Verbraucherzentrale
Bundesverband

Gutachten „KI-Rechtsgrundlage“

Vorschlag für eine Rechtsgrundlage für die Verarbeitung personenbezogener Daten im Kontext der Entwicklung und des Einsatzes von KI

Gutachter: Peter Hense, David Wagner (beide Spirit Legal Rechtsanwälte)

08.12.2025

Im Auftrag von:
Verbraucherzentrale Bundesverband e.V.
Rudi-Dutschke-Straße 17, 10969 Berlin

Team Digitales & Medien
T +49 30 25800-0
digitales @vzbv.de
vzbv.de

Stand:
Dezember, 2025

Gefördert durch:

Bundesministerium
der Justiz und
für Verbraucherschutz

aufgrund eines Beschlusses
des Deutschen Bundestages

Inhalt

Executive Summary

Die Europäische Kommission hat am 19. November 2025 mit dem Digital Omnibus einen Vorschlag zur Änderung der DSGVO vorgelegt, der Rechtsgrundlagen für das Training von KI-Systemen mit personenbezogenen Daten schaffen soll. Der VZBV hat dieses Gutachten in Auftrag gegeben, um den Entwurf aus Perspektive der deutschen Verbraucher zu bewerten und einen Regelungsvorschlag zu formulieren, der die Bedürfnisse der Verbraucher mit den legitimen Interessen der KI-Entwicklung und dem Ziel europäischer Wettbewerbsfähigkeit in Einklang bringt.

Zentrale Befunde

Der Kommissionsentwurf weist gravierende Defizite auf:

Zur Modifikation des Art. 6 Abs. 1 lit. f DSGVO (Art. 88c DSGVO-E):

- Die Anknüpfung an die Definition des „AI System“ aus dem AI Act schafft eine uferlose Sonderrechtszone. Die Formulierung „in the context of the development and operation“ erfasst praktisch jede Datenverarbeitung, die auch nur in einem vagen Zusammenhang mit KI steht.
- Der systemische Konflikt zwischen dem verarbeitungsbezogenen Ansatz der DSGVO und dem systembezogenen Ansatz des AI Act macht eine präzise datenschutzrechtliche Bewertung ohne klare Systemgrenzen unmöglich.
- Wesentliche Schutzanforderungen – namentlich technische Opt-out-Signale, „reasonable expectations“ und der Nutzen für die „society at large“ – finden sich ausschließlich in unverbindlichen Erwägungsgründen.
- Fristenregelungen, Verfahren für Dritte ohne Nutzerkonto, Schutz gegen Umgehung von Betroffenenrechten durch De-Identifizierung und Regelungen für Open-Source-Modelle fehlen vollständig.

Zur Ausnahme vom Verarbeitungsverbot sensibler Daten (Art. 9 Abs. 2 lit. k DSGVO-E):

- Die Vorschrift verkehrt die Schutzlogik des Datenschutzrechts: Je mehr Daten verarbeitet werden, desto einfacher wird die Rechtfertigung.
- Sie widerspricht der EuGH-Rechtsprechung zur weiten Auslegung des Art. 9 Abs. 1 DSGVO und zur proaktiven Prüfpflicht bei sensiblen Daten (Russmedia).
- Sie höhlt den Erforderlichkeitsgrundsatz aus, indem sie die Verarbeitung nicht erforderlicher sensibler Daten legitimiert.

Empfehlungen

Wir empfehlen:

- Ersatzlose Streichung des Art. 9 Abs. 2 lit. k DSGVO-E.
- Einführung einer Sui-generis-Regelung (neuer Art. 6a DSGVO) mit folgenden kumulativen Voraussetzungen:
  - Subsidiaritätsnachweis: Verarbeitungszweck kann nicht durch synthetische oder anonymisierte Daten erreicht werden
  - Spezifische Risikoinformation vor Beginn der Verarbeitung
  - Unbedingtes Widerspruchsrecht mit angemessener Frist, auch für Dritte ohne Nutzerkonto
  - Technische Schutzmaßnahmen gegen Reproduktion und Identifizierbarkeit
  - Einwilligungserfordernis für Daten von Kindern

I. Einleitung

Am 19. November 2025 legte die Europäische Kommission mit dem Digital Omnibus (COM(2025) 837) einen Vorschlag zur Änderung der Datenschutz-Grundverordnung vor. Zentrales Element ist die Schaffung erleichterter Rechtsgrundlagen für die Entwicklung und den Betrieb von KI-Systemen mit personenbezogenen Daten. Der Entwurf erweitert die Ausnahmen vom Verarbeitungsverbot des Art. 9 DSGVO und modifiziert mit Art. 88c DSGVO-E die Interessenabwägung nach Art. 6 Abs. 1 lit. f DSGVO für die Verarbeitung personenbezogener Daten im Zusammenhang mit der Entwicklung und dem Betrieb von KI-Systemen.

Die Kommission begründet ihren Vorschlag mit der Notwendigkeit, europäischen KI-Entwicklern den Zugang zu Trainingsdaten zu erleichtern. Er trifft auf eine Rechtslage, die in zentralen Fragen ungeklärt ist: Behördliche Stellungnahmen lassen Spielräume, gerichtliche Entscheidungen divergieren.

Das vorliegende Gutachten unterzieht den Kommissionsentwurf einer kritischen Analyse aus verbraucherschutzrechtlicher Perspektive. Es untersucht zunächst, ob ein Bruch mit dem technikneutralen Regelungsansatz der DSGVO gerechtfertigt sein kann (II.), und analysiert sodann die Modifikation des Art. 6 Abs. 1 lit. f DSGVO durch Art. 88c DSGVO-E sowie die Erwägungsgründe 30 und 31 (III.). Im Anschluss würdigt es die vorgeschlagenen Änderungen des Art. 9 DSGVO (IV.). Abschließend unterbreitet es einen alternativen Regelungsvorschlag, der die Interessen der Verbraucherinnen und Verbraucher wirksam schützt, ohne die Entwicklung von KI-Systemen pauschal zu unterbinden (V.).

II. Systematischer Bruch mit dem technikneutralen Regelungsansatz

1. Das Prinzip der Technikneutralität

Die Datenschutz-Grundverordnung folgt dem Grundsatz der Technikneutralität.[1] Sie reguliert nicht einzelne Technologien, sondern die Verarbeitung personenbezogener Daten unabhängig von den eingesetzten Mitteln. Dieser Ansatz, der sich seit der Datenschutz-Richtlinie 95/46/EG bewährt hat, gewährleistet, dass das Schutzniveau nicht von der verwendeten Technologie abhängt und neue technische Entwicklungen nicht zu Schutzlücken führen.

2. Rechtfertigung für einen technologiespezifischen Ansatz?

Der Digital Omnibus bricht mit diesem Grundsatz, indem er eine spezifische Rechtsgrundlage für das Training von KI-Systemen schafft. Die Frage, ob ein solcher Bruch gerechtfertigt sein kann, verdient eine differenzierte Betrachtung.

Für eine technologiespezifische Regelung lässt sich zunächst die erhebliche wirtschaftliche und politische Bedeutung anführen, die generativen KI-Systemen gegenwärtig beigemessen wird. Ob diese Bedeutung eine Sonderbehandlung im Datenschutzrecht rechtfertigt, ist eine politische Wertungsfrage, keine rechtliche Notwendigkeit.

Gewichtiger erscheint das Argument der Rechtsunsicherheit. Die Stellungnahme 28/2024 des EDSA formuliert zwar Anforderungen an das KI-Training, lässt aber erhebliche Interpretationsspielräume.[2] In der mitgliedstaatlichen Rechtsprechung hat dies zu divergierenden Entscheidungen geführt: Das OLG Köln bejahte mit Urteil vom 23. Mai 2025 die Zulässigkeit des KI-Trainings durch Meta auf Grundlage des berechtigten Interesses nach Art. 6 Abs. 1 lit. f DSGVO und entwickelte für Art. 9 DSGVO eine „tätigkeitsbezogene Reduktion” des Verarbeitungsverbots.[3] Die Entscheidung ist in der Literatur auf deutliche Kritik gestoßen, weil die implementierten Schutzmechanismen sich als defizitär erwiesen und das Gericht keine substanzielle Risikoanalyse vornahm.[4] Demgegenüber deutete das OLG Schleswig in seinem Urteil vom 12. August 2025 an, dass „durchaus ein Verstoß, insbesondere gegen Art. 9 DSGVO” vorliegen dürfte, und wies darauf hin, dass nichtregistrierte Dritte keine realistische Möglichkeit hätten, von der Verarbeitung zu erfahren und ihr zu widersprechen.[5]

---

[1] ErwGr. 15 DSGVO.

[2] EDSA, Opinion 28/2024 v. 17.12.2024, Rn. 59 ff.

[3] OLG Köln, Urt. v. 23.05.2025 – 15 UKl 2/25.

[4] *Pesch*, AI hot mess – Meta at German courts and the troubling state of EU regulation, CR-online.de Blog v. 7.9.2025, abrufbar unter: https://www.cr-online.de/blog/2025/09/07/ai-hot-mess-meta-at-german-courts-and-the-troubling-state-of-eu-regulation/ (zuletzt abgerufen am 3.12.2025). Speziell mit Blick auf Art. 9 DSGVO Keber, RDV Sonderheft, 2025, 13 f.; Glocker, RDi 2025, 427 (431 f.).

[5] OLG Schleswig, Urt. v. 12.08.2025 – 6 UKl 3/25.

Diese Divergenz zwischen zwei Oberlandesgerichten in einer wirtschaftlich und gesellschaftlich bedeutsamen Frage ist unbefriedigend. Eine höchstrichterliche Klärung steht aus. Sie ist allerdings nicht zeitnah zu erwarten: Beide Verfahren betrafen den einstweiligen Rechtsschutz; ein Hauptsacheverfahren wurde bislang nicht eingeleitet. Hinzu kommt, dass der Weg zum EuGH lang ist – die durchschnittliche Verfahrensdauer bei Vorabentscheidungsersuchen beträgt derzeit durchschnittlich 17,2 Monate.[6]

Doch auch der Gesetzgeber kann die Unsicherheit nicht ohne Weiteres beseitigen. Gesetzliche Regelungen sind notwendigerweise generell-abstrakt gefasst; Rechtsklarheit im eigentlichen Sinne entsteht typischerweise erst durch behördliche Stellungnahmen und Rechtsprechung, die das abstrakte Recht auf konkrete Sachverhalte anwenden. Auch eine neue gesetzliche Regelung bleibt auf diese Konkretisierung angewiesen.

Was eine gesetzliche Regelung gleichwohl leisten kann, ist zweierlei: Sie kann materielle Schutzanforderungen verbindlich festschreiben und so einen Mindeststandard gewährleisten, den Behörden und Gerichte nicht unterschreiten dürfen. Und sie kann ein Level Playing Field schaffen, indem sie alle Unternehmen, die KI-Systeme mit Daten europäischer Betroffener trainieren wollen, denselben Anforderungen unterwirft – unabhängig davon, in welchem Mitgliedstaat sie ihren Sitz haben oder vor welchem Gericht sie verklagt werden.

