bitkom

Stellungnahme

August 2025

Bitkom zur Konsultation der Bundesbeauftragten für den Datenschutz und die Informationsfreiheit zum datenschutzkonformen Umgang mit personenbezogenen Daten in KI-Modellen

Einleitung

Bitkom begrüßt die Initiative der Bundesbeauftragten für den Datenschutz und die Informationsfreiheit (BfDI), mit der Konsultation einen breiten, offenen Diskurs über die Entwicklung und Nutzung großer Sprachmodelle (LLMs) in einer Weise zu ermöglichen, die den risikobasierten Rahmen der DSGVO respektiert. Als Digitalverband bringt Bitkom seine Expertise gerne in diesen Prozess ein und unterstützt das Ziel, praxistaugliche, rechtssichere und innovationsfreundliche Lösungen für den datenschutzkonformen Einsatz von KI-Systemen zu entwickeln. Es ist wichtig klarzustellen, dass Modelle keine Datenbanken sind. Der Zweck des Modelltrainings besteht darin, Muster und Strukturen in den Daten zu verstehen. Modelle »speichern« personenbezogene Daten nicht in der Weise, wie es Datenbanken tun. Datenspeicherung ist der Prozess der Aufzeichnung und Aufbewahrung von Daten, damit sie in ihrem ursprünglichen Format abgerufen werden können. Im Gegensatz dazu werden Modellen während des Trainings die statistischen Beziehungen zwischen Daten vermittelt und diese Erkenntnisse werden in Modellgewichten kodiert.

Wir bedanken uns für die Möglichkeit der Teilnahme und legen nachfolgend unsere Stellungnahme entlang der gestellten Konsultationsfragen dar.

Vorbemerkung

Die in der Konsultation gewählte Fokussierung auf »große Sprachmodelle« (Large Language Models, LLMs) greift aus unserer Sicht zu kurz. Gleichzeitig sollte die Konsultation deutlich machen, dass die datenschutzrechtlichen Fragestellungen nicht pauschal für alle KI-Modelle gelten, sondern insbesondere dort relevant sind, wo das Risiko besteht, dass personenbezogene Daten aus dem Modell mit angemessenen Mitteln gewonnen werden könnten. Die datenschutzrechtlichen Fragestellungen, die in der Konsultation aufgeworfen werden, betreffen in der Praxis nicht nur LLMs, sondern grundsätzlich alle generativen KI-Modelle und zwar unabhängig von Modellgröße oder konkretem Einsatzbereich.

Zudem bleibt unklar, wie der Begriff »großes Sprachmodell« im Kontext der Konsultation konkret abgegrenzt werden soll. Die jüngst veröffentlichten Guidelines der Europäischen Kommission zu GPAI-Systemen (C(2025) 5045 final) zeigen, wie anspruchsvoll eine präzise technische und regulatorische Definition sein kann. Die Konsultation verzichtet jedoch auf eine solche Spezifizierung, was Interpretationsspielräume eröffnet und Unsicherheiten schafft.

Viele Unternehmen, insbesondere kleine und mittlere Unternehmen sowie sektorale Anwender, entwickeln und nutzen keine großskaligen Foundation Models, sondern greifen auf kleinere, angepasste Modelle zurück (z. B. kompakte Sprachmodelle, fine-getunte Vertical Models oder Embedded KI-Komponenten in Softwarelösungen). Für diese weit verbreiteten Anwendungsformen stellen sich datenschutzrechtlich ebenfalls relevante Fragen, etwa zum Einsatz von Trainingsdaten, zur Memorisierung oder zur Wahrnehmung von Betroffenenrechten.

Im Zusammenhang mit der Anonymität von Modellen ist außerdem anzumerken, dass – wie auch die EDSA-Stellungnahme 28/2024 betont – eine Einzelfallprüfung erforderlich ist, um festzustellen, ob ein KI-Modell als anonym gelten kann. Verantwortliche und Entwickler sollten nachweisen, dass die Wahrscheinlichkeit, personenbezogene Daten direkt aus ihren Modellen zu extrahieren, unerheblich ist. Entsprechende technische und organisatorische Maßnahmen sollten transparent dargelegt und überprüft werden können (vgl. Abschnitt 3.2.2 der Stellungnahme des EDPB). Dies fördert die Rechtssicherheit und Vertrauenswürdigkeit von KI-Anwendungen.

Vor diesem Hintergrund regen wir an, künftig auch kleinere, domänenspezifische KI-Modelle systematisch in die regulatorische Diskussion einzubeziehen, um praxistaugliche, differenzierende und umfassende Vorgaben für die Breite der Wirtschaft zu ermöglichen.

Anonymität des Modells

Wird das Training der KI mit anonymen Daten durchgeführt, ist die DSGVO auf das Training nicht anwendbar. Allerdings ist bei KI-Modellen eine vollständige

Anonymisierung angesichts der zum Training verwendeten Datenmengen in der Regel nicht zuverlässig möglich.

