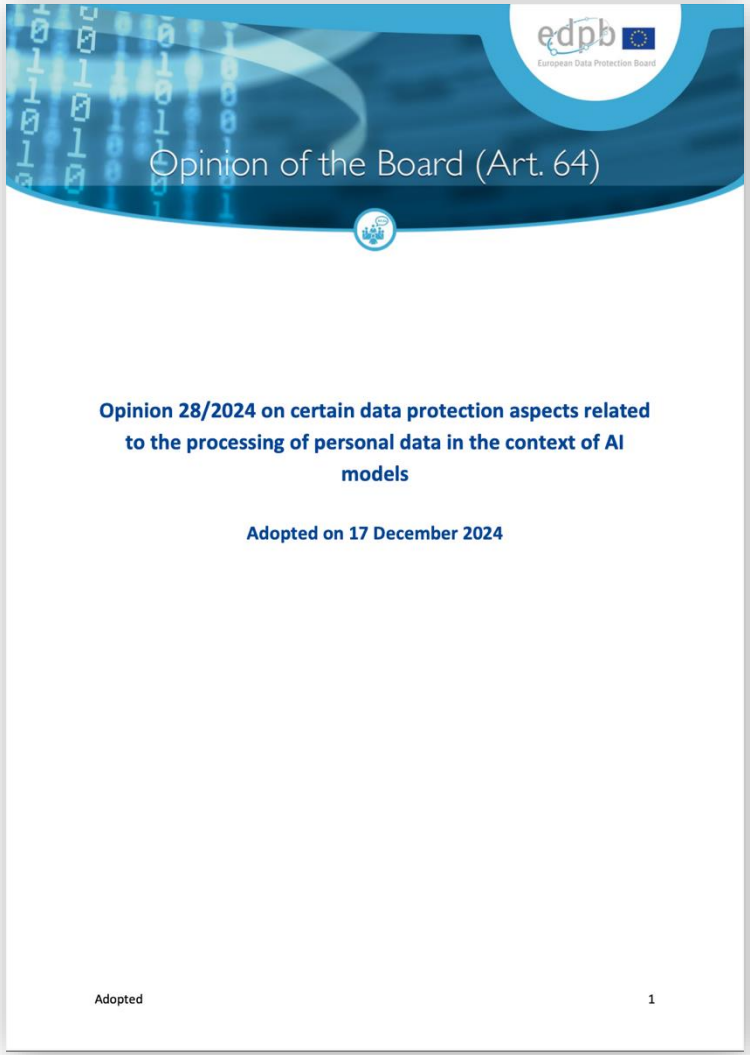
HÄRTING

KI und Datenschutz

Teil III: Die Position des EDSA zu Datenschutz und KI-Modellen

Stiftung Datenschutz | Datenschutz am Mittag | 30.01.2025

Sebastian Schulz | Rechtsanwalt



edpb
European Data Protection Board

Opinion of the Board (Art. 64)

Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models

Adopted on 17 December 2024

Adopted 1

Table of contents

Adopted 5

EDPB-Opinion 28/2024 – Background

- Verfahren nach Art. 64 Abs. 2 DSGVO
  - „Angelegenheit mit allgemeiner Geltung oder mit Auswirkungen in mehr als einem Mitgliedstaat"
  - eingereicht durch die Irish DPC
  - im Vorfeld: Stakeholder-Meeting; Austausch des EDSA mit dem EU AI Office

- begrenzter Scope: betrifft ausschließlich Teilaspekte des Datenschutzes im Zusammenhang mit der Entwicklung und dem Einsatz von KI-Modellen
  - 35 Seiten, 136 Randziffern
  - 4 Fragen, 4,5 Antworten
  - ausdrücklich keine Aussagen zu Art. 6 Abs. 4; Art. 9; Art. 22; Art. 25; Art. 35 DSGVO
  - EDSA arbeitet aktuell an weiteren Leitlinien zu spezifischen Fragen (z.B. Web Scraping)

Frage 1 – Datenschutzrechtliche Relevanz des KI-Modells

Question 1: 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)?

If the answer to question 1 is "yes":

i. At what stage of the processing operations leading to an AI Model is personal data no longer processed?

   a) How can it be demonstrated that the AI model does not process personal data?

ii. Are there any factors which would cause the operation of the final AI Model to no longer be considered anonymous?

   a) If so, how can the measures taken to mitigate, prevent or safeguard against these factors (so as to ensure the AI Model does not process personal data) be demonstrated?

