstößen gem. Art. 77 DSGVO, setzt aber keine persönliche Betroffenheit oder Verletzung einer natürlichen Person voraus (vgl. Hartmann 2026, Art. 85 Rn. 2ff.). Ein Verstoß durch ein Hochrisiko-KI-System ist ebenso wenig Voraussetzung für das Erheben einer solchen Beschwerde wie ein konkret entstandener Schaden. Es reicht die Annahme, dass ein Verstoß gegen die EU-KI-VO vorliegt.

Die Marktüberwachungsbehörde muss sich auf Basis der Marktüberwachungs- und Produktkonformitätsverordnung (VO (EU) 2019/1020, MÜ-VO) mit der Beschwerde auseinandersetzen. Konkretisierende Standards speziell für Beschwerden nach der EU-KI-VO fehlen allerdings, so dass der auf diesem Wege erreichbare Schutz zwischen den zuständigen Behörden variieren kann, solange keine vereinheitlichenden Standards erlassen werden. Das Beschwerdeverfahren ersetzt – selbstverständlich – ausdrücklich nicht andere Rechtsbehelfe und Klageverfahren und ist auch für Strafverfolgungsbehörden anwendbar.

Das Recht auf Erläuterung der Entscheidungsfindung im Einzelfall besteht gegenüber den Betreibenden von Hochrisiko-Systemen (Art. 86 EU-KI-VO). Es gilt ebenfalls für Strafverfolgungsbehörden; Geheimhaltungsinteressen sollten hier restriktiv gehandhabt werden. Im Gegensatz zu Art. 85 EU-KI-VO setzt dieses Recht allerdings eine Betroffenheit von der Entscheidung voraus.

5.4 Recht auf Anfechtung KI-basierter Entscheidungen (*right to contest*)

Das Recht auf Anfechtung umfasst die Möglichkeit, gegen Entscheidungen von Systemen Einspruch zu erheben oder sich effektiv zu beschweren, was angesichts von strukturellen Machtasymmetrien für die Gewährleistung von Handlungsfähigkeit und Fairness in digitalen Umgebungen von hoher Relevanz ist (vgl. Hirsbrunner u. a. 2025; Kleemann/Tahraoui 2026). Eine KI-basierte Entscheidung anzufechten, ist ein Recht, das in gemeinsamen Grundsätzen wie Transparenz, Zugänglichkeit und Autonomie verankert ist. Wenn die Anfechtbarkeit in das Design von Entscheidungssystemen

eingebettet ist, kann sie als Brücke zwischen Nutzenden, Betroffenen und Systemen dienen und Einzelpersonen Möglichkeiten bieten, sich aktiv mit Entscheidungen, die ihr Leben betreffen, auseinanderzusetzen und diese zu beeinflussen (Collins u. a. 2024, S. 96ff.). Wenn ein *right to contest* so verstanden und umgesetzt wird, kann es zu einem höheren Maß an Fairness, Gerechtigkeit und Rechtsstaatlichkeit beitragen, indem Fehler korrigiert, ungerechte Ergebnisse verhindert oder nachträglich geändert werden (Kaminski/Urban 2021, S. 1974f.).

Ein allgemeines, über den gerichtlichen Rechtsschutz im mitgliedstaatlichen Strafprozess- und Verwaltungsprozessrecht (z. B. Anfechtungsklage oder Beschwerde) hinausgehendes Recht auf Anfechtung KI-basierter Entscheidungen ergibt sich nicht explizit aus der EU-KI-VO. Es kann aber aus dem Grundrecht auf Privatsphäre (Art. 7 und 8 EU-GRCh) und auch aus Rechtsprechung des EuGH abgeleitet werden. Im Hinblick auf die staatliche KI-Nutzung in Deutschland ergänzt ein solches Recht die Rechtsschutzgarantie des Grundgesetzes (Art. 19 Abs. 4 GG). Auch die DSGVO respektive JI-Richtlinie lässt sich hierfür argumentativ heranziehen.

Datenschutzrechtlich bestimmt Art. 22 Abs. 1 DSGVO, dass Personen nicht einer ausschließlich automatisierten Entscheidung unterworfen werden dürfen und Verantwortliche „angemessene Maßnahmen [treffen müssen], um die Rechte und Freiheiten sowie die berechtigten Interessen der betroffenen Person zu wahren, wozu mindestens das Recht auf Erwirkung des Eingreifens einer Person seitens des Verantwortlichen, auf Darlegung des eigenen Standpunkts und auf *Anfechtung der Entscheidung* gehört“ (Abs. 3). Parallel dazu schreibt Art. 11 JI-Richtlinie für die Strafverfolgung vor, „dass eine ausschließlich auf einer automatischen Verarbeitung beruhende Entscheidung — einschließlich Profiling —, [...] verboten ist [...]“. Dazu heißt es in Erwägungsgrund 38 JI-Richtlinie konkretisierend: „In jedem Fall sollte eine solche Verarbeitung mit geeigneten Garantien verbunden sein, einschließlich [...] des Rechts [...] auf *Anfechtung der Entscheidung*.“ Ein Recht auf Anfechtung existiert demnach für gewisse Formen der Verarbeitung personenbezogener Daten, insbesondere für vollautomatisierte Verarbeitungsformen. Das dürfte auf einen erheblichen Teil KI-basierter Entscheidungen zutreffen, welche Einzelpersonen auf individueller Ebene betreffen.

Damit ein solches Recht effektiv ist, muss es zudem über ein bloßes Recht auf Berichtigung hinausgehen und zumindest die Verpflichtung umfassen, die Begründetheit der Beschwerde zu prüfen und die Entscheidung zu begründen. Die für die Datenverarbeitung Verantwortlichen müssen

dazu verpflichtet werden, die automatisierten Entscheidungen wirksam anfechtbar zu machen oder die Nutzung des algorithmischen Entscheidungssystems gänzlich einzustellen (Bayamhoğlu 2022, S. 1063).

Auch der EuGH hat in einer Entscheidung aus dem Februar 2025 ein „Recht auf Erklärung“ im Zusammenhang mit automatisierten Entscheidungen bestätigt.⁵ Das Urteil stellte klar, dass Gerichte und zuständige Behörden Zugang zu Informationen haben müssen, die durch Geschäftsgeheimnisse geschützt sind, wenn dies erforderlich ist, um diesen Schutz mit den Grundrechten der betroffenen Personen in Einklang zu bringen. Damit wird das Recht auf Anfechtung auch für KI-Systeme gegenüber gegenläufigen privatwirtschaftlichen Interessen gestärkt.