3. Gang der Untersuchung

Die grundsätzliche Berechtigung einer technologiespezifischen Regelung ist von der Frage zu unterscheiden, ob der konkrete Kommissionsentwurf den aufgezeigten Anforderungen genügt. Eine gesetzliche Regelung, die einen verlässlichen Schutzstandard und ein Level Playing Field schaffen will, muss materielle Anforderungen präzise formulieren und darf sich nicht in unverbindlichen Absichtserklärungen erschöpfen.

Vor diesem Hintergrund analysiert das Gutachten zunächst den vorgeschlagenen Art. 88c DSGVO-E, der die Interessenabwägung nach Art. 6 Abs. 1 lit. f DSGVO für die Verarbeitung im Zusammenhang mit der Entwicklung und dem Betrieb von KI-Systemen modifiziert, sowie die Erwägungsgründe 30 und 31 des Kommissionsentwurfs, die Leitlinien für die Interessenabwägung nach Art. 6 Abs. 1 lit. f DSGVO formulieren (III.). Sodann würdigt es den vorgeschlagenen Art. 9 Abs. 2 lit. k DSGVO, der eine neue Ausnahme vom Verarbeitungsverbot für sensible Daten schaffen soll (IV.). Abschließend entwickelt das Gutachten einen alternativen Regelungsvorschlag, der die in der Rechtsprechung zutage getretenen Defizite adressiert (V.).

[6] Gerichtshof der Europäischen Union, Jahresbericht 2024 – Rechtsprechungsstatistiken des Gerichtshofs, S. 21.

III. Art. 88c DSGVO-E Die Erwägungsgründe 30 und 31

1. Grundsätzliche Anerkennung des berechtigten Interesses

ErwGr. 30 bringt eine Selbstverständlichkeit zum Ausdruck: Die Verarbeitung personenbezogener Daten für das KI-Training kann auf ein berechtigtes Interesse im Sinne des Art. 6 Abs. 1 lit. f DSGVO gestützt werden. Allerdings setzt dies voraus, dass das Interesse rechtmäßig verfolgt wird; ein Interesse, dessen Wahrnehmung gegen geltendes Recht verstößt, ist kein berechtigtes Interesse im Sinne der Vorschrift.[7] Auch der EDSA erkennt in seiner Stellungnahme 28/2024 an, dass Art. 6 Abs. 1 lit. f DSGVO grundsätzlich als Rechtsgrundlage für das KI-Training in Betracht kommt; als Beispiele nennt er die Entwicklung KI-basierter Gesprächsassistenten, die Erkennung betrügerischer Inhalte sowie die Verbesserung der Bedrohungserkennung in Informationssystemen.[8] Die Entwicklung von KI-Systemen stellt ein wirtschaftlich legitimes Ziel dar, das nicht per se mit den Interessen der Betroffenen konfligiert. Die grundsätzliche Anerkennung des KI-Trainings als berechtigtes Interesse ist daher nicht zu beanstanden.

2. Art. 88c DSGVO-E: Modifikation der Interessenabwägung nach Art. 6 Abs. 1 lit. f DSGVO

2.1 Regelungsgegenstand und systematische Einordnung

Art. 88c DSGVO-E modifiziert die Interessenabwägung nach Art. 6 Abs. 1 lit. f DSGVO für die Verarbeitung personenbezogener Daten „im Zusammenhang mit der Entwicklung und dem Betrieb“ von KI-Systemen und KI-Modellen im Sinne des Art. 3 Nr. 1 der KI-Verordnung. Die Vorschrift erfasst damit einen breiteren Anwendungsbereich als das in den Erwägungsgründen primär behandelte Web Scraping: Sie gilt für sämtliche Verarbeitungsvorgänge, die der Entwicklung und dem Betrieb von KI-Systemen dienen – einschließlich des Trainings, der Validierung, des Testens und des produktiven Einsatzes.

Anders als Art. 9 Abs. 2 lit. k DSGVO-E schafft Art. 88c DSGVO-E keine eigenständige Rechtsgrundlage. Die Vorschrift verbleibt im Rahmen des Art. 6 Abs. 1 lit. f DSGVO und modifiziert dessen Anwendung in dreifacher Hinsicht: Sie bestätigt, dass die Verarbeitung für die Entwicklung und den Betrieb von KI-Systemen grundsätzlich auf ein berechtigtes Interesse gestützt werden kann („where appropriate“). Sie normiert spezifische Schutzanforderungen, die über die allgemeine Interessenabwägung hinausgehen. Und sie enthält einen Vorbehalt zugunsten speziellerer Vorschriften des Unionsrechts oder nationalen Rechts, die eine Einwilligung verlangen.

---

[7] EuGH, Urt. v. 4.10.2024 – C-621/22 (KNLTB), Rn. 39.

[8] EDSA, Opinion 28/2024 v. 17.12.2024, Rn. 69.

Damit ergänzt Art. 88c lediglich die Interessenabwägung nach Art. 6 Abs. 1 lit. f DSGVO um bereichsspezifische Anforderungen. Die Rechtsgrundlage bleibt Art. 6 Abs. 1 lit. f DSGVO; Art. 88c DSGVO-E konkretisiert nur deren Voraussetzungen für den Anwendungsfall der KI-Entwicklung. Allerdings hängt die Tragweite des Art. 88c DSGVO-E von einem Begriff ab, der seinerseits erhebliche Fragen aufwirft: dem „AI System“ im Sinne des Art. 3 Nr. 1 der KI-Verordnung.

2.2 Verengung des Anwendungsbereichs durch Neudefinition „personenbezogener Daten“

Der Kommissionsentwurf enthält neben den KI-spezifischen Vorschriften eine Ergänzung der Definition „personenbezogener Daten" in Art. 4 Nr. 1 DSGVO. Die vorgeschlagene Neufassung ergänzt die bestehende Definition um den Passus, dass Informationen nicht allein deshalb zu personenbezogenen Daten für einen Verantwortlichen werden, weil ein nachfolgender Empfänger über Mittel verfügt, die vernünftigerweise zur Identifizierung der betroffenen Person eingesetzt werden könnten („Such information does not become personal data for that entity merely because a subsequent recipient has means reasonably likely to be used to identify the natural person“). Die Kommission gibt vor, lediglich die Rechtsprechung des EuGH zu kodifizieren. Tatsächlich stützt sie sich jedoch allein auf eine selektive Lesart der Rechtssache C-413/23P (EDPS v SRB), obwohl der EuGH die Sache zur weiteren Prüfung zurückverwies und mindestens sechs weitere Entscheidungen in eine andere Richtung weisen.[9]

Die vorgeschlagene Neufassung führt einen subjektiven Ansatz ein, der auf die Identifizierungsmittel des jeweiligen Verantwortlichen abstellt. Die Formulierung steht in direktem Widerspruch zur Rechtssache C-479/22 P (OC), wonach Informationen bereits für den ursprünglichen Verantwortlichen unter die DSGVO fallen, wenn Empfänger über solche Mittel verfügen.[10] Der Entwurf ist damit unvereinbar mit einem Threat-Modelling-Ansatz, der nach dem Stand der Technik auch Maßnahmen gegen Reidentifizierung durch Dritte verlangt.[11]

Die Wechselwirkungen mit den nachfolgend untersuchten Vorschriften sind erheblich: Fällt ein Verantwortlicher mangels Personenbezugs nicht unter die DSGVO, gelten gerade jene Vorschriften nicht mehr – insbesondere Art. 32 DSGVO –, die eine ordnungsgemäße Pseudonymisierung gewährleisten.[12] Der Kommissionsentwurf operiert damit an zwei Stellschrauben gleichzeitig: Er verengt den Anwendungsbereich der DSGVO durch Neudefinition eines Kernbegriffs und weitet zugleich den privilegierten Bereich durch einen uferlosen Systembegriff aus.

---

[9] noyb, Preliminary Analysis: EU Commission's proposal to amend the GDPR, 2025, S. 4 ff. m.w.N., abrufbar unter: https://noyb.eu/sites/default/files/2025-12/noyb%20Digital%20Omnibus%20Report%20V1.pdf (zuletzt abgerufen am 3.12.2025).

[10] EuGH, Urt. v. 16.1.2024 – C-479/22 P (OC), Rn. 62 f.; vgl. auch noyb, Preliminary Analysis: EU Commission's proposal to amend the GDPR, 2025, S. 7. m.w.N., abrufbar unter: https://noyb.eu/sites/default/files/2025-12/noyb%20Digital%20Omnibus%20Report%20V1.pdf (zuletzt abgerufen am 3.12.2025).

[11] Stalla-Bourdillon, Déjà vu in data protection law: the risks of rewriting what counts as personal data, SSRN 2025, S. 10, abrufbar unter: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5197121 (zuletzt abgerufen am 3.12.2025); vgl.auch noyb, Preliminary Analysis: EU Commission's proposal to amend the GDPR, 2025, S. 16 f. m.w.N., abrufbar unter: https://noyb.eu/sites/default/files/2025-12/noyb%20Digital%20Omnibus%20Report%20V1.pdf (zuletzt abgerufen am 3.12.2025).S. 7 f.

[12] noyb, Preliminary Analysis: EU Commission's proposal to amend the GDPR, 2025, S. 8., abrufbar unter: https://noyb.eu/sites/default/files/2025-12/noyb%20Digital%20Omnibus%20Report%20V1.pdf (zuletzt abgerufen am 3.12.2025).

2.3 „AI Systems“ als uferlose Sonderrechtszone

Die in Art. 88c DSGVO-E verwendete Formulierung „Where the processing of personal data is necessary for the interests of the controller in the context of the development and operation of an AI system as defined in Article 3, point (1), of Regulation (EU) 2024/1689 or an AI model“ wirkt auf den ersten Blick gesetzestechnisch und nüchtern durch den Verweis auf eine bestehende Definition unter dem AI Act. Sie zielt jedoch auf die Schaffung eines Sonderrechts für jede Form automatisierter Datenverarbeitung, die mit der Übernahme dieser dem Datenschutzrecht systemfremden Definition einhergeht.

Der Begriff des AI System in Art. 3 Abs. 1 AI Act ist Gegenstand zahlreicher Publikationen, die sich im Wesentlichen mit den qualitativen Anforderungen der Definition auseinandersetzen („Was macht ein System zu einem AI System?“). Die gesetzliche Definition ist sehr weit gefasst und erfasst auch Standardsoftware und einfache digitale Automatisierungen. Diese uferlose Weite der Definition wurde im Trilog durch Erwägungsgrund 12 des AI Acts zu begrenzen versucht. Erwägungsgründe können jedoch den klaren Wortlaut des Gesetzes weder einschränken noch eine Auslegung entgegen dem Wortlaut erzwingen.[13] Folglich ist Erwägungsgrund 12 funktionslos, und es verbleibt bei einer äußerst weiten Definition des AI Systems. Diese weicht ab von der OECD-Definition, von den Definitionen internationaler Standardisierungsorganisationen wie ISO/IEC und IEEE, auch von denen europäischer Standardisierungsorganisationen wie CEN und CENELEC, und auch von den Ethikleitlinien für vertrauenswürdige KI der High-Level Expert Group on Artificial Intelligence (AI HLEG), auf die das Gesetz in Erwägungsgrund 27 des AI Acts ausdrücklich Bezug nimmt.[14] Diese Differenzen der verschiedenen Definitionen werden in der Praxis jedoch weitgehend ignoriert, da sie sich bei der Qualifikation von AI Systems kaum auswirken. Es ist davon auszugehen, dass jedes zumindest teilweise automatisierte System in den Anwendungsbereich des AI Acts fällt.