If the answer to question 1 is "no":

i. What are the circumstances in which that might arise?

   a) If so, how can the steps that have been taken to ensure that the AI Model is not processing personal data be demonstrated?

Sind KI-Modelle, die mit personenbezogenen Daten trainiert wurden, als personenbezogen einzuordnen? Oder sind sie stets als anonym zu betrachten?*

* betrachtet werden nur KI-Modelle, „die nicht darauf ausgelegt sind" personenbezogene Daten vorzuhalten (Rz. 30)

EDSA: Datenschutzrechtliche Relevanz des KI-Modells

- IdR keine Speicherung der pbD in den einzelnen Schichten des LLM
- **Aber**: case by case-Betrachtung erforderlich
  - PbD können im Modell enthalten sein, entweder infolge der „Beziehungen zwischen den im Modell enthaltenen Daten oder durch Abfrage des Modells". (Rz. 31, 37)
  - In realistischen Szenarien kann es möglich sein, aus dem Modell Informationen abzuleiten, wie z.B. die Ableitung von Zugehörigkeiten. (Rz. 38)
  - Die Bewertung sollte unter Berücksichtigung „aller Mittel, die nach vernünftigem Ermessen von dem für die Verarbeitung Verantwortlichen oder einer anderen Person zur Identifizierung von Personen eingesetzt werden können". (Rz. 41, 49)
  - Differenzierung zwischen öffentlich zugänglichem und unternehmensinternem Modell (Rz. 44, 46)

Personenbezug in LLM – Was sagt die DSGVO?

Art. 4 Nr. 1 DSGVO „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung ... identifiziert werden kann."

- EuGH, Urt. v. 19.10.16 – C-582/14 (Breyer); (theoretische) rechtliche Möglichkeit zur Herstellung von Personenbezug genügt
- EuGH, Urt. v. 09.11.23 – C-319/22 (VIN); Mittel die „vernünftigerweise" zur Herstellung von Personenbezug genutzt werden
- EuG, Urt. v. 26.04.23 – T-557/20: relativer Maßstab (pending C-413/23)

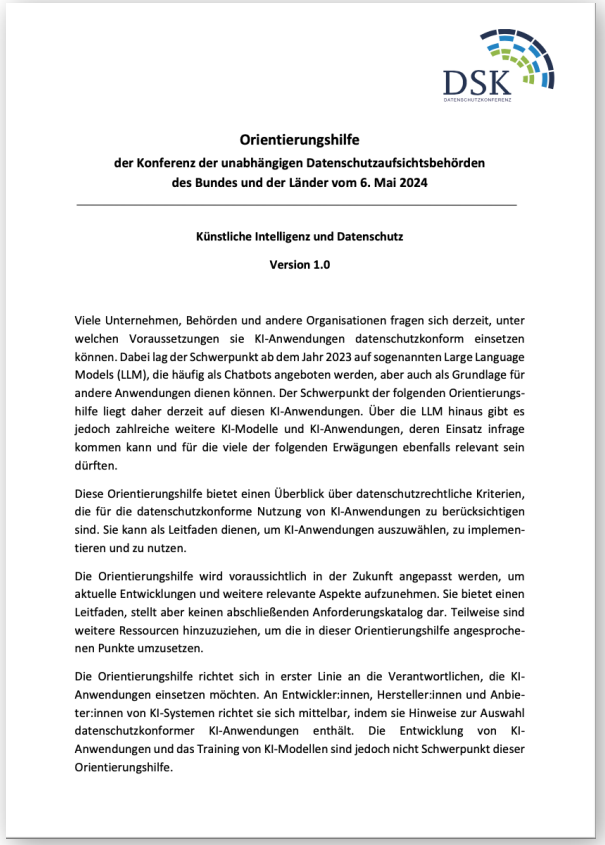
Personenbezug in LLM – Was sagt die DSGVO?

ErwG 26 S. 3: „Um festzustellen, ob eine natürliche Person identifizierbar ist, sollten **alle Mittel** berücksichtigt werden, die **von dem Verantwortlichen oder einer anderen Person nach allgemeinem Ermessen wahrscheinlich** genutzt werden, um die natürliche Person direkt oder indirekt zu identifizieren, wie beispielsweise das Aussondern."