Eine Person, die z. B. aufgrund eines Treffers in einem KI-gestützten Objektdetektionssystem oder einer KI-basierten Sprechenderkennung in das Visier der Strafverfolgung geraten ist, wird in der Praxis kaum Kenntnis davon erlangen, dass dieses System zur Ermittlung gegen sie beigetragen hat, geschweige denn, welche Parameter und Trainingsdaten dem Ergebnis zugrunde lagen. Das Recht auf Anfechtung kann hier Gefahr laufen, strukturell leer zu bleiben, solange keine Offenlegungspflichten gegenüber Betroffenen bestehen.

5.5 Weitere prozedurale *Accountability*-Vorkehrungen

Darüber hinaus enthält die EU-KI-VO weitere prozedurale *Accountability*-Vorkehrungen in unterschiedlicher Zuständigkeit, so die Pflicht der Anbietenden von Hochrisiko-KI-Systemen, ein Qualitätsmanagementsystem einzurichten (Art. 17 EU-KI-VO) und die in der Zuständigkeit des KI-Büros liegenden Überwachungsmaßnahmen gegenüber Anbietenden von KI-Modellen mit allgemeinem Verwendungszweck (Art. 89 EU-KI-VO). Auch die Pflicht der Anbietenden zur Beobachtung ihrer Systeme nach dem Inverkehrbringen (Art. 72 EU-KI-VO) sei hier erwähnt. Sie umfasst auch die Pflicht zur Erhebung und Dokumentation relevanter Daten über das jeweilige KI-System während seiner gesamten Lebensdauer (Abs. 2).

Die freiwillig zu befolgenden Verhaltenskodizes für KI-Systeme, die nicht in die Hochrisiko-Kategorie fallen, sind schließlich einer der wenigen, allerdings schwach ausgeprägten Beiträge zu *Accountability*-Vorkehrungen außerhalb des Hochrisiko-Bereichs.

5 EuGH, Urteil vom 27.2.2025, Rechtssache C-203/22, CK gg. Magistrat der Stadt Wien (Dun & Bradstreet Austria).

5.6 Einbindung Externer als *Accountability*-Foren?

Die Möglichkeit einer Überprüfung KI-basierter Entscheidungen durch Dritte kann auch durch weitere Überwachungsmechanismen wie *Audits* erleichtert werden, die darauf ausgelegt sind, Fehler oder Schäden durch KI-Systeme zu vermeiden. An vielen Stellen überlässt die EU-KI-VO Anbietenden und Betreibenden Spielräume bei der Frage, wie sie die ihnen obliegenden laufenden Kontrollaufgaben für ihre KI-Systeme wahrnehmen. Hier kommt die Einbindung externer Expertise, etwa zertifizierter Auditor:innen, in Betracht. Auch notifizierte Stellen (Art. 3 Nr. 22 EU-KI-VO) könnten hier vor dem Hintergrund ihrer Aufgaben und Informationspflichten (Art. 31 und 45 EU-KI-VO) eingebunden werden. Weiteren Stellen, z. B. staatlichen Behörden und zivilgesellschaftlichen Organisationen, könnten Überwachungsaufgaben übertragen werden, soweit sie über den erforderlichen Sachverstand verfügen. Beispielsweise schreibt Art. 14 Abs. 1 EU-KI-VO eine menschliche Aufsicht für risikoreiche Systeme vor und verlangt darüber hinaus gründliche Untersuchungen von Vorfällen, um die Behebung von Fehlern oder entstandenen Schäden zu gewährleisten (vgl. Laux 2024, S. 2857). Diese Aufgaben könnten im Interesse der Qualitätssicherung von neutralen Dritten wahrgenommen werden. Dabei sind allerdings Interessenkonflikte zu vermeiden, etwa die Vereinnahmung der Auditor:innen durch Industrieinteressen. Als bewährte Gegenmaßnahmen sehen viele *Accountability*-Foren z. B. eine Rotation der Auditor:innen und Prüfer:innen vor (vgl. Laux 2024, S. 2861).

Die Wahrnehmung dieser Aufgaben erfordert interdisziplinäre Expertise, da bei der Prüfung von KI-Systemen oftmals technische, rechtliche und oft auch sozialwissenschaftliche Fragen eng miteinander verknüpft sind. Forschende, zivilgesellschaftliche Akteur:innen und unabhängige Expert:innen können an solchen Audits beteiligt sein, um einen umfassenden, vielfältigen und letztlich effektiven Rechenschaftsrahmen zu etablieren. Das Risiko eines sogenannten *Audit Washing* besteht immer dann, wenn interne Audits oder Konformitätsbewertungen lediglich von den Entwickelnden und Anbietenden von KI-Systemen durchgeführt werden (Hartmann u. a. 2025).

6. *Accountability*- und *Compliance*-Vorkehrungen im Kontext bestehender *Accountability*-Foren in der EU und den Mitgliedstaaten

Wie gezeigt, hat die EU-KI-VO zusätzliche, stark fragmentierte *Accountability*- und *Compliance*-Vorkehrungen institutioneller und vor allem pro-

zeduraler Art etabliert, deren Anwendungsfeld weitgehend auf Hochrisiko-KI-Systeme beschränkt bleibt. Keine der KI-Behörden fungiert indes als durchsetzungsstarkes *Accountability*-Forum. Die bestehenden *Accountability*-Foren bleiben daher von großer Bedeutung, auch für die Kontrolle der KI-Nutzung, insbesondere der Europäische Datenschutzbeauftragte, die EU-Bürgerbeauftragte, die EU-Grundrechteagentur, der EuGH sowie die mitgliedstaatlichen Aufsichtsbehörden und Gerichte.

