KI und Datenschutz: EDSA-Stellungnahme zur Verarbeitung personenbezogener Daten in KI-Modellen

Am 18. Dezember 2024 hat der Europäische Datenschutzausschuss (EDSA) eine neue Stellungnahme zu bestimmten datenschutzrechtlichen Aspekten bei der Verarbeitung von personenbezogenen Daten im Zusammenhang mit KI-Modellen veröffentlicht. Die Stellungnahme enthält richtungsweisende Leitlinien für Entwickler und Nutzer von KI-Modellen und adressiert zentrale Themen wie die Frage, wann ein KI-Modell personenbezogene Daten enthält, die Bewertung des berechtigten Interesses als Rechtsgrundlage für die Entwicklung und Nutzung von KI-Modellen und Rechtsfolgen bei unrechtmäßiger Datenverarbeitung bei Entwicklung eines KI-Modells.

Mit der Stellungnahme 28/2024 vom 17. Dezember 2024 reagiert der EDSA auf eine Anfrage der irischen Aufsichtsbehörde (DPC), die den Datenschutzausschuss gebeten hatte, eine Stellungnahme nach Art. 64 Abs. 2 DSGVO mit allgemeiner Geltung zu erlassen. Im Zusammenhang mit der rasanten Entwicklung von Technologien zur Künstlichen Intelligenz („KI“) kam es in letzter Zeit immer häufiger zu datenschutzrechtlichen Fragestellungen – nicht zuletzt ausgelöst durch den Umstand, dass die im August 2024 in Kraft getretene KI-Verordnung vorsieht, dass die in ihr festgelegten Rechte und Pflichten neben der DSGVO Anwendung finden (Art. 2 Abs. 7 KI-VO).

Die Stellungnahme des EDSA bezieht sich auf KI-Modelle im Sinne der KI-VO, die das Ergebnis eines Trainings mit personenbezogenen Daten sind. Sie setzt klare Vorgaben, die ein Gleichgewicht zwischen Innovation und Datenschutz schaffen sollen.

Die wichtigsten Punkte der Stellungnahme haben wir hier für Sie zusammengefasst:

Wann ist ein KI-Modell „anonym“?

Für alle Informationen, die sich auf eine identifizierte oder identifizierbare Person beziehen, gelten die Grundsätze der DSGVO. Das gilt auch für pseudonymisierte Daten.

Liegen Daten vor, die sich nicht auf eine identifizierte oder identifizierbare Person beziehen (wie beispielsweise reine Firmendaten, z. B. Geschäftszahlen), müssen die Anforderungen der DSGVO nicht eingehalten werden. Das gleiche gilt bei anonymisierten Daten, die eine Verknüpfung zu einer bestimmten Person nicht (mehr) zulassen.

Sind KI-Modelle trainiert, beinhalten sie selbst i.d.R. keine personenbezogenen Daten in Klarform mehr: Sie beinhalten abstrahierte Datenrepräsentationen, die in Form von Vektoren oder numerischen Merkmalen im Modell vorliegen. Auch während der Anwendung arbeiten KI-Modelle meist auf einer höheren Abstraktionsebene: die Modelle analysieren die eingegeben Daten, indem sie Muster erkennen und Wahrscheinlichkeiten berechnen, um Output zu generieren. Dieser Umstand kann den Anschein erwecken, dass die in dem Model, insbesondere den großen Sprachmodellen (LLM) enthaltenen Datensätze anonym und nicht personenbezogen sind, da sie rein aus der Beziehung und den Zusammenhängen zwischen Token (Wortbausteinen) entstehen.

Allerdings können Daten, die in ein KI-Modell zu Trainingszwecken einfließen, personenbezogen sein und ebenso die Ergebnisse nach einem bestimmten Prompt. Wenn ich ein KI-Modelle nutze, kann ich mithin durch das Prompting personenbezogene Daten „generieren“. Sind dann nicht die in einem solchen KI-Modell auch in Ruhe vorliegenden Informationen zwangsläufig auch personenbezogene Daten?

Diesem Punkt widmet sich die erste Frage der DPC, die wissen möchte, ob das anwendungsfähige KI-Modell („final“), das mit personenbezogenen Daten trainiert wurde, den Anforderungen der Definition personenbezogener Daten gem. Art. 4 Abs. 1 DSGVO entspricht („Is the final AI Model, which has been trained using personal data, in all cases, considered not to meet the definition of personal data (as set out in Article 4(1) GDPR)?“).