ErwG 26 S. 4: Bei der Feststellung, ob Mittel nach allgemeinem Ermessen wahrscheinlich zur Identifizierung der natürlichen Person genutzt werden, sollten **alle objektiven Faktoren**, wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand, herangezogen werden, wobei die zum Zeitpunkt der Verarbeitung **verfügbare Technologie und technologische Entwicklungen** zu berücksichtigen sind.

Was stellt kein „vernünftigerweise" nutzbares (Dritt-)Wissen mehr dar?
→ Standard-Prompting < IT-Expert-Skills < Hacking
→ Bedeutung von Output-Filtern?

Positionen deutscher Aufsichtsbehörden – **DSK**



DSK
DATENSCHUTZKONFERENZ

Orientierungshilfe
der Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder vom 6. Mai 2024

Künstliche Intelligenz und Datenschutz

Version 1.0

Viele Unternehmen, Behörden und andere Organisationen fragen sich derzeit, unter welchen Voraussetzungen sie KI-Anwendungen datenschutzkonform einsetzen können. Dabei lag der Schwerpunkt ab dem Jahr 2023 auf sogenannten Large Language Models (LLM), die häufig als Chatbots angeboten werden, aber auch als Grundlage für andere Anwendungen dienen können. Der Schwerpunkt der folgenden Orientierungshilfe liegt daher derzeit auf diesen KI-Anwendungen. Über die LLM hinaus gibt es jedoch zahlreiche weitere KI-Modelle und KI-Anwendungen, deren Einsatz infrage kommen kann und für die viele der folgenden Erwägungen ebenfalls relevant sein dürften.

Diese Orientierungshilfe bietet einen Überblick über datenschutzrechtliche Kriterien, die für die datenschutzkonforme Nutzung von KI-Anwendungen zu berücksichtigen sind. Sie kann als Leitfaden dienen, um KI-Anwendungen auszuwählen, zu implementieren und zu nutzen.

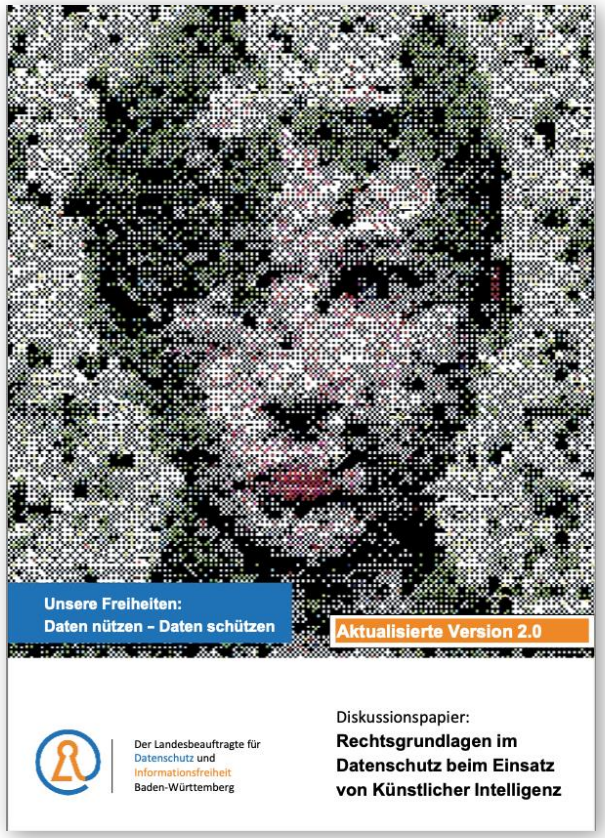
Die Orientierungshilfe wird voraussichtlich in der Zukunft angepasst werden, um aktuelle Entwicklungen und weitere relevante Aspekte aufzunehmen. Sie bietet einen Leitfaden, stellt aber keinen abschließenden Anforderungskatalog dar. Teilweise sind weitere Ressourcen hinzuzuziehen, um die in dieser Orientierungshilfe angesprochenen Punkte umzusetzen.