KI-Anwendungen benötigen große Mengen an Daten für das Training und Testen. Viele KI-Anwendungen verarbeiten auch für ihre Kernzwecke personenbezogene Daten, so dass auch insofern die Datenschutzaufsicht zuständig ist. Vor diesem Hintergrund ist es bemerkenswert, dass die Datenschutzaufsicht trotz ihrer einschlägigen Kompetenz und Erfahrungen in der EU-KI-VO nur am Rande vorkommt. Dies gilt in besonderem Maße für Mitgliedstaaten wie Deutschland, in denen die Funktion der KI-Marktüberwachungsbehörde entgegen Art. 74 Abs. 8 EU-KI-VO nicht bei den Datenschutzbehörden angesiedelt wurde. Die EU-KI-VO ermöglicht zwar, bestimmte Kontrollaufgaben an die Datenschutzaufsicht zu übertragen, insbesondere mit Blick auf die KI-Nutzung durch Strafverfolgungsbehörden, jedoch haben die Mitgliedstaaten dabei erhebliche Entscheidungsspielräume.

Der Europäische Datenschutzbeauftragte (*European Data Protection Supervisor*, EDPS) übernimmt die Funktion der Aufsichtsbehörde für die KI-Nutzung durch Organe, Einrichtungen und sonstige Stellen der EU (Art. 70 Abs. 9 und Art. 74 Abs. 9 EU-KI-VO). Hierunter fallen auch die für die Polizei- und Justizzusammenarbeit zuständigen EU-Agenturen, insbesondere Europol, Eurojust und eu-LISA. An dem KI-Gremium, das auf der Basis von Art. 65 EU-KI-VO eingerichtet wird, ist der EDPS allerdings nur mit Beobachterstatus beteiligt.

Andere EU- und mitgliedstaatliche *Accountability*-Foren wie die Ombudsinstitutionen und die Rechnungshöfe finden in der EU-KI-VO keine Erwähnung, können aber im Rahmen ihrer allgemeinen Zuständigkeit Kontrollaufgaben auch bezüglich der behördlichen KI-Nutzung wahrnehmen. Dies führt zur Notwendigkeit einer engen Abstimmung und Koordination zwischen den diversen *Accountability*-Foren, damit es nicht zu einer *Accountability*-Diffusion, Fragmentierung und Kompetenzgerangel kommt und Rechtssicherheit erreicht werden kann. Für die mitgliedstaatlichen Gerichte und den EuGH dürfte sich die Auslegung der EU-KI-VO ebenfalls zu einem neuen Arbeitsfeld entwickeln, sobald einschlägige Fälle dorthin gelangen.

7. Schlussfolgerungen und Ausblick

Dieser Beitrag hat gezeigt, dass die EU-KI-VO fragmentierte Mehrebenen-*Accountability*- und *Compliance*-Strukturen vorsieht, deren Anwendungsbereich sich weitgehend auf Hochrisiko-KI-Systeme beschränkt und die mit einigen Einschränkungen auch für die polizeiliche KI-Nutzung anwendbar sind.

Damit wird die Risikoklassifizierung eines KI-Systems zur entscheidenden Weichenstellung für die Intensität der prozeduralen Anforderungen und der Aufsichtsbefugnisse gegenüber der KI-Nutzung. Während der Aushandlung der EU-KI-VO kam es zu signifikanten Einschränkungen der Hochrisiko-Kategorie für Strafverfolgungsbehörden (vgl. Kleemann/Aden 2025) und damit auch zu einer Reduzierung der *Accountability*-Vorkehrungen für die polizeiliche KI-Nutzung. Bestrebungen im Rahmen des sogenannten *Digital Omnibus* (Europäische Kommission 2025), die EU-KI-VO und weitere Digitalgesetze (u. a. auch die DSGVO) zugunsten von Interessen an vereinfachten Verfahren zu ändern, berühren auch die Regeln für die Strafverfolgung. Allerdings fehlt es an Evidenz für die Notwendigkeit der Änderungen eines so neuen Rechtsakts wie der EU-KI-VO, dessen Anwendungspraxis noch begrenzt ist (zur Kritik vgl. Leufer 2025, Dachwitz 2025). Die Übergangsphase bis zur vollen Geltung der Hochrisiko-Anforderungen wurde schließlich verlängert (Art. 113 EU-KI-VO in der durch Verordnung (EU) 2026/1744 geänderten Fassung). Während die Europäische Kommission unter dem Druck von Interessengruppen und einiger mitgliedstaatlicher Regierungen versuchte, das Schutzniveau zu senken, schlug sie mit dem *Digital Omnibus* zugleich die Stärkung des KI-Büros und damit einer kommissionseigenen Dienststelle vor und setzte sich damit weitgehend durch (Art. 75a, 75b und 75c EU-KI-VO in der durch Verordnung (EU) 2026/1744 geänderten Fassung). Eine solche Zentralisierung auch im Bereich der Strafverfolgung wurde von einigen mitgliedstaatlichen Regierungen als zu starker Eingriff in ihre Kompetenzen kritisiert, weswegen sie eine Ausnahme von dem Aufgabenzuwachs für das KI-Büro befürworten (Bertuzzi 2026a). Dies wurde in Art. 75 in der durch Verordnung (EU) 2026/1744 geänderten Fassung festgeschrieben. Zudem schlossen sich 23 der 27 Mitgliedsstaaten einem Brief an, mit dem sie Einfluss auf die Leitlinien zur Einstufung von Hochrisiko-KI fordern, um zu einer möglichst engen Auslegung des Anwendungsbereichs der Hochrisiko-Anforderungen zu kommen; insbesondere der Bereich Strafverfolgung bedürfe „besonderer Überlegungen“, damit die Fähigkeiten von Strafverfolgungsbehörden nicht

untergraben würden (Bertuzzi 2026b). In einem Bereich, der ohnehin von Ausnahmen und „Hintertüren“ (siehe oben, Abschnitt 1) geprägt ist, sollte hier allerdings das Schutzniveau eher gesteigert als abgeschwächt werden.

Effektive *Accountability*- und *Compliance*-Vorkehrungen für die KI-Entwicklung und -Nutzung zeichnen sich durch eine ganzheitliche Perspektive auf technische Eigenschaften und den rechtlichen Rahmen sowie auf ihre Wechselwirkungen aus. Aufgrund der Fragmentierung der institutionellen Strukturen und der Aufsichtskompetenzen fehlt eine solche ganzheitliche Perspektive bisher. Wichtige Funktionen werden in diesem Kontext sogar den Anbietenden und Betreibenden von KI-Systemen überlassen, andere Aufsichtsfunktionen sind abhängigen Behörden wie dem KI-Büro in der Europäischen Kommission und Marktüberwachungsbehörden in den Mitgliedstaaten zugewiesen. Die EU-KI-VO gewährleistet daher aus sich heraus weder eine effektive Kontrolle der KI-Entwicklung und -Nutzung im Allgemeinen noch speziell der polizeilichen KI-Nutzung.