Unter dem AI Act bleibt dies zunächst ohne Konsequenzen, da dessen Regime an die nachfolgende Einordnung von AI Systems in eine vordefinierte Risikokategorie anknüpft (Prohibited, High-Risk, Transparenzpflichten etc.).

Die Integration der weiten Definition von AI Systems in die DSGVO führt hingegen zu einem flächendeckenden Sonderregime, dem alle Verarbeitungen (Art. 4 Nr. 2 DSGVO) „im Kontext von Entwicklung und Betrieb“ von AI Systems und AI Models besondere Erleichterungen unterworfen werden sollen. Art. 88c DSGVO-E ist dabei durch die Formulierung „in the context of the development and operation“ sprachlich so weit gefasst, dass eine Verarbeitung, die auch nur in einem vagen Zusammenhang mit „AI“ steht, darunter subsumiert werden könnte, ganz gleich, ob dieser Zusammenhang tatsächlich besteht oder nicht. Derzeit gibt es praktisch keine Verarbeitung von personenbezogenen Daten, die nicht in irgendeiner Form „im Kontext“ von AI erfolgt, da definitionsgemäß der gesamte Lebenszyklus der Datenverarbeitung eines AI Systems, von der Konzeption über die Gewinnung von Trainingsdaten, das Training, Testing und die Validation bis hin zu Operation und Decommissioning, „im Kontext“ der Entwicklung und des Betriebs von AI Systems stattfindet. Die nachstehende Abbildung verdeutlicht das:[15]

---

[13] Vgl. bereits EuGH, Urt. v. 19.11.1998 – C-162/97, Rn. 54 sowie EuGH, Urt. v. 26.10.2023 – C-307/22 (DW/FT), Rn. 44 m.w.N.

[14] Vgl. zum Ganzen Hense/Mustać, AI Act kompakt – Compliance, Management und Use Cases für die Unternehmenspraxis, 2024, S. 4 ff.

[15] CEN/CLC/TR 18115:2024 – Data governance and quality for AI within the European context, Figure 13, Data Lifecycle Framework.

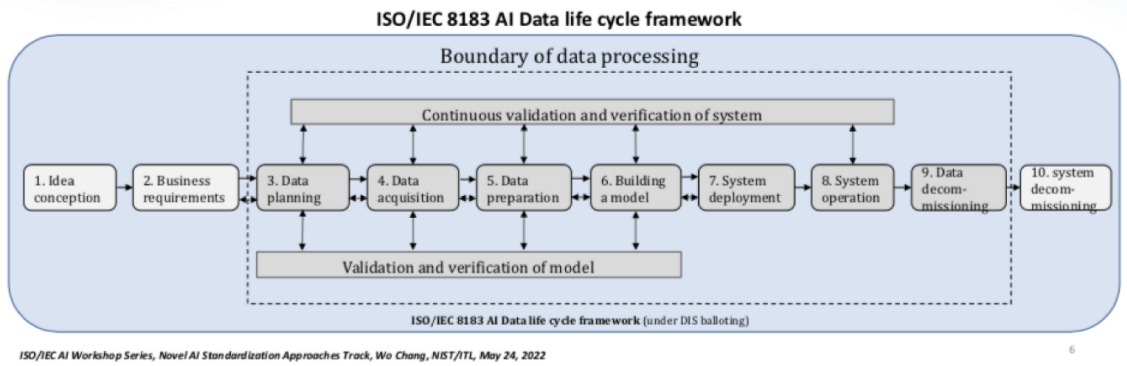

Abbildung 1: ISO/IEC 8183 AI Data life cycle framework (Quelle: CEN/CLC/TR 18115:2024)

Systemischer Konflikt zwischen DSGVO und AI Act

Ein grundlegendes systematisches Problem bei der Integration der Definition von AI Systems liegt darin, dass die DSGVO auf konkrete Verarbeitungen personenbezogener Daten und damit auf einen Anwendungskontext abstellt, nicht aber auf datenverarbeitende Systeme als solche. Das Produkthaftungs- und Produktsicherheitsrecht hingegen ist eher an Begriffen wie „Produkt“ oder „Device“ orientiert. „Systeme“ hingegen waren bisher kein relevanter Regelungsgegenstand des Unionsrechts, ihre Funktionalität und ihre Grenzen sind juristisch nicht einmal im Ansatz definiert. Während einzelne Verarbeitungsvorgänge im Rahmen einer Datenschutzfolgenabschätzung (DSFA) nach Art. 35 DSGVO risikobasiert bewertet werden können, ist dies bei Systemen nicht möglich, da Systeme abstrakt sind und offen: Sie existieren vom konkreten Datenfluss losgelöst. Systeme sind Infrastrukturen, Verarbeitungen deren konkrete Nutzung. Die Definition eines Systems ist statisch, weil sie das bewertete Objekt als feststehende Einheit zum Beispiel einer Konformitätsbewertung (vgl. Art. 43 AI Act) begreift, während Verarbeitungen als beweglicher Vorgang konzipiert sind. Der Systembegriff steht damit in direktem Konflikt zum risikobasierten Ansatz der DSGVO, weil er die Bewertungsebene von der variablen Handlung auf das statische Produkt verlagert, was das Kernkonzept der datenschutzrechtlichen Risikosteuerung unterläuft.

System-of-Systems-Problematik

Diese rechtliche Unsicherheit verschärft sich durch eine grundlegende technische Unschärfe des Gesetzesentwurfs: Der AI Act definiert nicht, wo ein AI System beginnt und wo es endet. Technische Grenzen und Interfaces komplexer KI-Architekturen bleiben rechtlich nicht adressiert, was im Kontext der Integration der Systemdefinition in die DSGVO zu einem Kontrollverlust bei der Rechtsanwendung führt, da die Instrumente der DSGVO sich asymmetrisch zum Anwendungsbereich des AI Acts auf AI Systems verhalten, was insbesondere europäische KMUs vor nicht zu bewältigende Herausforderungen stellt.

AI Systems sind keine kleinen monolithischen Einheiten, sondern komplexe Systems of Systems (SoS), also vernetzte Ensembles aus mehreren AI Subsystems und Computational Components, die jeweils eigene Funktionalitäten, Datenflüsse und Entscheidungslogiken implementieren. Ein System zur Kreditwürdigkeitsprüfung kann beispielsweise aus einem NLP-Modul für Dokumentenanalyse, einem klassifizierenden ML-Modell für Risikoscoring und einem erklärbaren AI-Modul (XAI) für die Begründungsgenerierung bestehen, wobei jede Komponente mit unterschiedlichen Daten, Algorithmen und Autonomiegraden operiert. Der AI Act benennt in Art. 3 Abs. 1 das AI System, als wäre es eine in der Realwelt aufgefundene, vordefinierte und

abgeschlossene Einheit. Das Gesetz übersieht dabei, dass in der Norm eine große Lücke klafft. Es fehlt an rechtlichen Kriterien für AI Models, AI Systems und andere Elemente als Komponenten innerhalb größerer Systeme. Diese Lücke kann derzeit nur durch die ingenieurswissenschaftliche Praxis geschlossen werden, insbesondere durch das System-of-Systems-Engineering (SoSE). Auch hierfür sind technische Standards unverzichtbar.[16]

Die nachfolgende Darstellung eines „Buy now, pay later“ (BNPL) Credit Scoring System auf Amazon AWS, das in praktisch Echtzeit Kreditentscheidungen im eCommerce trifft, verdeutlicht das Argument:[17]

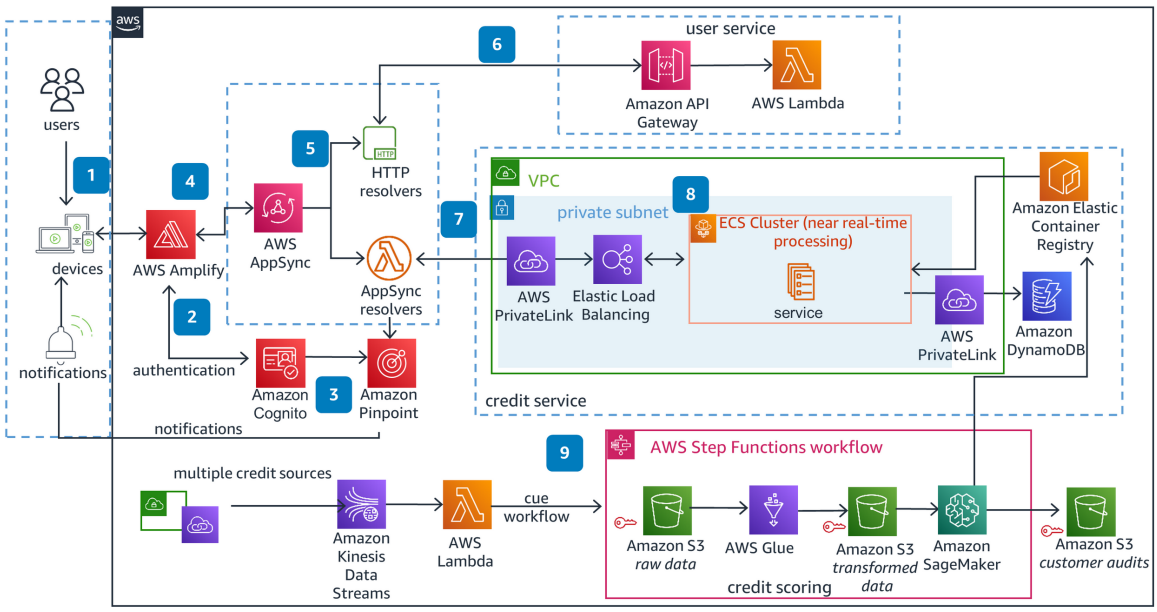

Abbildung 2: BNPL Credit Scoring System auf Amazon AWS (Quelle: AWS Solutions)

Die Darstellung der Systemarchitektur beschreibt ein verteiltes System, das die Schwierigkeiten bei der rechtlichen Abgrenzung von Verarbeitungstätigkeiten und Systemgrenzen unter gleichzeitiger Anwendung der DSGVO und des AI Act exemplarisch aufzeigt. Amazon SageMaker könnte als eigenständiges AI System gelten, da auf der Plattform Modelle trainiert und deployed werden. Gleichzeitig ist es eine Komponente des übergeordneten Kredit-Scoring-Systems, das wiederum auch als AI System qualifiziert werden könnte, weil es letztlich autonome Kreditentscheidungen trifft. Die AWS Lambda-Funktionen, die Glue-ETL-Jobs und die Kinesis-Streams fungieren als Komponenten, die je nach funktionalem Kontext entweder Teil eines größeren AI Systems sind oder eigenständige Datenverarbeitungssysteme ohne AI-Bezug darstellen.