Die Hamburgische Beauftragte für den Datenschutz und die Informationsfreiheit (HmbBfDI) hat im vergangenen Jahr in einem Diskussionspapier hierzu bereits die These zur Diskussion gestellt, dass die bloße Speicherung eines LLMs keine Verarbeitung personenbezogener Daten i. S. d. Art. 4 Nr. 2 DSGVO darstelle. Denn in LLMs selbst seien Tokens, Gewichte und Vektoren, aber keine personenbezogenen Daten gespeichert. Wir berichteten.

Der EDSA sieht das anders und stellt zunächst klar: Bei KI-Modellen, die speziell darauf ausgelegt sind, personenbezogene Daten derjenigen bereitzustellen, deren Daten für das Training genutzt wurden, stelle sich die Frage nach dem Personenbezug nicht; er sei zu bejahen. Beispiele seien generative Sprachmodelle, die mit Sprachaufnahmen einer Person feinabgestimmt würden, um ihre Stimme zu imitieren oder KI-Modelle, die personenbezogen Daten auf Abruf aus dem Training bereitstellten – bei diesen KI-Modellen sei die DSGVO wie gewohnt zu beachten (Rn. 29).

Zur Beantwortung der ersten Frage konzentriert sich der EDSA im Weiteren auf KI-Modelle, die **nicht** darauf ausgelegt sind, personenbezogene Daten aus speziellen Trainingsdaten bereitzustellen. Dabei ist er der Ansicht, dass Informationen aus dem Trainingssatz (einschließlich personenbezogener Daten), in den Parametern des KI-Modells aufgenommen bleiben könnten („may still remain 'absorbed'") – dargestellt durch mathematische Objekte (Rn. 31). Diese Daten unterschieden sich von den eingegebenen Trainingsdaten, enthielten aber stets die ursprünglichen Informationen, die letztendlich extrahiert oder anderweitig, direkt oder indirekt, aus dem KI-Modell gewonnen werden könnten. Der EDSA schlussfolgert daher: Wann immer Informationen, die sich auf identifizierte oder identifizierbare Personen beziehen, deren personenbezogene Daten zum Trainieren des KI-Modells verwendet wurden, mit vernünftigen Mitteln aus dem Modell heraus gewonnen werden können, kann davon ausgegangen werden, dass dieses KI-Modell nicht anonym ist – mit anderen Worten: personenbezogene Daten enthält. Ob dies tatsächlich so ist, müsse von Fall zu Fall beurteilt werden (Rn. 34).

Damit ein KI-Modell als anonym angesehen werden kann, müssten nach EDSA zwei Voraussetzungen erfüllt sein: (1) Die Wahrscheinlichkeit einer *direkten* Extraktion personenbezogener Daten, die bei der Entwicklung des Modells verwendet wurden, und

(2) die Wahrscheinlichkeit, personenbezogene Daten – absichtlich oder unabsichtlich – aus Abfragen zu erhalten, müssen – unter Berücksichtigung aller Mittel, die vernünftigerweise vom Verantwortlichen oder einer anderen Person eingesetzt werden können – gering („insignificant“) sein (Rn. 43). Damit bezieht sich der EDSA letztlich auf die Bestimmung relativ anonymer Daten (Erwägungsgrund 26 DSGVO).

Um die Anonymität eines KI-Modells zu überprüfen, sollen die Aufsichtsbehörden die vom Verantwortlichen bereitgestellten Unterlagen prüfen. In der Stellungnahme des EDSA sind Methoden aufgelistet, die von für die Verarbeitung Verantwortlichen für den Nachweis der Anonymität verwendet werden können. Zum Beispiel Methoden, die Verantwortliche in der Entwicklungsphase ergreifen können, um die Sammlung personenbezogener Daten für das Training zu begrenzen oder zu verhindern, die Identifizierbarkeit der genutzten Daten zu verringern oder den Schutz der Daten nach dem Stand der Technik zu gewährleisten (Rn. 44 ff.).

Unter welchen Voraussetzungen können berechtigte Interessen als Rechtsgrundlage für die Entwicklung eines KI-Modells dienen?