Absehbar ist die EU-KI-VO erst der Anfang der KI-Regulierung in Europa. Für die Rechtsprechung sowie für Empfehlungen und Entscheidungen der in der Verordnung vorgesehenen Behörden und Gremien verbleibt dabei ein großer Konkretisierungsbedarf, insbesondere zur Sicherung des in der EU-KI-VO an vielen Stellen geforderten hohen Schutzniveaus für „Grundrechte, Gesundheit und Sicherheit“, gerade in grundrechtssensiblen Bereichen wie der Strafverfolgung. Manches bleibt durch zukünftige weitere Gesetzgebung zu konkretisieren. Weiterer interdisziplinärer Forschungsbedarf besteht dabei unter anderem für die Übersetzung abstrakter Begriffe in handhabbare. In der EU-KI-VO, der KI-Rahmenkonvention des Europarats (CETS No. 225) und auch in *soft-law*-Instrumenten wie den OECD-Empfehlungen zu KI (OECD/LEGAL/0449) oder den UNESCO Empfehlungen zur Ethik von KI (SHS/BIO/PI/2021/1) wurden grundlegende Prinzipien für die KI-Entwicklung und -Nutzung definiert. Diese bedürfen der Konkretisierung hinsichtlich der Operationalisierung von Grundrechtsschutz in Form messbarer Indikatoren für Grundrechtseingriffe durch KI-Systeme, auch im Rahmen von Strafverfolgung.

Für KI-Systeme zur Objektdetektion oder Sprechendenerkennung in der Strafverfolgung könnte dies etwa bedeuten: Welche Fehlerquoten (hier insbesondere *False-Positive*-Raten) sind mit dem Verhältnismäßigkeitsgrundsatz noch vereinbar, auch unter Berücksichtigung der Intensität des jeweiligen Grundrechtseingriffs? Welche Mindestanforderungen an Trainingsdaten gelten, um diskriminierende Outputs zu vermeiden? Und ab welchem

Automatisierungsgrad ist eine menschliche Letztentscheidung zwingend erforderlich, bevor ein KI-Treffer zu einer operativen Maßnahme führt?

Für einzelne KI-Systeme könnte dies im Rahmen einer umfassend verstandenen Grundrechte-Folgenabschätzung geschehen. Konzepte von menschlicher Aufsicht und Intervention bedürfen weiterer Konkretisierung hinsichtlich einheitlicher Definitionen, Handhabung oder Umsetzung. Dazu zählen z. B. Art. 22 DSGVO, Art. 14 EU-KI-VO und weitere Ausformungen im mitgliedstaatlichen Recht. Offene Fragen verbleiben auch für die Haftung bei Fehlentscheidungen, im polizeilichen Kontext auch für die Rechenschaftspflicht für unrechtmäßige Polizeimaßnahmen, die auf KI-Entscheidungen zurückgehen.

Ein maßgeblicher Ansatzpunkt für effektive *Accountability*-Strukturen ist das Zusammenspiel aller beteiligten Akteur:innen im KI-Lebenszyklus. Technischen KI-Entscheidungen folgen echte soziale und rechtliche Auswirkungen. Sozio-technische Systeme müssen daher auch als solche behandelt werden. Ein Zurückziehen auf reine technische (Un-)Möglichkeiten oder rechtliche Vorgaben kann hier nicht funktionieren.

Literatur

- Aden, Hartmut (2025): Dilemmata der Polizeibeauftragten, *vorgänge. Zeitschrift für Bürgerrechte und Gesellschaftspolitik*, Heft 249 (= 64(1)), S. 3–14.
- Aden, Hartmut und Fährmann, Jan (2021): Argumente für einen besseren Musterentwurf für einheitlichere Polizeigesetze. Kritische Analyse von Entwicklungen im Polizeirecht aus rechtsstaatlicher und bürgerrechtlicher Perspektive. In: Martin H. W. Möllers und Robert Chr. van Ooyen (Hrsg.): *Jahrbuch Öffentliche Sicherheit 2020/2021*, Baden-Baden: Nomos und Frankfurt/Main: Verlag für Polizeiwissenschaft, S. 580–615.
- Bayamlioglu, Emre (2022): The right to contest automated decisions under the General Data Protection Regulation: Beyond the so-called “right to explanation”. *Regulation & Governance*, 16(4), S. 1058–1078. <https://doi.org/10.1111/rego.12391>.
- Bertuzzi, Luca (2026a): Can EU plans to centralize AI law enforcement endure pushback by member states? *MLex* vom 30.01.2026, URL: <https://www.mlex.com/mlex/articles/2435744/can-eu-plans-to-centralize-ai-law-enforcement-endure-pushback-by-member-states-> (besucht am 13.05.2026).
- Bertuzzi, Luca (2026b): EU countries press for law enforcement considerations in AI law guidelines. *MLex* vom 28.01.2026, URL: <https://www.mlex.com/mlex/articles/2434470/eu-countries-press-for-law-enforcement-considerations-in-ai-law-guidelines> (besucht am 13.05.2026).