Die Formulierung „in the context of the development and the operation“ erfasst nicht nur die dargestellten Kernkomponenten, sondern auch die in der obigen Systemarchitektur nur angedeuteten vor- und nachgelagerten Verarbeitungen: Webanalyse-Tools, die Nutzerverhalten auf Kreditvergleichsportalen tracken, E-Commerce-Checkout-Prozesse, die Zahlungsdaten und

[16] Hense/Mustać, SoS AI: An AI System of Systems Approach for EU AI Act Conformity, 2025.

[17] Guidance for Credit Decisioning Using Primary & Alternative Data on AWS, abrufbar unter: https://aws.amazon.com/de/solutions/guidance/credit-decisioning-using-primary-and-alternative-data-on-aws/ (zuletzt abgerufen am 3.12.2025).

Kaufhistorien generieren, sowie CRM-Systeme, die Kundeninteraktionen und Lead-Informationen verarbeiten. Diese Verarbeitungen würden ebenfalls alle unter den privilegierenden Anwendungsbereich von Art. 88c DSGVO-E fallen, sobald die erhobenen Daten auch nur potenziell für Kreditentscheidungen relevant sein könnten, was allein der Entscheidung des Verantwortlichen unterliegt. Art. 88c DSGVO-E würde damit eine Erleichterung nahezu jeder Datenverarbeitung bewirken und folglich eine generelle Absenkung des bislang hohen Schutzniveaus der DSGVO herbeiführen, das Art. 7 und 8 EU-Grundrechte-Charta vorschreiben und das zum über Jahrzehnte hart erkämpften europäischen Acquis gehört.

Praktische Unanwendbarkeit der Datenschutzgrundsätze

Ohne präzise technische und regulatorische Grenzziehung ist eine datenschutzrechtliche Bewertung dessen, was ein AI System sein soll, unmöglich. Um die gesetzgeberische Definition eines AI Systems mit den Mitteln der Datenschutzgrundsätze aus Art. 5 DSGVO auszufüllen, insbesondere im Rahmen der für Art. 88c DSGVO-E erforderlichen Interessenabwägung die betroffenen Interessen adäquat zu definieren und die verpflichtenden Informationen über die Verarbeitung inklusive der Interessenabwägung transparent machen zu können, müssten Verantwortliche zuvor die funktionalen Grenzen, also welche Systemkomponente welche Teilaufgabe erfüllt, rechtssicher dokumentieren. Ferner müssten sie Dateninterfaces definieren, also welche Daten zwischen Subsystemen fließen und ob es sich dabei um personenbezogene Daten handelt, sodann die Entscheidungsverantwortung bestimmen, also welches Subsystem welche autonome Entscheidung trifft, sowie eine Lebenszykluszuordnung vornehmen, also welche Phase von Entwicklung, Training, Testing, Validation, Operation bis Decommissioning welches Subsystem betrifft. Erst durch diese technische Präzisierung wird das gesetzlich vorausgesetzte AI System überhaupt identifizierbar. Bleiben die Systemgrenzen jedoch unklar oder sind sie streitig, ist auch die Verarbeitung im Sinne von Art. 4 Nr. 2 DSGVO nicht mehr eindeutig bestimmbar.

Fällt der Datenfluss zwischen zwei AI Subsystems verschiedener Anbieter noch unter „Entwicklung und Betrieb“ eines Gesamtsystems? Sind Data Annotation oder ein Data Preprocessing Module Teil des durch Art. 88c DSGVO-E begünstigten AI Systems oder eigenständige, nicht von der Norm privilegierte Verarbeitungen? Ist zeitlich vorgelagertes Webscraping Teil des Systementwicklungskontextes der Norm oder aufgrund der zeitlich nachgelagerten datenschutzrechtlichen Zweckbestimmung doch nicht rechtlich begünstigt? Die DSGVO erfordert für jede konkrete Verarbeitung volle Rechtmäßigkeit, eine gesonderte Rechtsgrundlage, eine klare Zweckbindung, Datensicherheit und Risikobewertung. Ohne klare Systemgrenzen wird aber das risikobasierte Modell der DSGVO technisch und rechtlich operativ unanwendbar. Dieses operative Risiko verbleibt aufgrund der Beweislast für Systemkonformität (AI Act) und Rechtmäßigkeit der Verarbeitung (DSGVO) beim Anbieter und Verantwortlichen von AI Systems: Die gesetzgeberische Bezugnahme auf AI Systems führt also zu einer erheblichen Rechtsunsicherheit, nicht zu einer Erleichterung.

2.4 Zuordnung der Schutzanforderungen: Normtext und Erwägungsgründe

Der Kommissionsentwurf verteilt die materiellen Anforderungen an das KI-Training auf zwei Regelungsebenen: den verbindlichen Normtext des Art. 88c DSGVO-E und die unverbindlichen Erwägungsgründe 30 und 31. Die nachfolgende Übersicht ordnet die wesentlichen Schutzanforderungen ihrer jeweiligen Regelungsebene zu:

Tabelle 1: Schutzanforderungen für KI: Normtext, Erwägungsgründe, Regelungslücken

Schutzanforderung	Normtext (Art. 88 c)	Erwägungsgründe	Regelungslücke
Unbedingtes Widerspruchsrecht	✓	✓	–
Erweiterte Transparenz („enhanced transparency“)	✓ (unbestimmt)	✓ (unbestimmt)	Konkretisierung fehlt
Datenminimierung bei Quellenauswahl	✓	–	–
Datenminimierung beim Training und Testen	✓	–	–
Schutz gegen Offenlegung residualer Daten / Regurgitation / Data Leakage	✓	✓	–
Technische Opt-out-Signale	–	✓	–
„Reasonable expectations“ als Abwägungskriterium	–	✓	–
„Society at large“ als Abwägungskriterium	–	✓	–
Bias-Erkennung und -Beseitigung	–	✓	–
Sichere und genaue Outputs	–	✓	–
Privacy-preserving Techniken (z.B. Differential Privacy)	–	✓ (unbestimmt)	Konkretisierung fehlt
Kinderschutz	✓ (allgemeiner Hinweis)	–	Konkretisierung fehlt

Schutzanforderung	Normtext (Art. 88 c)	Erwägungsgründe	Regelungslücke
Fristenregelungen für Information und Widerspruch	–	–	✓
Schutz gegen Umgehung von Betroffenenrechten durch De-Identifizierung	–	–	✓
Regelungen für Open-Source-Modelle	–	–	✓

2.5 Bewertung der im Normtext verankerten Anforderungen

Die bisherige Rechtslage überlässt die Zulässigkeit des KI-Trainings der allgemeinen Interessenabwägung nach Art. 6 Abs. 1 lit. f DSGVO, ohne spezifische Schutzanforderungen zu normieren. Art. 88c DSGVO-E schafft demgegenüber bereichsspezifische Anforderungen, die über die allgemeine Interessenabwägung hinausgehen und verbindliche Rechtspflichten begründen, deren Verletzung sanktionierbar ist.

Das unbedingte Widerspruchsrecht („unconditional right to object“) geht über das allgemeine Widerspruchsrecht des Art. 21 Abs. 1 DSGVO hinaus, das eine Begründung aus der „besonderen Situation“ des Betroffenen verlangt. Die Normierung entspricht der Forderung des EDSA und trägt dem Umstand Rechnung, dass die Verarbeitung für das KI-Training irreversible Folgen haben kann.

Die Anforderung der Datenminimierung bei der „selection of sources“ sowie beim „training and testing“ stellen ausschnittsweise klar, dass der allgemeine Grundsatz des Art. 5 Abs. 1 lit. c DSGVO auch bei dem spezifischen Anwendungsfall der KI-Entwicklung anzuwenden ist. Die Pflicht zum Schutz gegen die Offenlegung residualer Daten adressiert das Risiko, dass Trainingsdaten in Modellausgaben reproduziert werden.

Gleichwohl bleiben die im Normtext verankerten Anforderungen in wesentlichen Punkten unbestimmt. Der Begriff der „enhanced transparency“ wird nicht konkretisiert; die Vorschrift benennt weder die zu erteilenden Informationen noch die Form ihrer Bereitstellung. Die Minimierungsanforderungen geben Verantwortlichen keine hinreichende Orientierung, welche konkreten Maßnahmen zu ergreifen sind. Der Kinderschutz erschöpft sich in dem allgemeinen Hinweis, dass die Interessen des Betroffenen „in particular where the data subject is a child“ zu berücksichtigen sind, ohne spezifische Schutzanforderungen zu normieren.

3. Unverbindliche Verortung wesentlicher Anforderungen in den Erwägungsgründen

Die Übersicht in Abschnitt IV. 2.3 zeigt, dass der Kommissionsentwurf wesentliche Schutzanforderungen ausschließlich in den Erwägungsgründen verortet. Dies betrifft namentlich die technischen Opt-out-Signale, die Kriterien der „reasonable expectations“ und des Nutzens für die „society at large“ sowie sämtliche Konkretisierungen der im Normtext nur abstrakt formulierten Anforderungen.

Nach ständiger Rechtsprechung des EuGH dienen Erwägungsgründe der Auslegung, begründen aber keine selbständigen Rechtspflichten.18 Die Unterscheidung zwischen verbindlichem Normtext und erläuternden Erwägungsgründen ist dogmatisch zwingend; andernfalls würde die Grenze zwischen Gesetz und Gesetzesbegründung verwischt. Die Kommission lagert damit wesentliche Schutzanforderungen in einen rechtlich unverbindlichen Bereich aus. Das Ergebnis ist ein regulatorischer Kunstgriff, der den Anschein von Schutz erweckt, ohne effektiven Schutz zu gewährleisten, und der Gefahr läuft, als bloße Symbolgesetzgebung zu verpuffen.

Soweit der Kommissionsentwurf materielle Anforderungen ausschließlich in den Erwägungsgründen verortet, untergräbt dies deren Verbindlichkeit. Verantwortliche können sich darauf berufen, dass die in ErwGr. 31 formulierten Anforderungen keine eigenständigen Rechtspflichten begründen. Aufsichtsbehörden und Gerichte können zwar auf die Erwägungsgründe als Auslegungshilfe zurückgreifen; ein Verstoß gegen die dort formulierten Anforderungen begründet aber keinen unmittelbaren Rechtsverstoß.

Besonders problematisch ist die vollständige Regelungslücke in Bezug auf zentrale Schutzanforderungen, die auch von den Erwägungsgründen nicht aufgegriffen werden: Fristenregelungen für Information und effektive Widerspruchsverfahren für Dritte ohne Nutzerkonto, Schutz gegen Umgehung von Betroffenenrechten durch De-Identifizierung und Regelungen für Open-Source-Modelle finden sich weder im Normtext noch in den Erwägungsgründen. Diese Lücken kann auch eine extensive Auslegung der bestehenden Vorschriften auf Basis der vorgesehenen Erwägungsgründe nicht ohne Weiteres schließen.

4. Inhaltliche Defizite der Erwägungsgründe

Auch inhaltlich bleiben die Erwägungsgründe hinter dem zurück, was für einen wirksamen Schutz der Betroffenen erforderlich wäre. Der Kommissionsentwurf greift zwar einzelne Anforderungen des EDSA auf, formuliert sie aber so unbestimmt, dass sie in der Praxis leerzulaufen drohen.