Frage 1

Nach Erwägungsgrund 26 Satz 3 DSGVO sollten bei der Prüfung, ob eine natürliche Person identifizierbar ist, 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. Unter Berücksichtigung der in der EDSA Stellungnahme 28/2024 Rn. 35ff. gelisteten Vorgehen, unter welchen Umständen könnte ein LLM als anonym erachtet werden?

Die Frage, ob ein LLM als anonym gelten kann, hängt nicht allein von den Trainingsdaten ab, sondern vom Gesamtzusammenhang, demnach der Modellarchitektur, der Wahrscheinlichkeit einer Reidentifikation und den realistisch verfügbaren Mitteln, um eine konkrete Person zu identifizieren.

Darüber hinaus ist hervorzuheben, dass in der Trainingsvorbereitung mittels Bots und Crawler vor allem Inhalte von öffentlichen Websites (»Web-Scraping«) sammeln. Auch wenn sich darunter personenbezogene Daten befinden können, handelt es sich häufig um Informationen, die bereits von den betroffenen Personen selbst öffentlich zugänglich gemacht wurden oder deren Veröffentlichung im jeweiligen Kontext erwartet werden durfte (z. B. LinkedIn-Profile oder Angaben von Wissenschaftler*innen zu ihrer beruflichen Position in Fachzeitschriften). Insofern ist das Risiko der Re-Identifikation bei solchen bereits öffentlich verfügbaren Daten tendenziell geringer und bewegt sich eher im Rahmen der Erwartungen der betroffenen Personen.

Bitkom teilt die Auffassung, dass eine vollständige Anonymisierung der für das Training verwendeten Daten bei heutigen Modellen nur schwer zuverlässig möglich ist. Modelle »speichern« personenbezogene Daten nicht in der Weise, wie dies Datenbanken tun. Datenspeicherung ist der Prozess der Aufzeichnung und Aufbewahrung von Daten, damit diese in ihrem ursprünglichen Format abgerufen werden können. Im Gegensatz dazu werden Modellen während des Trainings die statistischen Beziehungen zwischen Daten vermittelt, und diese Erkenntnisse werden in Modellgewichten kodiert. Die EDPB-Stellungnahme 28/2024 erkennt an, dass die Bewertung der Identifizierbarkeit kontext- und risikobasiert erfolgen muss, ohne dass hierfür ein vollständiger Ausschluss sämtlicher personenbezogener Daten erforderlich ist. Personenbezogene Daten können in bestimmten Fällen bewusst zur Minderung von Bias oder zur Verbesserung der Modellleistung genutzt werden, in anderen Fällen ist ihr Vorkommen unbeabsichtigt aus der Zusammenstellung großer, vielfältiger Datensätze. Wir unterstützen ausdrücklich den Ansatz der EDPB-Stellungnahme 28/2024 zur Definition der Anonymität von KI-Modellen sowie die von der BfDI zitierte EDSA-Definition, welche zu Recht den Schwerpunkt auf die Wahrscheinlichkeit einer Reidentifikation legt. Nach unserer Auffassung kann ein Modell dann als anonym gelten, wenn

- es nicht gezielt zur Reproduktion personenbezogener Daten konzipiert wurde,
- eine Identifikation natürlicher Personen mit zumutbaren Mitteln nicht möglich ist,
- das Modell keine stabilen personenbezogenen Inhalte ausgibt – weder explizit noch implizit.

Wir regen an, den Schwellenwert für die Annahme von Anonymität kontext- und risikobasiert zu definieren. Dabei sollte in bestimmten, klar kontrollierten Szenarien (z. B. intern betriebene Modelle mit strengen technischen und organisatorischen Maßnahmen) ein niedrigerer Schwellenwert gelten.

Zusätzlich zu den bereits genannten Kriterien sollten folgende Gesichtspunkte berücksichtigt werden:

1. Bei internen Modellen, die ausschließlich autorisiertem Personal zugänglich sind und unter strengen technischen und organisatorischen Maßnahmen betrieben werden, ist die Wahrscheinlichkeit einer Re-Identifikation deutlich geringer als bei öffentlich zugänglichen Modellen.
2. Der Einsatz moderner, datenschutzfreundlicher Verfahren, wie Differential Privacy, Regularisierung, Deduplizierung, Datenbereinigung, Dropout, Loss Masking, Output-Filterung, Prompt-Shielding, Zugriffskontrollen und Audit-Logging, kann das Risiko von ungewollter Wiedergabe und Inferenzangriffen signifikant verringern.
3. LLMs kodieren Daten in hochdimensionalen statistischen Repräsentationen. Eine Extraktion personenbezogener Daten ist in den meisten kontrollierten Szenarien technisch nicht mit zumutbaren Mitteln möglich, insbesondere wenn
   - kein Zugriff auf Trainingsdaten während der Inferenz besteht,
   - kein Feintuning mit identifizierbaren Personendaten erfolgt,
   - das Modell gegen Extraktionsangriffe getestet ist.
4. Die DSGVO fordert Rechenschaftspflicht, keine absolute Garantie. Wenn Verantwortliche Datenschutz-Folgenabschätzungen durchführen, TOMs und Risikobewertungen dokumentieren, auf die Wiedergabe von personenbezogenen Daten testen und Transparenz gegenüber Betroffenen auf der Anwendungsebene wahren, sollte dies für die Anerkennung der Anonymität ausreichen, auch bei verbleibenden theoretischen Restrisiken.