- Bosman, Timo (2025): Kommentierung zu Art. 25 EU-KI-VO. In: Schefzig, Jens und Kilian, Robert (Hrsg.): BeckOK KI-Recht, 5. Edition
Stand: 01.11.2025, München: C.H. Beck.
- Bovens, Mark (2007): Analysing and assessing accountability: A conceptual framework. *European Law Journal*, 13(4), S. 447–468.
- Bovens, Mark; Goodin, Robert E. und Schillemans, Thomas (2014): Public Accountability. In: Mark Bovens, Thomas Schillemans und Robert E. Goodin (Hrsg.): *The Oxford Handbook of Public Accountability*. Oxford: Oxford University Press, S. 1–20.
- Collins, Robert Patrick; Redström, Johan und Rozendaal, Marco (2024): The Right to Contestation: Towards Repairing Our Interactions with Algorithmic Decision Systems, *International Journal of Design*, 18, S. 95–106.
- Dachwitz, Ingo (2025): „Größter Rückschritt für digitale Grundrechte in der Geschichte der EU“. *Netzpolitik* vom 13.11.2025, URL: <https://netzpolitik.org/2025/digitaler-omnibus-groesster-rueckschritt-fuer-digitale-grundrechte-in-der-geschichte-der-eu/> (besucht am 08.02.2026).
- Europäische Kommission (2025): Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Änderung der Verordnungen (EU) 2024/1689 und (EU) 2018/1139 im Hinblick auf die Vereinfachung der Umsetzung harmonisierter Vorschriften für künstliche Intelligenz (Digital-Omnibus-Verordnung zur KI), COM(2025) 836 final 2025/0359 (COD). URL: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=celex%3A52025PC0836> (besucht am 20.07.2026).
- European Center for Not-for-Profit Law (ECNL) (2026): Four things to know about the EU AI Act Advisory Forum. URL: <https://ecnl.org/news/four-things-know-about-eu-ai-act-advisory-forum> (besucht am 20.07.2026).
- Fährmann, Jan; Aden, Hartmut und Arzt, Clemens (2022): Transparenz der polizeilichen Datenverarbeitung: Defizite und technische Lösungsansätze: In: Michael Friedewald, Michael Kreutzer und Marit Hansen (Hrsg.): *Selbstbestimmung, Privatheit und Datenschutz - Gestaltungsoptionen für einen europäischen Weg*, Wiesbaden: Springer, S. 303–326.
- Fischer, Maximilian T.; Schlegel, Udo; Keim, Daniel A.; Altmann, Stefan; Grothe, Claudia; Reuter, Paul; Coleman, Graham; Geierhos, Manuela; Maoro, Falk; Kluin, Marco; Weinbuch, Monika; Aden, Hartmut; Kleemann, Steven; Tahraoui, Milan; Louban, Anna; Arndt, Maximilian; Schönrock, Sabrina; Brandner, Lou T.; Hirsbrunner, Simon D.; Loh, Wulf und Yilmaz, Yusuf (2025): *DIN SPEC 91517:2025–05. Anforderungen an vertrauenswürdige KI-Methoden in polizeilichen Anwendungen*. Berlin: DIN Media GmbH, <https://doi.org/10.31030/3612025>.
- Garcia-Godinez, Miguel (2025): A default framework for fundamental rights impact assessment. *AI Ethics*, 5, S. 5149–5163, <https://doi.org/10.1007/s43681-025-00761-1>.
- Guerra, Georgia (2019): An Interdisciplinary Approach for Comparative Lawyers: Insights from the Fast-Moving Field of Law and Technology. *German Law Journal*, 19(3), S. 579–601, <https://doi.org/10.1017/S2071832200022793>.
- Hartmann, David; Laranjeira de Pereira, José Renato; Streitbürger, Chiara und Berendt, Bettina (2025): Addressing the regulatory gap: moving towards an EU AI audit ecosystem beyond the AI Act by including civil society. *AI Ethics*, 5, S. 3617–3638.

- Hartmann, Sarah (2026): Kommentierung zu Art. 85 EU-KI-VO. In: Martini, Mario und Wendehorst, Christiane (Hrsg.): KI-VO: Verordnung über künstliche Intelligenz, Kommentar. 2. Auflage, München: C.H. Beck.
- Hirsbrunner, Simon; Kleemann, Steven und Tahraoui, Milan (2025): Contestation in artificial intelligence as a practice: from a system-centered perspective of contestability towards normative contextualization, situative critique and organizational culture. *Frontiers in Communication* 10. <https://doi.org/10.3389/fcomm.2025.1638257>.
- Kaminski, Margot E. und Urban, Jennifer M. (2021): The Right to Contest AI, in: *Columbia Law Review*, 121, S. 1957–2048.
- Kleemann, Steven und Aden, Hartmut (2025): Die Nutzung Künstlicher Intelligenz durch Strafverfolgungsbehörden – „Hintertüren“ der Verordnung der Europäischen Union über Künstliche Intelligenz. In: Wilfried Honekamp, Stefanie Kemme und Jens Struck (Hrsg.): *Auswirkungen von Künstlicher Intelligenz auf die zukünftige Polizeiarbeit. Technologische Potenziale, rechtliche Rahmenbedingungen, kriminologisch-sozialwissenschaftliche Erkenntnisse*, Wiesbaden: Springer, S. 3–30.
- Kleemann, Steven und Aden, Hartmut (2026): Die Grundrechte-Folgenabschätzung nach dem Artificial Intelligence Act der Europäischen Union – Chancen einer qualitativen Konkretisierung für die polizeiliche KI-Nutzung. *Die Polizei*, 117(4a) (Sonderheft), S. 226–230.
- Kleemann, Steven und Tahraoui, Milan (2026): Responsibility and Accountability for the use of AI in Law Enforcement in the European Union: Lost in Negotiations?. *MenschenRechtsMagazin*, 30(2), S. 117–143. <https://doi.org/10.60935/mrm2025.30.28>.
- Krahn, Noelle V. (2024): The European Union’s Artificial Intelligence Act – A Status Report, *Transatlantic Law Journal (TLJ)*, 3, S. 120–125.
- Laux, Johann (2024): Institutionalised distrust and human oversight of artificial intelligence: Towards a democratic design of AI governance under the European Union AI Act. *AI & Society*, 39(6), S. 2853–2866.
- Leufer, Daniel (2025): Digital Rights Are on the Chopping Block in the European Commission’s Omnibus. *Tech Policy Press* vom 19.11.2025, URL: <https://www.techpolicy.press/digital-rights-are-on-the-chopping-block-in-the-european-commissions-omnibus/> (besucht am 13.05.2026).
- Martini, Mario und Botta, Jonas (2025): Polizeiliche Datenanalyse mittels KI – Unions-, verfassungs- und polizeirechtliche Leitplanken für Palantir & Co. *Die Öffentliche Verwaltung (DÖV)*, 78(24), S. 1033–1044.
- Martinez Demarco, Sol; Tahraoui, Milan und Kleemann, Steven (2026): The siloed logic and the implementation of the Artificial Intelligence Act in the law enforcement context: legal and ethical analysis of the applicability of accountability and responsibility to high-risk AI systems. In: Marie Eneman, Diana Miranda, Jan Ljungberg und William Webster (Hrsg.): *AI Surveillance and the Future of Policing: Opportunities, Threats, Perspectives and Cases*. London: Routledge, i.E.
- Novelli, Claudio; Taddeo, Mariarosaria und Floridi, Luciano (2024): Accountability in artificial intelligence: What it is and how it works. *AI & Society*, 39, S. 1871–1882.