4.1 „Society at large” als Abwägungskriterium

ErwGr. 31 stellt darauf ab, ob „the interest pursued by the controller is beneficial for the data subject and society at large”. Diese Formulierung wirft systematische Bedenken auf.

Wenn Benefits für die Gesellschaft insgesamt bei der datenschutzrechtlichen Interessenabwägung eine Rolle spielen sollen, müssten konsequenterweise auch die gesellschaftlichen Risiken einbezogen werden. Das wäre jedoch systembrüchig: Die Risiken von KI-Systemen adressiert nach

[18] Vgl. bereits EuGH, Urt. v. 19.11.1998 – C-162/97, Rn. 54.

der Konzeption des europäischen Gesetzgebers die KI-Verordnung, nicht das Datenschutzrecht. Umgekehrt erscheint es verfehlt, auf abstrakte Hoffnungen für die Gesellschaft abzustellen, ohne konkrete Belege für einen tatsächlichen Nutzen zu fordern. Der Kommissionsentwurf erlaubt es Verantwortlichen, sich auf vage Versprechungen zu berufen, während die korrespondierenden Risiken in einem anderen Regelungsregime behandelt werden. Diese Aufspaltung führt dazu, dass die Interessenabwägung nur die positiven Aspekte berücksichtigt und die negativen ausblendet.

Das Kriterium erweist sich zudem als Deckmantel für kommerzielle Interessen. Denn KI-Entwickler verwenden personenbezogene Daten kostenlos, um Produkte zu entwickeln, die sie anschließend auf dem Markt verkaufen; Betroffenen Personen bleibt dabei regelmäßig keine echte Wahlmöglichkeit.[19] Die pauschale Legitimationsformel „society at large“ löst diese Asymmetrie zwischen Datengebern und Datennehmern nicht auf, sondern verschärft sie – indem sie den privaten Gewinn hinter dem Gemeinwohlbegriff verschwinden lässt.

Bei Foundation-Modellen verschärft sich das Problem zusätzlich. Ob ein solches Modell der Gesellschaft tatsächlich nützt, hängt von der konkreten Einbindung in Anwendungen ab, die zum Zeitpunkt des Trainings noch gar nicht feststehen. Ein und dasselbe Modell kann im Kundendienst unterstützen oder zur Grundlage für Entscheidungen mit erheblichen Auswirkungen auf Betroffene werden – etwa bei Einstellungsentscheidungen oder in Verwaltungsverfahren.[20] Ohne Kenntnis der konkreten Einsatzzwecke lässt sich ein Nutzen für die „society at large” schlicht nicht beurteilen.

4.2 Zementierung bestehender Marktverhältnisse

ErwGr. 31 stellt ferner auf „reasonable expectations of the data subject based on their relationship with the controller” ab. Diese Formulierung bevorzugt Anbieter, die bereits über eine eigene Nutzerbasis verfügen: Eine „reasonable expectation” kann nur dort entstehen, wo bereits eine Beziehung zum Verantwortlichen besteht. Große Plattformen wie Meta, Google oder Microsoft, die über Milliarden von Nutzerkonten verfügen, können auf dieser Grundlage argumentieren, ihre Nutzer hätten mit einer Verwendung ihrer Daten für KI-Training rechnen müssen.[21] Newcomer ohne bestehende Nutzerbasis können sich auf dieses Argument nicht stützen.

Der Entwurf enthält zudem keine Klarstellung, dass eine solche Erwartung nicht rückwirkend für vergangene Interaktionen begründet werden kann. Dieses Defizit ermöglicht es Verantwortlichen, nachträglich eine „reasonable expectation” zu konstruieren, obwohl Nutzer zum Zeitpunkt ihrer Interaktion mit dem Dienst keinerlei Anlass hatten, eine Verwendung ihrer Daten für KI-Training zu erwarten.

4.3 Unzureichende Transparenzanforderungen

ErwGr. 31 fordert „enhanced transparency”, ohne zu konkretisieren, was darunter zu verstehen ist. Der EDSA formuliert demgegenüber präzise Anforderungen: Alle relevanten Informationen sind in

---

[19] Mustać, Data Altruism by Default: An Alternative to Consent for Personal Data Processing in Machine Learning, CEP Project 2024, S. 1, abrufbar unter: https://cep-project.org/publications/data-altruism-by-default-an-alternative-to-consent-for-personal-data-processing-in-machine-learning-winner/ (zuletzt abgerufen am 3.12.2025).

[20] Pesch, AI hot mess – Meta at German courts and the troubling state of EU regulation, CR-online.de Blog v. 7.9.2025, abrufbar unter: https://www.cr-online.de/blog/2025/09/07/ai-hot-mess-meta-at-german-courts-and-the-troubling-state-of-eu-regulation/ (zuletzt abgerufen am 3.12.2025).

[21] Kritisch gegenüber der Entstehung solcher Erwartungen mit Blick auf die Vergangenheit etwa Brink, RDV Sonderheft, 2025, 11, 12.

zugänglicher, verständlicher und benutzerfreundlicher Form bereitzustellen; alternative Informationsformen wie Medienkampagnen, grafische Visualisierungen, Transparenzkennzeichnungen und Modellkarten können die Transparenz erhöhen; die Informationen sollen über die Anforderungen der Art. 13 und 14 DSGVO hinausgehen, mit besonderer Berücksichtigung von Kindern und schutzbedürftigen Personen.

Die unbestimmte Formulierung des Kommissionsentwurfs gibt Verantwortlichen keine hinreichende Orientierung und Betroffenen keinen durchsetzbaren Anspruch. Die gegenwärtige Praxis zeigt, wohin das führt: Nutzer müssen Informationen „Stück für Stück” zusammensuchen, indem sie Datenschutzrichtlinien durchforsten, die auf Foren verlinken, die wiederum auf FAQs verweisen.[22] Transparenzanforderungen verkommen so zu einer Verkaufsnarrative – einer Geschichte, die den Nutzern erzählt wird, basierend auf Erwartungen darüber, was sie verstehen können und was sie hören wollen.[23]

4.4 Technische Schutzmechanismen ohne Nutzerzugang

ErwGr. 31 fordert die Beachtung von „technical indications embedded in a service limiting the use of data for AI development by third parties”. Gemeint sind technische Signale wie robots.txt-Dateien, Meta-Tags oder HTTP-Header, mit denen Webseitenbetreiber Crawlern das Auslesen ihrer Inhalte untersagen können. Diese Mechanismen stammen aus der Frühzeit des Internets und wurden ursprünglich entwickelt, um Suchmaschinen die Indexierung bestimmter Seiten zu verbieten.

ErwGr. 31 überträgt dieses Konzept auf das KI-Training; Art. 88c DSGVO-E erwähnt technische Opt-out-Signale hingegen nicht: Verantwortliche sollen technische Opt-out-Signale respektieren. Das klingt nach einem pragmatischen Ansatz – doch er leidet an einem grundlegenden Konstruktionsfehler. Die genannten Schutzmechanismen stehen nur Webseitenbetreibern zur Verfügung, nicht aber den Nutzern geschlossener Plattformen. Wer Inhalte auf Facebook, Instagram oder LinkedIn einstellt, kann keine robots.txt-Datei konfigurieren; diese Kontrolle liegt ausschließlich beim Plattformbetreiber.[24] Der Entwurf setzt implizit voraus, dass Nutzer Zugang zu technischen Schutzmechanismen haben – eine Annahme, die bei proprietären Systemen regelmäßig nicht zutrifft.

Das führt zu einer paradoxen Situation: ErwGr. 31 fordert die Beachtung technischer Opt-out-Signale, gibt aber gerade denjenigen, deren Daten verarbeitet werden, keine Möglichkeit, solche Signale zu setzen. Der vermeintliche Schutzmechanismus schützt Plattformbetreiber, nicht Betroffene.

---

[22] Mustać, Data Altruism by Default: An Alternative to Consent for Personal Data Processing in Machine Learning, CEP Project 2024, S. 4, abrufbar unter: https://cep-project.org/publications/data-altruism-by-default-an-alternative-to-consent-for-personal-data-processing-in-machine-learning-winner/ (zuletzt abgerufen am 3.12.2025).

[23] Mustać, Data Altruism by Default: An Alternative to Consent for Personal Data Processing in Machine Learning, CEP Project 2024, S. 2, abrufbar unter: https://cep-project.org/publications/data-altruism-by-default-an-alternative-to-consent-for-personal-data-processing-in-machine-learning-winner/ (zuletzt abgerufen am 3.12.2025). Mustać, Data Altruism by Default, 2024, S. 2.

[24] Mustać, Data Altruism by Default: An Alternative to Consent for Personal Data Processing in Machine Learning, CEP Project 2024, S. 13, abrufbar unter: https://cep-project.org/publications/data-altruism-by-default-an-alternative-to-consent-for-personal-data-processing-in-machine-learning-winner/ (zuletzt abgerufen am 3.12.2025).

4.5 Defizitäre Opt-out-Regelungen

ErwGr. 31 verlangt „an unconditional right to object”. Art. 88c DSGVO-E übernimmt diese Anforderung in den verbindlichen Normtext. Die Formulierung klingt vielversprechend, greift aber zu kurz. Ein Widerspruchsrecht, das nur die künftige Verarbeitung betrifft, verfehlt die Realität des KI-Trainings: Sind die Daten einmal in ein Modell eingeflossen, prägen sie dessen Verhalten dauerhaft. Das Widerspruchsrecht (Art. 21 DSGVO) kann daher nicht isoliert vom Löschungsanspruch (Art. 17 DSGVO) gedacht werden. Ein wirksamer Schutz setzt voraus, dass Betroffene nicht nur der künftigen Verwendung widersprechen, sondern auch die Entfernung bereits verarbeiteter Daten verlangen können. Der Kommissionsentwurf blendet diese Verzahnung aus.

Schon das Widerspruchsrecht selbst wirft erhebliche Umsetzungsfragen auf. Ein wirksames Opt-out muss für Nutzer des Dienstes ebenso einfach zugänglich sein wie für Dritte, deren Daten andere Nutzer auf der Plattform bereitgestellt haben. Bei sozialen Netzwerken gelangen regelmäßig auch personenbezogene Daten Dritter in den Trainingsdatensatz – etwa wenn Nutzer Fotos hochladen, auf denen andere Personen abgebildet sind. Der Kommissionsentwurf adressiert diese Konstellation nicht.

Das Opt-out muss zwei Anforderungen genügen: Der Verantwortliche muss es verständlich erklären, und die betroffene Person muss es ohne Aufwand ausüben können. Die gegenwärtige Praxis verfehlt beides – sie verlangt von Nutzern, Informationen aktiv zu suchen, aus verschiedenen Quellen zusammenzutragen und sich durch mehrstufige Prozesse zu arbeiten.[25]

Es fehlen zudem Fristenregelungen. Wann muss das Training angekündigt werden? Ab wann wird ein Widerspruch wirksam? Ankündigungen kurz vor Trainingsbeginn lassen Betroffenen faktisch keine Möglichkeit, rechtzeitig zu reagieren.[26]

Die Löschungsproblematik verschärft diese Defizite. Das Opt-out darf nicht dadurch umgangen werden, dass der Anbieter die Daten de-identifiziert und anschließend erklärt, den Betroffenen im Datensatz nicht mehr identifizieren zu können. Diese Konstellation ist praktisch bedeutsam: Verantwortliche können Trainingsdaten so transformieren, dass eine direkte Zuordnung zu einzelnen Personen nicht mehr möglich ist, während die Daten gleichwohl das Modellverhalten prägen. Beruft sich der Verantwortliche darauf, den Betroffenen im transformierten Datensatz nicht mehr identifizieren zu können, laufen sowohl Widerspruchsrecht als auch Löschungsanspruch faktisch leer. Der Kommissionsentwurf enthält keine Vorkehrungen gegen diese Umgehungsstrategie. Erforderlich wäre eine Dokumentationspflicht, die den Zusammenhang zwischen Ursprungsdaten und Widerspruch auch nach der Transformation nachvollziehbar hält.