Die Orientierungshilfe richtet sich in erster Linie an die Verantwortlichen, die KI-Anwendungen einsetzen möchten. An Entwickler:innen, Hersteller:innen und Anbieter:innen von KI-Systemen richtet sie sich mittelbar, indem sie Hinweise zur Auswahl datenschutzkonformer KI-Anwendungen enthält. Die Entwicklung von KI-Anwendungen und das Training von KI-Modellen sind jedoch nicht Schwerpunkt dieser Orientierungshilfe.

ausdrücklich darauf hin, dass Anwendende sich nicht auf die Richtigkeit der Ergebnisse verlassen können, sondern diese überprüfen müssen. Hinsichtlich personenbezogener Daten besteht bei Unrichtigkeit jedoch ein Recht der betroffenen Personen auf Berichtigung. Diese Berichtigung muss in einer KI-Anwendung umsetzbar sein, zum Beispiel durch Korrektur von Daten oder durch ein Nachtraining/Fine Tuning.

28 Machen betroffene Personen von ihrem Recht auf Löschung gemäß Art. 17 Abs. 1 DS-GVO Gebrauch, ist zu beachten, dass einige KI-Anwendungen gegebenenfalls durch die Verknüpfung unterschiedlicher Daten einen Personenbezug herstellen können. Es ist daher besonders wichtig, dass bei der Löschung personenbezogener Daten darauf geachtet wird, dass eine Wiederherstellung des Personenbezugs dauerhaft unmöglich ist. Dies kann je nach KI-Anwendung auf verschiedenen Wegen umgesetzt werden.

Positionen deutscher Aufsichtsbehörden – **BaWü**



3. Bereitstellen von Anwendungen der Künstlichen Intelligenz

Ob es sich bei einem Bereitstellen von mit personenbezogenen Daten trainierten KI-System um eine Verarbeitung personenbezogener Daten handelt, bedarf einer differenzierten Bewertung. Auf der einen Seite kann die kostenfreie und kontogebundene Bereitstellung solcher Anwendungen eine Verarbeitung der zuvor zu ihrer Entwicklung und Fortentwicklung verwendeten personenbezogenen Daten darstellen. Ob dies der Fall ist, hängt davon ab, inwieweit die Trainingsdaten als noch in dem KI-System „enthalten" anzusehen sind, da sie bei der Nutzung der Anwendung weiter verarbeitet werden.[28] Auf der anderen Seite werden im Rahmen der Nutzung erhobene personenbezogene Daten unter Umständen durch das KI-System ebenfalls weiterverarbeitet, insbesondere durch ein weiteres Training der Anwendung. Eine solche Verarbeitung bedürfte einer separaten Rechtsgrundlage.

Positionen deutscher Aufsichtsbehörden – **HH**

Der Hamburgische Beauftragte für Datenschutz und Informationsfreiheit

Diskussionspapier: Large Language Models und personenbezogene Daten

Dieses Diskussionspapier bildet den derzeitigen Wissens- und Erkenntnisstand beim Hamburgischen Beauftragten für Datenschutz und Informationsfreiheit (HmbBfDI) zur Frage der Anwendbarkeit der Datenschutz-Grundverordnung (DSGVO) auf Large Language Models[1] (LLMs) ab. Das Papier ist ein *Debattenimpuls*. Es soll Unternehmen und Behörden dabei unterstützen, datenschutzrechtliche Komplexe besser zu verorten. Zu diesem Zweck werden vorliegend relevante technische Aspekte von LLMs erläutert, vor dem Hintergrund der Rechtsprechung des Gerichtshofs der Europäischen Union (EuGH) zum Personenbezug bewertet und daraus resultierende Folgen für die Praxis beleuchtet. Hieraus lassen sich drei grundlegende Thesen ableiten:

1. **Die bloße Speicherung eines LLMs stellt keine Verarbeitung im Sinne des Art. 4 Nr. 2 DSGVO dar. Denn in LLMs werden keine personenbezogenen Daten gespeichert. Soweit in einem LLM-gestützten KI-System personenbezogene Daten verarbeitet werden, müssen die Verarbeitungsvorgänge den Anforderungen der DSGVO entsprechen. Dies gilt insbesondere für den Output eines solchen KI-Systems.**

2. **Mangels Speicherung personenbezogener Daten im LLM können die Betroffenenrechte der DSGVO nicht das Modell selbst zum Gegenstand haben. Ansprüche auf Auskunft, Löschung oder Berichtigung können sich jedoch zumindest auf Input und Output eines KI-Systems der verantwortlichen Anbieter:in oder Betreiber:in beziehen.**