Werden personenbezogene Daten verarbeitet, ist stets eine Rechtsgrundlage erforderlich, welche die Verarbeitung gestattet. Diese Rechtsgrundlagen sind unter anderem in Art. 6 DSGVO aufgelistet. Eine dieser Rechtsgrundlagen ist das „berechtigte Interesse“; dieses erlaubt es Unternehmen, personenbezogene Daten zu verarbeiten, wenn die Verarbeitung zur Wahrung der berechtigten Interessen des Verantwortlichen oder eines Dritten erforderlich ist, sofern nicht die Interessen der betroffenen Person überwiegen (Art. 6 Abs. 1 lit. f) DSGVO).

In Bezug auf die zweite und dritte Frage der DPC widmet sich der EDSA in seiner Stellungnahme dem „berechtigten Interesse“ als geeignete Rechtsgrundlage für die Verarbeitung personenbezogener Daten bei Entwicklung und Einsatz von KI-Modellen.

Er stellt klar, dass Unternehmen diese Rechtsgrundlage durchaus für die Verarbeitung von personenbezogenen Daten im Zusammenhang mit KI-Modellen heranziehen können, weist jedoch darauf hin, dass der bekannte Drei-Stufen-Test angewendet werden müsse (EDSA, Guidelines 01/24). Welche drei Kriterien nach EDSA kumulativ erfüllt sein müssen, um von einem berechtigten Interesse des

Verantwortlichen ausgehen zu können, haben wir hier schon einmal dargestellt: Legitime Ziele, Erforderlichkeit der konkreten Verarbeitung zur Zielerreichung und eine Interessenabwägung im Einzelnen.

Mit Blick auf die Verarbeitung personenbezogener Daten bei Entwicklung und Einsatz von KI-Modellen spielt laut EDSA in der Abwägung insbesondere eine Rolle, welche berechtigten Erwartungen eine betroffene Person im Hinblick auf die Datenverarbeitung haben darf. Für sie kann es schwierig sein, die Vielfalt potentieller Anwendungsmöglichkeiten und Verarbeitungsaktivitäten zu erfassen und die Komplexität der Arbeitsweise eines KI-Modells zu verstehen. Um beurteilen zu können, ob die betroffene Person vernünftigerweise erwarten kann, dass ihre personenbezogenen Daten verarbeitet werden, sind daher die diesen bereitgestellten Informationen als auch der Kontext der Verarbeitung von Bedeutung: Sind die personenbezogenen Daten öffentlich zugänglich oder nicht? In welcher Beziehung steht die betroffene Person zu dem Verantwortlichen der Datenverarbeitung? Aus welcher Quelle stammen die durch das KI-Modell verarbeiteten Daten? Wie wird das KI-Modell möglicherweise in Zukunft weiterverwendet?

Der EDSA schlägt in seiner Stellungnahme außerdem eine Reihe milderer Maßnahmen für die Entwicklungs- und Einsatzphase eines KI-Modells vor, um die Auswirkungen der Datenverarbeitung auf betroffene Personen zu begrenzen (Rn. 96 ff.). Diese sollten auf die Umstände des Falls und die Merkmale des KI-Modells, einschließlich seiner beabsichtigten Nutzung, zugeschnitten sein. Hierunter fallen etwa technische Maßnahmen, z. B. das Ersetzen personenbezogener Daten durch Fake-Daten (insbesondere im Rahmen des LLM-Trainings), und Transparenzmaßnahmen, wie die Veröffentlichung leicht zugänglicher Informationen – über jene nach Art. 13, 14 DSGVO hinaus – zu Kriterien der Datenerhebung und verwendeten Datensätzen.

Welche Folgen hat die unrechtmäßige Datenverarbeitung in der Entwicklungsphase des KI-Modells für spätere Verarbeitungsschritte?

Zu guter Letzt widmet sich der EDSA den Folgen einer unrechtmäßigen Verarbeitung personenbezogener Daten im

Rahmen der Entwicklungs- und Trainingsphase für die Rechtmäßigkeit der nachfolgenden Verarbeitung bei Einsatz des KI-Modells. Zur Beantwortung unterscheidet er drei verschiedene Szenarien:

In Szenario 1 werden (nicht anonymisierte) personenbezogene Daten im KI-Model gespeichert und durch denselben Verantwortlichen weiterverarbeitet. Hier ist im Einzelfall zu entscheiden, ob Entwicklungs- und anschließende Bereitstellungsphase unterschiedliche Zwecke verfolgen (und damit unterschiedliche Verarbeitungstätigkeiten darstellen) und inwieweit die Unrechtmäßigkeit der ersten Verarbeitung die Rechtmäßigkeit der nachfolgenden beeinflusst. Beruht die nachfolgende Verarbeitung auf einem berechtigten Interesse gem. Art. 6 Abs. 1 lit. f DSGVO, ist die mögliche Unrechtmäßigkeit einer vorangegangenen Verarbeitung im Rahmen der Abwägungsentscheidung zu berücksichtigen (z.B. mit Blick auf die Risiken für die betroffene Person oder die Tatsache, dass sie mit einer späteren Verarbeitung nicht rechnen musste). In solchen Fällen kann die Unrechtmäßigkeit der ursprünglichen Datenverarbeitung in der Entwicklungsphase Einfluss auf die Rechtmäßigkeit der nachfolgenden Verarbeitung bei Einsatz des KI-Modells haben (Rn. 122, 123).

In Szenario 2 werden personenbezogene Daten in einem KI-Modell durch einen Verantwortlichen gespeichert (auch hier nicht anonymisiert) und durch einen anderen Verantwortlichen – bei Bereitstellung des Modells – weiterverarbeitet. Nach EDSA sollte Letzterer eine angemessene Bewertung durchführen, um sicherzustellen, dass das KI-Modell nicht durch eine unrechtmäßige Verarbeitung personenbezogenen Daten entwickelt wurde (Rn. 129 ff.)

In Szenario 3 werden personenbezogene Daten unrechtmäßig verarbeitet, jedoch werden diese anschließend anonymisiert, bevor sie durch denselben oder einen anderen Verantwortlichen weiterverarbeitet werden. Wenn nachgewiesen werden kann, dass der nachfolgende Einsatz eines KI-Modells keine Verarbeitung personenbezogener Daten beinhaltet, ist die DSGVO laut EDSA nicht anwendbar und die Unrechtmäßigkeit der anfänglichen Verarbeitung hat keine Auswirkungen auf die nachfolgende Nutzung des KI-Modells (Rn. 134). Wenn ein Verantwortlicher im Anschluss an die Anonymisierung eines KI-Modells erneut

personenbezogene Daten bei Einsatz des KI-Modells erhebt, gilt in Bezug auf diese Erhebung und Verarbeitung wiederum die DSGVO. In diesen Fällen beeinflusst die ursprüngliche Unrechtmäßigkeit einer Datenverarbeitung in der Entwicklungsphase die spätere Datenverarbeitung ebenfalls nicht (Rn. 135). Dies wird für die Praxis ein wesentliches Szenario werden, wenn etwa Dritte die von anderen entwickelten Modelle weiterverwenden und so eine „Vergiftung" der nachfolgenden Verarbeitungsprozesse vermeiden können (Vermeidung „fruit of the poisonous tree").

Die Stellungnahme des EDSA ist ein erster Wegweiser, der die spezifischen Herausforderungen des Datenschutzes, die KI-Modelle mit sich bringen, auf europäischer Ebene adressiert und mehr Klarheit in Bereichen schafft, die bislang von Unsicherheit geprägt waren. Wir werden in diesem Jahr noch intensiv über KI und Datenschutz diskutieren – vielleicht ja gemeinsam in Berlin am 21. Februar 2025 auf dem DAV KI-Forum, in dem unsere Partnerin Dr. Kristina Schreiber unter Moderation von Frederik Richter, LL.M., Stiftung Datenschutz, auf dem Panel mitdiskutieren wird (https://anwaltverein.de/de/der-dav/dav-veranstaltungen/dav-forum/dav-ki-forum).

Für alle weiteren Fragen rund um das Datenschutzrecht stehen Ihnen gerne zur Verfügung

Dr. Kristina Schreiber
+49 221 65065-337
kristina.schreiber@loschelder.de

Dr. Simon Kohm
+49 221 65065-200
simon.kohm@loschelder.de

Dennis Pethke, LL.M.
+49 221 65065-337
dennis.pethke@loschelder.de

Rebecca Moßner
+49 221 65065-465
rebecca.mossner@loschelder.de

Impressum

LOSCHELDER RECHTSANWÄLTE
Partnerschaftsgesellschaft mbB
Konrad-Adenauer-Ufer 11
50668 Köln

Tel. +49 (0)221 65065-0, Fax +49 (0)221 65065-110
info@loschelder.de
www.loschelder.de