Völlig ungeklärt bleibt schließlich, wie ein Opt-out Betroffene schützen soll, wenn KI-Modelle als Open Source veröffentlicht werden. Bei frei verfügbaren Modellen kann jedermann das trainierte Modell herunterladen und nutzen. Ein nachträglicher Widerspruch oder Löschungsantrag gegenüber dem ursprünglichen Verantwortlichen ist dann faktisch wirkungslos – das Modell existiert bereits in unzähligen Kopien, auf die der Betroffene keinen Zugriff hat.

---

[25] Mustać, Data Altruism by Default: An Alternative to Consent for Personal Data Processing in Machine Learning, CEP Project 2024, S. 3 f., abrufbar unter: https://cep-project.org/publications/data-altruism-by-default-an-alternative-to-consent-for-personal-data-processing-in-machine-learning-winner/ (zuletzt abgerufen am 3.12.2025).

[26] Pesch, AI hot mess – Meta at German courts and the troubling state of EU regulation, CR-online.de Blog v. 7.9.2025, abrufbar unter: https://www.cr-online.de/blog/2025/09/07/ai-hot-mess-meta-at-german-courts-and-the-troubling-state-of-eu-regulation/ (zuletzt abgerufen am 3.12.2025).

4.6 Unbestimmte Minimierungsanforderungen

Was „appropriate measures to effectively minimise risks” konkret bedeutet, bleibt offen. Art. 88c DSGVO-E verlangt zwar Datenminimierung bei der „selection of sources“ sowie beim „training and testing“, ohne jedoch konkrete Maßnahmen zu benennen. Der EDSA fordert deutlich präzisere Maßnahmen: Dokumentation aller Trainingsdatenquellen; Prüfung der Trainingsdaten auf statistische Verzerrungen; Ausschluss nicht autorisierter Inhalte; Entfernung unnötiger personenbezogener Daten; Einsatz methodischer Entscheidungen, die die Identifizierbarkeit signifikant reduzieren; Implementierung datenschutzerhaltender Techniken wie Differential Privacy.[27]

Diese Konkretisierungen ermöglichen eine effektive Kontrolle durch Aufsichtsbehörden und Betroffene. Der Kommissionsentwurf belässt es bei der unbestimmten Formulierung „appropriate measures” und gibt damit weder Verantwortlichen noch Aufsichtsbehörden hinreichende Orientierung.

Der EDPS warnt in seiner Guidance for Risk Management of Artificial Intelligence Systems vom 11. November 2025 vor der wahllosen Sammlung und Speicherung personenbezogener Daten: Die Ansammlung großer Datenmengen ohne klare Kriterien könne zur Akkumulation von Informationen führen, die für die Ziele des KI-Systems nicht notwendig sind und dem Grundsatz der Datenminimierung widersprechen.[28]

4.7 Fehlender Kinderschutz

Der Entwurf lässt Daten von Kindern weitgehend unberücksichtigt. Art. 88c DSGVO-E enthält lediglich den allgemeinen Hinweis, dass die Interessen des Betroffenen „in particular where the data subject is a child“ zu berücksichtigen sind, ohne konkrete Schutzanforderungen zu normieren. Das wiegt besonders schwer, weil auch Volljährige und Institutionen Daten von Minderjährigen veröffentlichen können, die dann in den Trainingsdatensatz eingehen – einschließlich sensibler Daten wie Fotografien.[29] Der EDSA fordert ausdrücklich, dass bei öffentlich zugänglichen Informationen Kinder und schutzbedürftige Personen besonders berücksichtigt werden müssen. Kinder können die Tragweite der Datenverarbeitung typischerweise nicht überblicken und bedürfen daher eines besonderen Schutzes.[30] Der Kommissionsentwurf enthält keine entsprechenden konkreten Vorgaben.

5. Zwischenergebnis

Art. 88c DSGVO-E verankert wesentliche Schutzanforderungen im verbindlichen Normtext: ein unbedingtes Widerspruchsrecht, erweiterte Transparenz, Datenminimierung bei Quellenauswahl sowie beim Training und Testen und den Schutz gegen die Offenlegung residualer Daten.

---

[27] EDSA, Opinion 28/2024, Rn. 51 f., 58, 105.

[28] EDPS, Guidance for Risk Management of Artificial Intelligence systems, 11.11.2025, S. 30 f.

[29] Vgl. OLG Köln, Urt. v. 23.05.2025 – 15 UKl 2/25, Rn. 87.

[30] Zur besonderen Schutzbedürftigkeit von Kindern, die auch im Kontext von KI-Training in besonderer Weise Berücksichtigung finden muss OLG Köln, Urt. v. 23.05.2025 – 15 UKl 2/25, Rn. 87.

Gegenüber der bisherigen Rechtslage, die keine spezifischen Anforderungen an das KI-Training normiert, ergänzt Art. 88c DSGVO-E die Interessenabwägung nach Art. 6 Abs. 1 lit. f DSGVO um bereichsspezifische Schutzanforderungen.

Gleichwohl bleibt der Regelungsrahmen unvollständig. Die im Normtext verankerten Anforderungen sind unbestimmt formuliert und bedürfen der Konkretisierung. Wesentliche Gesichtspunkte – namentlich die Kriterien der „reasonable expectations“ und des Nutzens für die „society at large“, die technischen Opt-out-Signale sowie sämtliche Konkretisierungen der abstrakten Normvorgaben – finden sich ausschließlich in den unverbindlichen Erwägungsgründen 30 und 31.

Diese greifen die vom EDSA formulierten Schutzanforderungen nur selektiv auf und bleiben in zentralen Punkten hinter ihnen zurück. Die unbestimmten Formulierungen geben Verantwortlichen keine Orientierung und Betroffenen keine durchsetzbaren Ansprüche.

Fristenregelungen, Verfahren für Dritte ohne Nutzerkonto, Schutz gegen Umgehung von Betroffenenrechten durch De-Identifizierung und Regelungen für Open-Source-Modelle fehlen sowohl im Normtext als auch in den Erwägungsgründen. Die in der Rechtsprechung zutage getretenen Defizite – insbesondere die Unwirksamkeit der Schutzmechanismen und die Schutzlosstellung nichtregistrierter Dritter – adressiert der Kommissionsentwurf auch mit der Modifikation des Art. 6 Abs. 1 lit. f DSGVO durch Art. 88c DSGVO-E nicht.

IV. Art. 9 Abs. 2 lit. k und Abs. 5 DSGVO-E

1. Tatbestandsvoraussetzungen und Regelungskonzeption

Der neue Art. 9 Abs. 2 lit. k DSGVO-E schafft eine Ausnahme vom Verarbeitungsverbot für sensible Daten im Kontext der Entwicklung und des Betriebs von KI-Systemen und KI-Modellen. Die Ausnahme greift nur unter den Bedingungen des ebenfalls neu eingefügten Art. 9 Abs. 5 DSGVO-E.

Die Regelung folgt einer Stufenlogik: Der Verantwortliche muss zunächst geeignete technische und organisatorische Maßnahmen implementieren, um die Erhebung und sonstige Verarbeitung sensibler Daten zu vermeiden. Identifiziert er gleichwohl sensible Daten in den Trainings-, Test- oder Validierungsdatensätzen oder im KI-System selbst, muss er diese entfernen. Erfordert die Entfernung unverhältnismäßigen Aufwand, muss er die Daten zumindest wirksam davor schützen, zur Erzeugung von Outputs verwendet, offengelegt oder Dritten anderweitig zugänglich gemacht zu werden.

ErwGr. 33 begrenzt den Anwendungsbereich der Ausnahme ausdrücklich auf „residuale” sensible Daten – also solche, die trotz Vermeidungsmaßnahmen im Datensatz verbleiben, obwohl sie für den Verarbeitungszweck nicht notwendig sind. Werden sensible Daten hingegen gezielt verarbeitet, bleibt der Verantwortliche auf die bestehenden Ausnahmetatbestände des Art. 9 Abs. 2 lit. a–j DSGVO verwiesen.

2. Dogmatische Einwände

2.1 Verkehrung der Schutzlogik

Das Datenschutzrecht ist seit jeher Individualrecht; sein Schutzniveau bemisst sich nach den Risiken für den Einzelnen, nicht nach den Interessen des Verantwortlichen.[31] Art. 24 DSGVO normiert diesen Grundsatz: Je umfangreicher die Verarbeitung, desto höher die Anforderungen an die Schutzmaßnahmen.[32] Der Kommissionsentwurf verkehrt diese Logik ins Gegenteil: Er privilegiert gerade die massenhafte Verarbeitung unstrukturierter Datensätze, bei der eine Einzelfallprüfung praktisch unmöglich ist. Je mehr Daten verarbeitet werden, desto einfacher wird die Rechtfertigung.

Dass der Verantwortliche es in der Hand haben soll, durch die Potenzierung der Eingriffe das Schutzniveau des Einzelnen herabzusenken, erscheint kontra-intuitiv. Die praktische Unmöglichkeit, bei der Massendatenerhebung jede Einzelkategorie zu prüfen, darf nicht dazu führen, dass der grundrechtliche Schutz insgesamt ausgehöhlt wird.

2.2 Widerspruch zur EuGH-Rechtsprechung

Die Rechtsprechung des EuGH zur weiten Auslegung des Art. 9 Abs. 1 DSGVO steht dem Kommissionsentwurf entgegen. Der Gerichtshof stellte klar, dass es gerade nicht darauf ankommt, ob der Verantwortliche mit dem Ziel handelt, Informationen zu erhalten, die unter eine der in Art. 9 DSGVO genannten besonderen Kategorien fallen.[33] Enthält ein Datensatz sowohl sensible als auch nicht-sensible Daten, die zum Zeitpunkt der Erhebung nicht voneinander getrennt werden können, ist die Verarbeitung dieses Datensatzes als im Sinne von Art. 9 Abs. 1 DSGVO untersagt anzusehen, sofern er mindestens ein sensibles Datum umfasst und keine der Ausnahmen des Art. 9 Abs. 2 DSGVO greift.[34]

Diese Rechtsprechung ist dogmatisch zwingend: Andernfalls könnte der Schutz sensibler Daten durch bloße Vermischung mit nicht-sensiblen Daten umgangen werden. Der Kommissionsentwurf setzt sich über diese Rechtsprechung hinweg, indem er gerade die untrennbare Vermischung zum Anknüpfungspunkt einer Privilegierung macht.