- Novelli, Claudio; Hacker, Philipp; Morley, Jessica; Trondal, Jarle und Floridi, Luciano (2025): A Robust Governance for the AI Act: AI Office, AI Board, Scientific Panel, and National Authorities. *European Journal of Risk Regulation*. 16(2), S. 566–590. <https://doi.org/10.1017/err.2024.57>.
- Schuett, Jonas (2024): Risk Management in the Artificial Intelligence Act. *European Journal of Risk Regulation*, 15(2), S. 367–385. <https://doi.org/10.1017/err.2023.1>.
- Thomaidou, Anna und Limniotis, Konstantinos (2025): Navigating Through Human Rights in AI: Exploring the Interplay Between GDPR and Fundamental Rights Impact Assessment. *Journal of Cybersecurity and Privacy*, 5(1), 7, <https://doi.org/10.3390/jcp5010007>.
- Töpfer, Eric; John, Sonja und Aden, Hartmut (2023): *Parlamentarische Polizeibeauftragte: Menschenrechtliche Empfehlungen für die Stellen in Bund und Länder*. Berlin: Deutsches Institut für Menschenrechte und Forschungsinstitut für Öffentliche und Private Sicherheit, <https://doi.org/10.4393/opushwr-4265>

Mitarbeiterinnen und Mitarbeiter dieses Bandes

Dr. Hartmut Aden ist Professor für Öffentliches Recht, Europarecht, Politik- und Verwaltungswissenschaft am Fachbereich Polizei und Sicherheitsmanagement der Hochschule für Wirtschaft und Recht Berlin sowie stv. Direktor des Forschungsinstituts für Öffentliche und Private Sicherheit (FÖPS Berlin) der Hochschule. E-Mail: Hartmut.Aden@hwr-berlin.de

Bruno Albert ist wissenschaftlicher Mitarbeiter im interdisziplinären Forschungsprojekt „Staatliche Eingriffe in private Endgeräte zur Strafverfolgung (SEpES)“ am Bayerischen Forschungsinstitut für Digitale Transformation (bidt) in München. E-Mail: bruno.albert@bidt.digital

Pascal Altenkamp war bis Ende Oktober 2025 wissenschaftlicher Mitarbeiter am Institut für Digitales Management und Neue Medien der Munich School of Management an der Ludwig-Maximilians-Universität München und Teil des Forschungsprojekts „Demokratie-X: Analyse der Tragfähigkeit einer privatheitsschonenden, fairen und gemeinwohlorientierten Plattform für Nachrichten“.

Franziska Baum ist wissenschaftliche Mitarbeiterin an der Juniorprofessur „Soziologie mit Schwerpunkt Technik“ und an der Professur für „Soziologie mit Schwerpunkt Arbeit, Wirtschaft und Organisation (Prof. Dr. Tanja Carstensen) der TU Chemnitz, und ebenfalls Mitarbeiterin im Projekt *Simplications*.

Joel Baumann ist Professor für Neue Medien im Fachbereich Visuelle Kommunikation an der Kunsthochschule Kassel (seit 2003) und war dort drei Jahre als Prorektor sowie acht Jahre als Rektor tätig. Seine Forschungsschwerpunkte liegen im Bereich der Wissenschaftskommunikation sowie der kritischen Auseinandersetzung mit KI-Systemen, insbesondere im Kontext intimer Mensch-Maschine-Beziehungen und der Anwendung von KI in Kunst und Design. Im drittmittelgeförderten Forschungsprojekt SENTIMENT, das sichere intime Selbstoffenbarung mit KI-Systemen untersucht, verantwortet er den künstlerischen Bereich. E-Mail: jbaumann@uni-kassel.de

Dr. Arne Berger ist Teilprojektleiter in *Simplications* und Professor für Mensch-Computer Interaktion an der Hochschule Anhalt.

Dr. Andreas Bischof ist Koordinator von *Simplications* und Juniorprofessor für „Soziologie mit Schwerpunkt Technik“ an der TU Chemnitz.

Hans Brorsen ist Rechtsanwalt und Mitgründer der Valid Technologies GmbH. E-Mail: hans@valid.tech

Florian Dehling ist wissenschaftlicher Mitarbeiter am Institut für Informatik der Justus-Liebig-Universität Gießen und forscht dort in der Gruppe für Daten und Anwendungssicherheit. E-Mail: florian.dehling@uni-giessen.de

Dr. Felix Freiling ist Inhaber des Lehrstuhls für IT-Sicherheit an der Friedrich-Alexander-Universität Erlangen-Nürnberg, Sprecher des DFG-Graduiertenkollegs 2475 "Cyberkriminalität und forensische Informatik" und Mitglied im Direktorium des Bayerischen Forschungsinstituts für Digitale Transformation (bidt). E-Mail: felix.freiling@fau.de

Dr. Michael Friedewald leitet das Geschäftsfeld „Informations- und Kommunikationstechnik“ am Fraunhofer-Institut für System- und Innovationsforschung ISI in Karlsruhe. Er ist Koordinator der „Plattform Privatheit“. E-Mail: michael.friedewald@isi.fraunhofer.de

Dr. Christian Geminn ist Privatdozent für Öffentliches Recht und Recht der digitalen Gesellschaft sowie Geschäftsführer der Projektgruppe verfassungsverträgliche Technikgestaltung (provet) im Wissenschaftlichen Zentrum für Informationstechnik-Gestaltung (ITeG) an der Universität Kassel und geschäftsführender Gesellschafter der Datenrecht Beratungsgesellschaft (DRBG). Er ist Mitglied der „Plattform Privatheit“ und Koordinator des Projekts PriDI sowie Teilprojektverantwortlicher in den Projekten SENTIMENT, DEEP-PRISMA, PRETINA und DiversPrivat. E-Mail: c.geminn@uni-kassel.de

Moritz Griesel ist Wissenschaftlicher Mitarbeiter am Lehrstuhl für Bürgerliches Recht, Wettbewerbs- und Immaterialgüterrecht, Medien- und Informationsrecht (Prof. Dr. Andreas Wiebe, LL.M (Vigina)) an der Georg-August-Universität Göttingen. E-Mail: moritz.griesel@jura.uni-goettingen.de