3. **Das Training von LLMs mit personenbezogenen Daten muss datenschutzkonform erfolgen. Dabei sind auch die Betroffenenrechte zu beachten. Ein ggf. datenschutzwidriges Training wirkt sich aber nicht auf die Rechtmäßigkeit des Einsatzes eines solchen Modells in einem KI-System aus.**

[1] Gemeint sind hierbei allein die Modelle als wichtiger, aber nicht alleiniger Bestandteil eines KI-Systems (z. B. eines LLM-basierten Chatbots).

www.datenschutz-hamburg.de
E-Mail: mailbox@datenschutz.hamburg.de
Ludwig-Erhard-Straße 22 - D-20459 Hamburg - Tel.: 040 - 4 28 54 - 40 40 - Fax: 040 - 4 28 54 – 40 00
Vertrauliche Informationen sollten auf elektronischem Weg nur verschlüsselt an uns übermittelt werden.
Unser öffentlicher PGP-Schlüssel ist im Internet verfügbar (Fingerprint: 0932 5798 33C1 8C21 6C90 E770 0BDD BAE4 3377 5707).

1. **Die bloße Speicherung eines LLMs stellt keine Verarbeitung im Sinne des Art. 4 Nr. 2 DSGVO dar. Denn in LLMs werden keine personenbezogenen Daten gespeichert. Soweit in einem LLM-gestützten KI-System personenbezogene Daten verarbeitet werden, müssen die Verarbeitungsvorgänge den Anforderungen der DSGVO entsprechen. Dies gilt insbesondere für den Output eines solchen KI-Systems.**

2. **Mangels Speicherung personenbezogener Daten im LLM können die Betroffenenrechte der DSGVO nicht das Modell selbst zum Gegenstand haben. Ansprüche auf Auskunft, Löschung oder Berichtigung können sich jedoch zumindest auf Input und Output eines KI-Systems der verantwortlichen Anbieter:in oder Betreiber:in beziehen.**

3. **Das Training von LLMs mit personenbezogenen Daten muss datenschutzkonform erfolgen. Dabei sind auch die Betroffenenrechte zu beachten. Ein ggf. datenschutzwidriges Training wirkt sich aber nicht auf die Rechtmäßigkeit des Einsatzes eines solchen Modells in einem KI-System aus.**

Foto Adobe Stock/fleartbettung F.A.Z.

Wahrscheinlichkeiten reichen nicht

Von Thomas Fuchs und Markus Wünschelbaum

Trinkt der bayerische Ministerpräsident Markus Söder gerne Bier?" ChatGPT antwortet auf diese Frage derzeit wahrscheinlich mit „Ja, ge Datenfragmente, eine gezielte Suche nach Informationen über bestimmte Personen im LLM ist nicht möglich. In der Praxis erfordern solche Angriffe nierens der Modelle, der Eingabe von Nutzerdaten in KI-Systeme und vor allem während der Ausgabe sowie deren Verwendung. Ansprüche auf Löschung,

Auf Wahrscheinlichkeiten kommt es an

Von Rolf Schwartmann und Moritz Köhler

Es gibt keine Sprache ohne Information. Da LLM statistische Wahrscheinlichkeiten zur Funktionsweise der menschlichen Sprache repräsen- Versicherungsnummer zielsicher auf Grundlage einer wahrscheinlichen Aneinanderreihung von Tokens ermittelt? Gerade die neuen Sprachmodelle enthalten sind. Die oftmals angeführte Differenzierung zwischen KI-System und LLM ist zwar in beiden Diskussionen für das Verständnis der technischen Hinter-

Und wer muss das alles nach- und beweisen?

EDSA: Der Verantwortliche.

- To conduct their assessment, SAs should **review the documentation provided by the controller to demonstrate the anonymity** of the model (Exec Summary).
- Therefore, for a SA to agree with the controller that a given AI model may be considered anonymous, it should check at least **whether it has received sufficient evidence** that, with reasonable means: (i) personal data, related to the training data, cannot be extracted out of the model; and (ii) any output produced when querying the model does not relate to the data subjects whose personal data was used to train the model (Rz. 38)

Vereinbar mit Rechenschaftspflicht aus Art. 5 Abs. 2 DSGVO?

- HambBfDI: DSGVO verlangt von der DPA „handfeste Beweise" für das Vorliegen von pbD
- Parallele zur Nachweispflicht bei Anonymisierung?