2.3 Schutzlosstellung des Einzelnen

Die vorgeschlagene Ausnahme nimmt dem Betroffenen den Rechtsverstoß als Anknüpfungspunkt für Rechtsbehelfe. Greift die Ausnahme, liegt kein Verstoß gegen Art. 9 Abs. 1 DSGVO vor – obwohl sensible Daten verarbeitet werden. Der Kontrollverlust über die eigenen Daten, den der EuGH als ersatzfähigen immateriellen Schaden anerkannt hat,[35] bleibt ohne Kompensation, weil es an einem Rechtsverstoß fehlt. Die Regelung immunisiert den Verantwortlichen, ohne dem Betroffenen einen Ausgleich zu gewähren.

---

[31] Vgl. Art. 1 Abs. 2 DSGVO.

[32] Vgl. Art. 24 Abs. 1 Satz 1 DSGVO.

[33] EuGH, Urt. v. 4.7.2023 – C-252/21 (Meta/Bundeskartellamt), Rn. 69; EuGH, Urt. v. 1.8.2022 – C-184/20 (Vyriausioji tarnybinės etikos komisija), Rn. 127.

[34] EuGH, Urt. v. 4.7.2023 – C-252/21 (Meta/Bundeskartellamt), Rn. 89

[35] EuGH, Urt. v. 14.12.2023 – C-340/21 (Natsionalna agentsia za prihodite), Rn. 80 ff.

3. Adaption der OLG Köln-Logik

Der Kommissionsentwurf adaptiert die Kernlogik des Urteils des OLG Köln vom 23. Mai 2025.[36] Das Gericht entwickelte dort eine „tätigkeitsbezogene Reduktion” des Art. 9 DSGVO: Das Verarbeitungsverbot für sensible Daten bedürfe bei massenhafter automatisierter Datenverarbeitung einer „Aktivierung” durch einen Antrag des betroffenen Dritten auf Herausnahme seiner Daten. Ohne einen solchen Antrag bestehe keine Pflicht, sensible Daten vorab zu identifizieren und auszusondern.

Der Kommissionsentwurf folgt dieser Logik in zweifacher Hinsicht: Erstens erlaubt Art. 9 Abs. 2 lit. k DSGVO-E die Verarbeitung sensibler Daten, die „trotz” Vermeidungsmaßnahmen im Datensatz verbleiben – der Verantwortliche schuldet also keine Erfolgsgarantie. Zweitens verlagert Art. 9 Abs. 5 DSGVO-E die Schutzmechanismen auf die Output-Ebene: Können sensible Daten nicht entfernt werden, genügt es, sie vor der Verwendung in Outputs zu schützen. ErwGr. 33 flankiert diese Konstruktion, indem er ausdrücklich nur „residuale” sensible Daten erfasst und damit die Verantwortung für eine effektive Vorabfilterung relativiert.

Das OLG Köln stützte seine Konstruktion auf eine Parallele zur EuGH-Rechtsprechung zum Suchmaschinenbetreiber Google.[37] Diese Parallele trägt nicht.[38] Der EuGH begründete sein Urteil mit den Besonderheiten einer Suchmaschine, bei der Ergebnisse automatisch erzeugt und nur im Fall eines Auslistungsantrags manuell geprüft werden. Er argumentierte mit dem begrenzten Verantwortungsbereich der Suchmaschinenbetreiber, deren beschränkten Befugnissen und Einflussmöglichkeiten sowie dem Umstand, dass die Listung in den Suchergebnissen erst auf Hinweis Betroffener zutage tritt.[39]

Beim KI-Training ist die Situation fundamental anders: Eine spätere Auslistung ist nicht möglich, da die Trainingsdaten unwiderruflich in das KI-Modell eingeflossen sind.[40] Eine Suchanfrage, die den Betroffenen ermöglicht, die Verarbeitung ihrer Daten zu erkennen und einen Auslistungsantrag zu stellen, steht bei einem KI-Trainingsdatensatz nicht zur Verfügung.[41] Zudem sind Nutzer sozialer Netzwerke – anders als Webseitenbetreiber – nicht für die Einbringung ihrer Daten in das KI-Modell verantwortlich; sie können nicht Verantwortlicher im Sinne des Art. 4 Nr. 7 DSGVO sein, wenn sie mit der Datenverarbeitung nicht einmal rechnen.

Das Urteil des EuGH in der Rechtssache Russmedia vom 2. Dezember 2025 stützt diese Kritik.[42] Der Gerichtshof entschied, dass der Betreiber eines Online-Marktplatzes als (gemeinsam) Verantwortlicher im Sinne von Art. 26 DSGVO vor der Veröffentlichung von Anzeigen verpflichtet, mittels geeigneter technischer und organisatorischer Maßnahmen Anzeigen mit sensiblen Daten zu identifizieren, die Identität des inserierenden Nutzers zu prüfen und die Veröffentlichung zu

---

[36] OLG Köln, Urt. v. 23.5.2025 – 15 UKl 2/25.

[37] OLG Köln, Urt. v. 23.5.2025 – 15 UKl 2/25 Rn. 102, unter Verweis auf EuGH, Urt. v. 24.09.2019 – C-136/17 (GC ua/CNIL), Rn. 45 ff.

[38] [38] *Pesch*, AI hot mess – Meta at German courts and the troubling state of EU regulation, CR-online.de Blog v. 7.9.2025, abrufbar unter: https://www.cr-online.de/blog/2025/09/07/ai-hot-mess-meta-at-german-courts-and-the-troubling-state-of-eu-regulation/ (zuletzt abgerufen am 3.12.2025). Speziell mit Blick auf Art. 9 DSGVO Keber, RDV Sonderheft, 2025, 13 f.; Glocker, RDi 2025, 427 (431 f.).

[39] EuGH, Urt. v. 24.09.2019 – C-136/17 (GC ua/CNIL), Rn. 45 ff.

[40] Pesch, AI hot mess – Meta at German courts and the troubling state of EU regulation, CR-online.de Blog v. 7.9.2025, abrufbar unter: https://www.cr-online.de/blog/2025/09/07/ai-hot-mess-meta-at-german-courts-and-the-troubling-state-of-eu-regulation/ (zuletzt abgerufen am 3.12.2025).

[41] Pesch, AI hot mess – Meta at German courts and the troubling state of EU regulation, CR-online.de Blog v. 7.9.2025, abrufbar unter: https://www.cr-online.de/blog/2025/09/07/ai-hot-mess-meta-at-german-courts-and-the-troubling-state-of-eu-regulation/ (zuletzt abgerufen am 3.12.2025).

[42] EuGH, Urt. v. 2.12.2025 – C-492/23 (Russmedia).

verweigern, wenn keine ausdrückliche Einwilligung der betroffenen Person nachgewiesen wird.[43] Der EuGH stellte zudem klar, dass sich ein solcher Betreiber in Bezug auf Verstöße gegen die DSGVO nicht auf die Haftungsprivilegierung der Art. 12 bis 15 der E-Commerce-Richtlinie berufen kann; diese Bestimmungen könnten nicht in die Regelung der DSGVO eingreifen.[44]

Die Entscheidung betrifft zwar Online-Marktplätze und nicht das KI-Training. Gleichwohl lässt sich aus ihr ein allgemeines Prinzip ableiten: Bei der Verarbeitung sensibler Daten verlangt der EuGH proaktive Schutzmaßnahmen vor der Verarbeitung. Der Verantwortliche muss Inhalte mit sensiblen Daten identifizieren und darf sich nicht darauf verlassen, dass Betroffene nachträglich die Löschung beantragen. Diese Pflicht zur Vorabprüfung steht in einem Spannungsverhältnis zur OLG Köln-Konstruktion, die der Kommissionsentwurf adaptiert. Ob die Russmedia-Grundsätze auf das KI-Training übertragbar sind, wird die weitere Rechtsprechung klären müssen. Das Risiko, dass der Kommissionsentwurf an dieser Judikatur scheitert, ist jedenfalls nicht von der Hand zu weisen.

4. Aufweichung des Erforderlichkeitsgrundsatzes

Art. 9 Abs. 5 DSGVO-E erlaubt die Verarbeitung sensibler Daten, die für die Entwicklung des KI-Systems gerade nicht erforderlich sind. Die Vorschrift setzt tatbestandlich voraus, dass die sensiblen Daten trotz Vermeidungsmaßnahmen im Datensatz verbleiben, obwohl sie für den Verarbeitungszweck nicht notwendig sind. Damit durchbricht der Entwurf den Erforderlichkeitsgrundsatz, der zu den tragenden Säulen des Datenschutzrechts gehört.[45]

Diese Durchbrechung bei Art. 9 DSGVO wirft die Frage auf, ob sie auf Art. 6 Abs. 1 lit. f DSGVO ausstrahlt. Die Logik liegt nahe: Sensible Daten unterliegen nach Art. 9 Abs. 1 DSGVO einem grundsätzlichen Verarbeitungsverbot und genießen damit ein höheres Schutzniveau als nicht-sensible Daten. Wenn der Gesetzgeber ausgerechnet bei sensiblen Daten auf das Erforderlichkeitskriterium verzichtet, liegt der Schluss a maiore ad minus auf der Hand – erst recht dürfte dann bei nicht-sensiblen Daten die Erforderlichkeit der Verarbeitung für das berechtigte Interesse großzügiger ausgelegt werden.

ErwGr. 30 und 31, die Art. 6 Abs. 1 lit. f DSGVO betreffen, deuten in diese Richtung: Sie erkennen das KI-Training pauschal als berechtigtes Interesse an, ohne die Erforderlichkeit der konkreten Datenverarbeitung zu thematisieren. Damit steht der Kommissionsentwurf in einem Spannungsverhältnis zur aktuellen EuGH-Rechtsprechung. In der Rechtssache Mousse hat der Gerichtshof Art. 6 Abs. 1 lit. f DSGVO ausdrücklich in Verbindung mit Art. 5 Abs. 1 lit. c DSGVO ausgelegt und betont, dass die Erforderlichkeit nur zu bejahen ist, wenn das berechtigte Interesse nicht in zumutbarer Weise ebenso wirksam mit anderen Mitteln erreicht werden kann, die weniger stark in die Grundrechte der betroffenen Personen eingreifen.[46] Die Datenverarbeitung dürfe nicht lediglich von Nutzen sein, sondern müsse objektiv unerlässlich sein.[47]

Der Kommissionsentwurf höhlt diesen Grundsatz aus, indem er die Verarbeitung nicht erforderlicher Daten legitimiert – bei Art. 9 ausdrücklich, bei Art. 6 durch beredtes Schweigen.

---

[43] EuGH, Urt. v. 2.12.2025 – C-492/23, Rn. 97, 102, 105 f.

[44] EuGH, Urt. v. 2.12.2025 – C-492/23, Rn. 131, 135 f.

[45] Vgl. Art. 5 Abs. 1 lit. c DSGVO (Datenminimierung).

[46] EuGH, Urt. v. 9.1.2025 – C-394/23 (Mousse), Rn. 48.

[47] EuGH, Urt. v. 9.1.2025 – C-394/23 (Mousse), Rn. 33 f.