Dr. Daniel Guagnin leitet den Bereich „Netze und Gesellschaft“ am nexus Institut. Er forscht und lehrt zu gesellschaftlichen Auswirkungen von Technik und ihrer partizipativen Gestaltung. Er berät Unternehmen zu KI-Governance und Datenschutz. Email-Adresse: soziologe@guagnin.de

Dr. Andreas Hammer ist Postdoc am Lehrstuhl für IT-Sicherheitsinfrastrukturen der FAU Erlangen-Nürnberg. Er beschäftigt sich mit den Forschungsbereichen IT-Forensik und Human Factors in IT-Security. E-Mail: andreas.hammer@fau.de

Antonios Hazim ist Softwareentwickler bei [tech@spree](mailto:tech@spree.de). Zum Zeitpunkt der Entstehung dieses Bandes war er am nexus Institut im Bereich „Netze und Gesellschaft“ an der Forschung zu Themen des Datenschutzes, der KI und der partizipativen Technikentwicklung beteiligt. E-Mail: antonios.hazim@spree.de

Dr. Dirk Heckmann ist Inhaber des Lehrstuhls für Recht und Sicherheit der Digitalisierung an der TU München, Mitglied des Bayerischen Verfassungsgerichtshofs, Direktor des TUM Center for Digital Public Services und Mitglied des Bayerischen Forschungsinstituts für Digitale Transformation (bidt). E-Mail: dirk.heckmann@tum.de

Dr. Jessica Heesen ist Vorstandsmitglied des Internationalen Zentrums für Ethik in den Wissenschaften der Universität Tübingen und leitet dort die Forschungsgruppe *Medienethik, Technikphilosophie & KI*. Sie ist Mitglied des wissenschaftlichen Beirats der „Plattform Privatheit“ und ist an verschiedenen Forschungsprojekten der Plattform beteiligt. E-Mail: jessica.heesen@uni-tuebingen.de

Dr. Nils B. Heyen ist Soziologe und wissenschaftlicher Projektleiter in der Abteilung Neue Technologien am Fraunhofer-Institut für System- und Innovationsforschung ISI in Karlsruhe. Er leitet das im Rahmen der Plattform Privatheit geförderte Projekt „PRIMA – Privatheitsmanagement bei Gesundheits-Apps entlang der Patient Journey“. E-Mail: nils.heyen@isi.fraunhofer.de

Eva Isabel Hoffmeister ist Doktorandin der Wirtschaftsinformatik am Institut für Digitales Management und Neue Medien der Munich School of Management an der Ludwig-Maximilians-Universität München. Sie ist Teil

des Forschungsprojekts „Demokratie-X: Analyse der Tragfähigkeit einer privatheitsschonenden, fairen und gemeinwohlorientierten Plattform für Nachrichten“. E-Mail: eva.hoffmeister@lmu.de

Dr. Murat Karaboga ist wissenschaftlicher Mitarbeiter am Fraunhofer-Institut für System- und Innovationsforschung ISI in Karlsruhe. E-Mail: murat.karaboga@isi.fraunhofer.de

Joanna Klauser ist wissenschaftliche Mitarbeiterin im interdisziplinären Forschungsprojekt „Staatliche Eingriffe in private Endgeräte zur Strafverfolgung (SEpES)“ am Bayerischen Forschungsinstitut für Digitale Transformation (bidt) in München. E-Mail: joanna.klauser@bidt.digital

Steven Kleemann ist wissenschaftlicher Mitarbeiter beim Deutschen Institut für Menschenrechte und Doktorand an der Universität Potsdam. E-Mail: steven.kleemann@uni-potsdam.de

Jana Koch war im Rahmen des Projekts PRIMA für den Projektpartner C&S Computer und Software GmbH tätig. Inzwischen arbeitet sie als wissenschaftliche Mitarbeiterin an der Medizinischen Fakultät OWL der Universität Bielefeld. E-Mail: jana.koch@uni-bielefeld.de

Sebastian Koch ist wissenschaftlicher Mitarbeiter am Institut für Informatik der Justus-Liebig-Universität Gießen und forscht dort in der Gruppe für Daten und Anwendungssicherheit. E-Mail: sebastian.koch@uni-giessen.de

Dr. Lindrit Kqiku ist wissenschaftlicher Mitarbeiter der Forschungsgruppe Computer Sicherheit und Privatheit an der Georg-August-Universität Göttingen. E-Mail: kqiku@cs.uni-goettingen.de

Thomas Kunz ist wissenschaftlicher Mitarbeiter am Fraunhofer-Institut für Sichere Informationstechnologie SIT und am Nationalen Forschungszentrum für angewandte Cybersicherheit ATHENE. E-Mail: thomas.kunz@sit.fraunhofer.de

Dr. Albrecht Kurze ist Teilprojektleiter in *Simplications* und wissenschaftlicher Mitarbeiter an der Professur Medieninformatik der TU Chemnitz.

Emma Carlotta Laumann ist Datenanalystin im Zentrum für Geopolitik, Geoökonomie und Technologie der Deutschen Gesellschaft für Auswärtige Politik (DGAP). Sie hat einen Forschungsmaster in Sozialwissenschaften mit Schwerpunkt auf quantitativer Methodik von der Universität Amsterdam (UvA) und einen Bachelor-Abschluss in Politikwissenschaft von der Vrije Universiteit Amsterdam (VU).

Dr. Tanja Lehmann ist wissenschaftliche Mitarbeiterin an der Juniorprofessur „Soziologie mit Schwerpunkt Technik“ der TU Chemnitz und im BMFTR-geförderten Projekt *Simplications*.