Fragen 2 und 3 – Interessenabwägung als Rechtsgrundlage

Question 2: Where a data controller is relying on legitimate interests as a legal basis for personal data processing to create, update and/or develop an AI Model, how should that controller demonstrate the appropriateness of legitimate interests as a legal basis, both in relation to the processing of third-party and first-party data?

i. What considerations should that controller take into account to ensure that the interests of the data subjects, whose personal data are being processed, are appropriately balanced against the interests of that controller in the context of:

   a) Third-party data

   b) First-party data

Question 3: Post-training, where a data controller is relying on legitimate interests as a legal basis for personal data processing taking place within an AI Model, or an AI System of which an AI Model forms part, how should a controller demonstrate the appropriateness of legitimate interests as a legal basis?

Unter welchen Umständen können sich Entwickler und Anbieter von KI-Modellen auf Art. 6 Abs. 1 lit. f DSGVO als Rechtsgrundlage für das Training und den Einsatz von KI-Modellen berufen, soweit hierbei personenbezogene Daten verarbeitet werden?

EDSA: Interessenabwägung als Rechtsgrundlage

- **Kernbotschaft**: Art. 6 Abs. 1 S. 1 lit. f DSGVO denkbare Rechtsgrundlage
- im Wesentlichen Verweise auf Opinion 01/2024 zu Art. 6 Abs. 1 S. 1 lit. f DSGVO
- Drei-Stufen-Prüfung
  - Berechtigtes Interesse: weites Verständnis okay
  - Erforderlichkeit
    - Verarbeitung pbD kann erforderlich sein, um Bias zu vermeiden
    - kein milderes Mittel
    - Abwägungsentscheidung (Kriterien ab Rz. 79)
    - umfangreiche Auflistung von möglichen mitigierenden Maßnahmen (ab Rz. 96)

Abwägungsentscheidung – Kriterien (Auszug)

- For example, if a processing activity entails **benefits for the data subject**, these may be taken into account in the balancing test. (Rz. 82)
- For example, the **use of web scraping in the development phase** may lead - in the absence of sufficient safeguards - to **significant impacts on individuals**, due to the large volume of data collected, the large number of data subjects, and the indiscriminate collection of personal data. (Rz. 86)
- **Reasonable expectations:** To this end, the information provided to data subjects may be considered to assess whether data subjects can reasonably expect their personal data to be processed. **Aber**: „not sufficient in itself" (Rz. 92)
- **Insgesamt**: tendenziell keine Berücksichtigung von ohnehin verpflichtenden Maßnahmen (z.B. aus Artt. 13, 25, 32)

Frage 4 – Infizierende Wirkung rechtswidrigen Trainings

Question 4: If an AI Model has been found to have been created, updated or developed using unlawfully processed personal data, what is the impact of this, if any, on the lawfulness of the continued or subsequent processing or operation of the AI model, either on its own or as part of an AI System, where:

i. The AI Model, either alone or as part of an AI System, is processing personal data?

ii. Neither the AI Model, nor the AI Model as part of an AI System, is processing personal data?

Welche Auswirkungen hat ein datenschutzwidriges Training eines KI-Modells auf dessen Einsatz in der Praxis?

EDSA: Infizierende Wirkung rechtswidrigen Trainings

- **Grundsätzliches**
  - Verarbeitung in der Entwicklungsphase kann zu Verstößen gegen „andere Bestimmungen" der DSGVO führen, z.B. Artt. 13, 14, 25 (Rz. 111)
  - datenschutzrechtliche Rollenkonstellation der beteiligten Akteure (Anbieter, Betreiber) von maßgeblicher Bedeutung (case by case, Rz. 112)
  - Art. 17 Abs. 1 lit. d → Löschverpflichtung bei unrechtmäßiger Verarbeitung im Rahmen des Trainings (Rz. 115)
- Position differenziert nach **drei Szenarien**

EDSA: Infizierende Wirkung rechtswidrigen Trainings

Szenario 1

Ein für die Verarbeitung Verantwortlicher **verarbeitet unrechtmäßig personenbezogene Daten zur Entwicklung des Modells**, die **personenbezogenen Daten werden in dem Modell gespeichert** und **anschließend von demselben Verantwortlichen** verarbeitet (z. B. im Zusammenhang mit dem Einsatz des Modells, Rz. 119f.).