5. Zwischenergebnis

Art. 9 Abs. 2 lit. k DSGVO-E durchbricht den Schutz sensibler Daten zugunsten der KI-Entwicklung. Die Regelung verkehrt die Schutzlogik des Datenschutzrechts, widerspricht der EuGH-Rechtsprechung zur weiten Auslegung des Art. 9 Abs. 1 DSGVO und stellt den Einzelnen schutzlos. Der Kommissionsentwurf übernimmt die umstrittene Konstruktion des OLG Köln, wonach das Verarbeitungsverbot für sensible Daten bei massenhafter Datenverarbeitung erst durch einen Antrag des Betroffenen „aktiviert” werden muss. Das Russmedia-Urteil des EuGH bestätigt, dass Plattformbetreiber proaktive Prüfpflichten bei sensiblen Daten treffen; auch die Haftungsprivilegierung der E-Commerce-Richtlinie verdrängen datenschutzrechtliche Pflichten nicht. Schließlich wirft die Aufgabe des Erforderlichkeitsgrundsatzes bei Art. 9 Abs. 5 DSGVO-E die Frage auf, ob diese Aufweichung auf die Interessenabwägung nach Art. 6 Abs. 1 lit. f DSGVO ausstrahlt – eine Entwicklung, die der EuGH in der Mousse-Entscheidung durch seine enge Auslegung der Erforderlichkeit gerade zu verhindern sucht.

V. Regelungsempfehlungen

1. Die Notwendigkeit einer eigenständigen Regelung

Das Training von Systemen der Künstlichen Intelligenz stellt einen eigenständigen Regelungsfall dar, der sich von den herkömmlichen Verarbeitungssituationen des Art. 6 Abs. 1 lit. f DSGVO grundlegend unterscheidet. Die Verarbeitung großer Datenmengen für das KI-Training weist Besonderheiten auf, die eine Herauslösung aus der Dogmatik des Art. 6 Abs. 1 lit. f und eine eigenständige Regelung erfordern: Die Daten fließen unwiderruflich in das Modell ein, eine nachträgliche Löschung ist technisch nicht oder nur eingeschränkt möglich, und die Trainingsdaten können in Modellausgaben reproduziert werden.

Wir empfehlen daher eine Sui-generis-Regelung, die diesen eigenständigen Regelungsfall durch konkrete, kumulative Voraussetzungen adressiert.

2. Streichung des Art. 9 Abs. 2 lit. k DSGVO-E

Wir empfehlen die ersatzlose Streichung des vorgeschlagenen Art. 9 Abs. 2 lit. k DSGVO-E. Die Vorschrift verkennt die Schutzlogik des Art. 9 DSGVO, steht im Widerspruch zur Rechtsprechung des EuGH[48] und stellt den Einzelnen schutzlos. Verantwortliche sind auf die bestehenden Ausnahmetatbestände des Art. 9 Abs. 2 lit. a–j DSGVO zu verweisen.

[48] Vgl. EuGH, Urt. v. 2.12.2025 – C-492/23 (Russmedia), Rn. 97, 102, 105 f.

3. Normvorschlag: Sui-generis-Regelung für KI-Training

3.1 Vorschlag für einen neuen Art. 6a DSGVO

Artikel 6a

Verarbeitung für die Entwicklung und Verwendung von Systemen der Künstlichen Intelligenz

(1) Die Verarbeitung personenbezogener Daten für die Entwicklung und Verwendung von Systemen der Künstlichen Intelligenz ist abweichend von Artikel 6 Absatz 1 Buchstabe f rechtmäßig, wenn die Bedingungen der Absätze 2 bis 6 erfüllt sind.

(2) Die Verarbeitung personenbezogener Daten für die Entwicklung und Verwendung von Systemen der Künstlichen Intelligenz ist nur rechtmäßig, wenn der Verantwortliche nachweist, dass die festgelegten, eindeutigen Verarbeitungszwecke nicht durch nach dem Stand der Technik verfügbare Technologien zum Schutz personenbezogener Daten, wie synthetische oder anonymisierte Daten erreicht werden können.

(3) Der Verantwortliche informiert die betroffene Person mit angemessener Frist von nicht weniger als drei Monaten vor Beginn der Verarbeitung über die spezifischen Risiken der Verarbeitung für das Training von Systemen nach Absatz 1, insbesondere darüber, dass

a) personenbezogene Daten in diese Systeme einfließen können,

b) eine nachträgliche Löschung aus diesen Systemen technisch nicht oder nur eingeschränkt möglich ist, und

c) die für die Entwicklung dieser Systeme genutzten Daten in den Systemausgaben reproduziert werden können.

(4) Der Verantwortliche räumt der betroffenen Person ein voraussetzungsloses Widerspruchsrecht ein. Der Widerspruch kann jederzeit vor Beginn der Verarbeitung nach Absatz 1 erklärt werden und bedarf keiner Begründung.

(5) Der Verantwortliche hat nachzuweisen, dass er ausschließlich Daten verarbeitet von betroffenen Personen, die nach Absatz 3 informiert wurden und die ihr Widerspruchsrecht nach Absatz 4 effektiv ausüben konnten.

(6) Der Verantwortliche trifft unter Berücksichtigung des Stands der Technik und der Implementierungskosten geeignete technische und organisatorische Maßnahmen, um

a) die Reproduktion personenbezogener Daten einer nach Absatz 3 informierten betroffenen Person, die ihr Widerspruchsrecht nicht ausgeübt hat, in den Systemausgaben von Systemen nach Absatz 1 zu verhindern, und

b) die Identifizierbarkeit betroffener Personen in Systemen nach Absatz 1 zu minimieren.

Satz 1 lässt Verarbeitungen personenbezogener Daten unberührt, die in anderer Weise rechtmäßig verarbeitet werden.

(7) Die Verarbeitung personenbezogener Daten von Kindern unter 18 Jahren ist nach Absatz 1 nur mit ausdrücklicher Einwilligung der Träger elterlicher Verantwortung zulässig. Mit Erreichen der Volljährigkeit haben betroffene Personen ein voraussetzungsloses Widerspruchsrecht gegen Verarbeitungen nach Absatz 1, das sie binnen 12 Monaten nach Erreichen der Volljährigkeit gegenüber dem Verantwortlichen ausüben können.

(6) Dieser Artikel lässt Artikel 9 unberührt.

3.2 Vorschlag für einen neuen Erwägungsgrund 30a DSGVO

(30a) Die Entwicklung und Verwendung von Systemen der Künstlichen Intelligenz stellt einen eigenständigen Regelungsfall dar, der sich von den herkömmlichen Verarbeitungssituationen des Artikel 6 Absatz 1 Buchstabe f grundlegend unterscheidet. Die Daten fließen typischerweise unwiderruflich in das Modell ein, eine nachträgliche Löschung ist technisch nicht oder nur eingeschränkt möglich, und die Trainingsdaten können in Modellausgaben reproduziert werden. Artikel 6a trägt diesen Besonderheiten Rechnung, indem er den Regelungsfall aus der allgemeinen Dogmatik des Artikel 6 Absatz 1 Buchstabe f herauslöst und konkrete Voraussetzungen normiert.

Die Verarbeitung personenbezogener Daten für die Entwicklung und Verwendung von KI-Systemen ist nur zulässig, wenn der Verarbeitungszweck nicht anderweitig erreicht werden kann. Die bloße Nützlichkeit personenbezogener Daten genügt nicht; der Verantwortliche hat nachzuweisen, dass die festgelegten Verarbeitungszwecke nicht durch nach dem Stand der Technik verfügbare Technologien zum Schutz personenbezogener Daten erreicht werden können.

Die betroffene Person muss die spezifischen Folgen der Verarbeitung für die Entwicklung und Verwendung von KI-Systemen kennen, um eine informierte Entscheidung über die Ausübung des Widerspruchsrechts treffen zu können. Die Risikoinformation hat daher insbesondere darauf hinzuweisen, dass personenbezogene Daten unwiderruflich in das Modell einfließen können, dass eine nachträgliche Löschung technisch nicht oder nur eingeschränkt möglich ist und dass die Trainingsdaten in Modellausgaben reproduziert werden können.

Das Widerspruchsrecht muss vor Beginn der Verarbeitung effektiv ausgeübt werden können. Berechtigte Erwartungen der betroffenen Personen können nur entstehen, wenn die Verarbeitung für die Entwicklung und Verwendung von KI-Systemen vertraglich vereinbart wurde oder der Verantwortliche die Verarbeitung mit angemessener Frist angekündigt hat. Eine rückwirkende Konstruktion berechtigter Erwartungen, um bestehende Datenbestände nachträglich zu legitimieren, ist unzulässig. Das Widerspruchsrecht steht auch Personen zu, die nicht Nutzer des Dienstes sind. Der Verantwortliche hat entsprechende Verfahren bereitzustellen, die einen niedrigschwelligen Zugang ohne Registrierungspflicht ermöglichen.

Das Widerspruchsrecht darf nicht dadurch unterlaufen werden, dass der Verantwortliche die Trainingsdaten de-identifiziert und sich anschließend darauf beruft, die betroffene Person nicht mehr identifizieren zu können. Der Verantwortliche hat durch geeignete Dokumentation sicherzustellen, dass der Zusammenhang zwischen Ursprungsdaten und Widerspruch auch nach einer Transformation der Daten nachvollziehbar bleibt.

Die technischen Maßnahmen zum Schutz vor Reproduktion personenbezogener Daten und zur Minimierung der Identifizierbarkeit sind technologieneutral zu verstehen. Zu den gegenwärtig verfügbaren Techniken zählen Differential Privacy, Federated Learning und Machine Unlearning. In bestimmten Fällen kann die Ausgabe personenbezogener Daten durch das Modell zulässig sein, etwa wenn die betroffene Person eingewilligt hat oder die Daten offensichtlich selbst öffentlich gemacht hat.

Die Veröffentlichung von KI-Modellen als Open Source ist grundsätzlich wünschenswert, geht aber mit besonderen Herausforderungen einher, da ein nachträglicher Widerspruch gegenüber dem ursprünglichen Verantwortlichen faktisch wirkungslos ist, wenn das Modell bereits in zahlreichen Kopien existiert. Der Verantwortliche hat dies bei der Entscheidung über eine Open-Source-Veröffentlichung

zu berücksichtigen und gegebenenfalls geeignete Vorkehrungen zu treffen, etwa durch Dokumentation der Herkunft der Trainingsdaten oder vertragliche Bindung nachgelagerter Nutzer.

Kinder verdienen bei der Verarbeitung ihrer personenbezogenen Daten besonderen Schutz, da sie die Tragweite und die Risiken der Datenverarbeitung für die Entwicklung und Verwendung von KI-Systemen typischerweise nicht überblicken können. Die Verarbeitung ihrer Daten bedarf daher der ausdrücklichen Einwilligung der Träger der elterlichen Verantwortung. Dies gilt auch dann, wenn die Daten durch Dritte in den Datensatz eingebracht wurden. Mit Erreichen der Volljährigkeit sollen die Betroffenen selbst entscheiden können, ob ihre Daten weiterhin verwendet werden dürfen; ihnen steht daher ein voraussetzungsloses Widerspruchsrecht zu, das sie innerhalb einer angemessenen Frist nach Erreichen der Volljährigkeit ausüben können.

VI. Abbildungsverzeichnis

VII. Tabellenverzeichnis