Dr. Christian Lindenmeier hat am Lehrstuhl für IT-Sicherheitsinfrastrukturen der FAU Erlangen-Nürnberg promoviert und beschäftigt sich mit der digitalen Forensik in der Strafverfolgung sowie der Sicherheit von mobilen Endgeräten. E-Mail: christian.lindenmeier@fau.de

Dr. Luigi Lo Iacono ist Professor für IT-Sicherheit an der Justus-Liebig-Universität Gießen und dort auch der gesamtuniversitäre Informationssicherheitsbeauftragte. E-Mail: luigi.lo_iacono@uni-giessen.de

Dr. Matthias Meitzler ist wissenschaftlicher Mitarbeiter in der Forschungsgruppe Medienethik, Technikphilosophie & KI am Internationalen Zentrum für Ethik in den Wissenschaften (IZEW) der Universität Tübingen. E-Mail: matthias.meitzler@uni-tuebingen.de

Lisa Mühl ist wissenschaftliche Mitarbeiterin und Doktorandin in der Forschungsgruppe INTITEC (Intimacy With and Through Technology) an der Universität Duisburg-Essen. Im BMFTR-geförderten Projekt SENTIMENT erforscht sie, wie Intimität in Interaktionen mit LLM-basierten Kommunikationssystemen wie ChatGPT entsteht und welche Rolle Selbstoffenbarung, Personalisierung und Systemeigenschaften dabei spielen, einschließlich der psychologischen Untersuchung von Privacy-by-Design Interventionen im Sinne des EU AI Act. E-Mail: lisa.muehl@uni-due.de

Dr. Katja Muñoz ist Senior Research Fellow am Zentrum für Geopolitik, Geoökonomie und Technologie der Deutschen Gesellschaft für Auswärtige Politik (DGAP). Ihre Forschungsarbeit befasst sich mit der kritischen Schnittstelle zwischen künstlicher Intelligenz, Demokratie und Sicherheit,

mit besonderem Schwerpunkt auf der komplexen Beziehung zwischen sozialen Medien und politischen Dynamiken.

Maria Pawelec ist wissenschaftliche Mitarbeiterin in der Forschungsgruppe Medienethik, Technikphilosophie & KI am Internationalen Zentrum für Ethik in den Wissenschaften (IZEW) der Universität Tübingen. E-Mail: maria.pawelec@uni-tuebingen.de

Dr. Sabine Pfeiffer ist Inhaberin des Lehrstuhls für Soziologie (Technik – Arbeit – Gesellschaft) an der Friedrich-Alexander-Universität Erlangen-Nürnberg, Sprecherin des DFG-Schwerpunktprogramms 2267 „Digitalisierung der Arbeitswelten“ und Mitglied im Direktorium des Bayerischen Forschungsinstituts für Digitale Transformation (bidt). E-Mail: sabine.pfeiffer@fau.de

Dr. Daniela Pöhn ist kommissarische wissenschaftlich-technische Geschäftsführerin am Forschungsinstitut Cyber Defence and Smart Data (FI CODE) der Universität der Bundeswehr München (UniBw M) und Privatdozentin an der UniBw M. E-Mail: daniela.poehn@unibw.de

Dr. Delphine Reinhardt ist Professorin an der Georg-August-Universität Göttingen und leitet dort die Forschungsgruppe Computer Sicherheit und Privatheit. E-Mail: reinhardt@cs.uni-goettingen.de

Dr. Alexander Roßnagel ist Hessischer Beauftragter für Datenschutz und Informationsfreiheit, Universitätsprofessor für öffentliches Recht i.R. und Direktor des Wissenschaftlichen Zentrums für Informationstechnik-Gestaltung (ITeG) der Universität Kassel sowie Sprecher der Plattform Privatheit. E-Mail: a.rossnagel@uni-kassel.de

Volkan Sayman ist wissenschaftlicher Mitarbeiter im nexus Institut, Gastforscher an der TU Berlin und als Technik- und Raumsoziologe Teil der Nachwuchsforschungsgruppe ELEVATE, die den sozialen und ökologischen Einsatz von Drohnen für zivile Zwecke interdisziplinär untersucht. E-Mail: sayman@nexusinstitut.de

Katharina Schilke ist wissenschaftliche Mitarbeiterin am Lehrstuhl für Rechtsinformatik an der Universität des Saarlandes. E-Mail: katharina.schilke@uni-saarland.de; legalinf.de/schilke

Claudia Schlüfter ist wissenschaftliche Mitarbeiterin in der Abteilung Neue Technologien am Fraunhofer-Institut für System- und Innovationsforschung ISI in Karlsruhe. Sie untersucht unter anderem die Wirkungen von Gesundheits-Apps auf die digitale Transformation des Gesundheitssystems. E-Mail: claudia.schluefter@isi.fraunhofer.de

Dr. Diana Schneider ist wissenschaftliche Mitarbeiterin der Abteilung Neue Technologien am Fraunhofer-Institut für System- und Innovationsforschung ISI in Karlsruhe. Sie forscht unter anderem zu ethischen und sozialen Implikationen von Digitalisierung und künstlicher Intelligenz im Sozial- und Gesundheitswesen. E-Mail: diana.schneider@isi.fraunhofer.de

Dr. Guido Scorza is a lawyer, journalist, and adjunct professor specializing in technology law and privacy, and a former member of the Board of the Italian Data Protection Authority (Garante). E-Mail: guido@guidoscorza.it

Marina Steenbergen ist wissenschaftliche Mitarbeiterin der Abteilung IPR am Fraunhofer-Institut für Sichere Informationstechnologie SIT und am Nationalen Forschungszentrum für angewandte Cybersicherheit ATHENE. E-Mail: marina.steenbergen@sit.fraunhofer.de

Dr. Markus Uhlmann ist Postdoc am Fachgebiet Soziologische Theorien an der Universität Kassel. Er ist Teil des Forschungsprojekts „Demokratie-X: Analyse der Tragfähigkeit einer privatheitsschonenden, fairen und gemeinwohlorientierten Plattform für Nachrichten“. E-Mail: markus.uhlmann@uni-kassel.de

Ulrich Waldmann ist wissenschaftlicher Mitarbeiter am Fraunhofer-Institut für Sichere Informationstechnologie SIT und am Nationalen Forschungszentrum für angewandte Cybersicherheit ATHENE. E-Mail: ulrich.waldmann@sit.fraunhofer.de

Dr. Klara-Aylin Wenten ist wissenschaftliche Mitarbeiterin am Fachgebiet für soziologische Theorie an der Universität Kassel und forscht zum Wechselverhältnis zwischen Emotionen, KI und Macht. Zuvor war sie an der Martin-Luther-Universität Halle-Wittenberg und an der Technischen Universität München tätig, wo sie zur Arbeitssteuerung im Rahmen des agilen Managements promoviert hat. Sie ist zudem Mitglied des Zentrums emanzipatorische Technikforschung. E-Mail: Klara-Aylin.Wenten@uni-kassel.de

Harald Zwingelberg ist Jurist am Unabhängigen Landeszentrum für Datenschutz, der Dienststelle der Landesbeauftragten für Datenschutz Schleswig-Holstein. Er leitet das Referat für Datenschutz-Forschungsprojekte. E-Mail: hzwingelberg@datenschutzzentrum.de