- **prozessual**: Maßnahmen von DPAs gegen wegen der Rechtswidrigkeit der initialen Verarbeitung können auf den Einsatz des Modells durchschlagen
- **materiellrechtlich**: wenn Verarbeitung bei Einsatz des Modells auf Grundlage von Art. 6 Abs. 1 lit. f basiert, muss im Rahmen der Interessenabwägung die Rechtswidrigkeit des vorangegangenen Verarbeitungsschrittes berücksichtigt werden

Fazit: Infektion möglich

EDSA: Infizierende Wirkung rechtswidrigen Trainings

Szenario 2

Ein für die Verarbeitung Verantwortlicher verarbeitet **unrechtmäßig personenbezogene Daten**, **um das Modell zu entwickeln**, die **personenbezogenen Daten werden in dem Modell gespeichert** und **von einem anderen Verantwortlichen im Zusammenhang mit dem Einsatz des Modells verarbeitet** (Rz. 124f.).

- getrennte Bewertung nach Verantwortlichkeiten erforderlich (case by case)
- Anordnungen der DPA gegen den Erstverantwortlichen (Entwicklung) können auf den Zweitverantwortlichen (Einsatz) durchschlagen; Hinweis auf Art. 19
- Zweitverantwortlicher soll eine angemessene Bewertung durchführen, „dass das KI-Modell nicht durch unrechtmäßige Verarbeitung von Daten entwickelt wurde".

Fazit: Infektion (tendenziell) möglich

EDSA: Infizierende Wirkung rechtswidrigen Trainings

Szenario 3

Ein für die Verarbeitung Verantwortlicher verarbeitet **unrechtmäßig personenbezogene Daten, um das Modell zu entwickeln**, stellt anschließend aber sicher, dass das **Modell anonymisiert** wird, bevor **derselbe oder ein anderer Verantwortlicher** eine weitere Verarbeitung von Daten im Zusammenhang mit dem Einsatz einleitet (Rz. 133f.).

- Wenn nachgewiesen werden kann, dass der spätere Betrieb des KI-Modells keine Verarbeitung personenbezogener Daten nach sich zieht, ist die DSGVO auf das Modell nicht anwendbar.
- Die bloße Behauptung der Anonymität des Modells reicht jedoch nicht aus (vgl. Frage 1).

Fazit: keine Infektion

Zusammenfassung der Position des EDSA

1. Damit ein **KI-Modell als anonym** angesehen werden kann, **muss** sowohl
(1.) die Wahrscheinlichkeit einer direkten (einschließlich probabilistischen) Extraktion personenbezogener Daten von Personen, deren Daten zur Entwicklung des Modells verwendet wurden, als auch
(2.) die Wahrscheinlichkeit absichtlich oder unabsichtlich solche personenbezogenen Daten aus Abfragen zu erhalten, **sehr gering sein**.

2. Die **allgemeine Interessenabwägungsklausel ist grundsätzlich eine taugliche Rechtsgrundlage** für die Verarbeitung von pbD sowohl im Rahmen des Trainings als auch des Einsatzes des KI-Modells.

3. Wenn ein **KI-Modell datenschutzwidrig trainiert** wurde und im Rahmen der Anwendung von Personenbezug auszugehen ist, schlägt die Rechtswidrigkeit u.U. (case by case) auf das Modell durch.

4. Wenn ein **KI-Modell datenschutzwidrig trainiert** wurde, im Rahmen der Anwendung aber **kein Personenbezug** mehr herstellbar ist, dann ist das KI-Modell dem Datenschutzrecht entzogen (und nutzbar).

WIE GEHT ES WEITER?

O-Ton
Thomas Fuchs

EDSA Art. 64 (2) Stellungnahme zu KI-Modellen erwartet im Dezember 2024

Wichtiger Schritt für die Rechtssicherheit –
aber noch nicht das Ende der Fahnenstange

Wahrscheinlichkeiten reichen nicht

Bloße Möglichkeiten können staatliche Eingriffe durch Maßnahmen wie bspw. Geldbußen nicht rechtfertigen

Modelle werden durch die KI-VO reguliert

Produkte (bspw. Modelle) sind bereits reguliert, wenn sie ein wahrscheinliches Risiko für die Gesellschaft begründen.

HÄRTING

Sebastian Schulz

Partner

schulz@haerting.de

www.linkedin.com/in/sebastian-schulz-HAERTING