Frage 2

Welche technischen Maßnahmen setzen Sie bereits ein bzw. planen Sie einzusetzen, um die Memorisierung von Daten zu verhindern (wie z.B. Deduplikation, Verwendung anonymer bzw. anonymisierter Trainingsdaten, Fine-Tuning ohne personenbezogene Daten, Differential Privacy, etc.)? Welche Erfahrungen haben Sie damit gemacht?

Viele unserer Mitglieder setzen bereits eine Vielzahl technischer Maßnahmen ein, um die Rekonstruktion von Trainingsdaten, insbesondere personenbezogene Daten, in

Modellausgaben zu vermeiden oder deren Auftreten zu minimieren. Dazu gehören unter anderem:

- **Deduplikation:** Mehrfache Vorkommen derselben Inhalte werden vor dem Training entfernt, um Überrepräsentationen zu verhindern.
- **Datenbereinigung und Reduktion strukturierter Identifier:** Vorverarbeitungsschritte sorgen dafür, dass sensible Merkmale entfernt oder abstrahiert werden.
- **Einsatz unstrukturierter Daten:** Solche Daten erschweren eine Reidentifikation erheblich, da sie meist keine direkten Identifikatoren enthalten.
- **Fine-Tuning mit kontrollierten Datenquellen:** Personalisierte Anpassungen von Modellen erfolgen gezielt mit anonymisierten oder pseudonymisierten Daten.
- **Ausgabefilter und Prompt-Blocking:** Modelle werden nachträglich so gesteuert, dass problematische Inhalte unterdrückt werden.
- **Experimentelle Ansätze wie Differential Privacy:** Diese werden punktuell eingesetzt, sind aber derzeit noch nicht auf großskalige Trainingsverfahren übertragbar.

Darüber hinaus werden moderne, datenschutzfreundliche Verfahren wie Loss Masking, Output Filtering, Prompt Shielding, Zugriffskontrollen, Audit Logging, Red Teaming, synthetische Daten und automatisierte Risiko-Messsysteme eingesetzt. Anbieter setzen z.B. auch umfassendes Red Teaming für generative KI-Produkte um und teilen die gewonnenen Erkenntnisse mit der Branche. Ein Lebenszyklus-Ansatz, der Maßnahmen in allen Phasen berücksichtigt, ist essenziell. Diese Maßnahmen zeigen Wirkung, insbesondere in Verbindung mit geeigneten Governance-Strukturen und internem Monitoring.

Wir empfehlen einen Lebenszyklus-Ansatz, der Maßnahmen in allen Phasen berücksichtigt:

- **Datenminimierung und De-Identifizierung:** Robuste Bereinigung und Filterung vor dem Training, ggf. Umwandlung sensibler Daten in synthetische Daten mithilfe von Loss-Masking-Verfahren.
- **Trainingskontrollen und Governance:** Wenn die Verwendung von Datensätzen aus risikoreichen Quellen erforderlich ist, sind geeignete Sicherheitsvorkehrungen zu treffen.
- **Modellarchitektur und Nach-Trainings-Maßnahmen:** Modelle so auslegen, dass sie Muster statt einzelner Datenpunkte lernen; nachträgliche Schutzmaßnahmen wie Output-Filter, Prompt-Shielding und Grounding-Mechanismen einsetzen.
- **Red Teaming und adversariale Tests:** Systematische Prüfungen, ob personenbezogene Daten extrahierbar sind.

Automatisierte Risiko-Messsysteme: Kontinuierliche, teils automatisierte Bewertung des Risikos, dass ein Modell personenbezogene Daten generiert, inkl. Rückkopplung für Modellverbesserungen.

Frage 3

Wie schätzen Sie das Risiko ein, dass personenbezogene Daten aus einem LLM extrahiert werden? Erläutern Sie Ihre Einschätzung möglichst anhand konkreter Beispiele, Einzelfälle oder empirischer Beobachtungen.

Bitkom hält das Risiko einer gezielten Extraktion (Zur Klarstellung: Mit »Extraktion« ist hier böswillige Exfiltration (z. B. durch Jailbreaking-Techniken) gemeint.) personenbezogener Daten aus einem LLM unter realistischen Bedingungen für sehr gering. Eine Reihe von Faktoren begründet diese Einschätzung:

- Die Ausgabe eines Modells basiert auf Wahrscheinlichkeitsverteilungen, nicht auf gespeicherten Daten.
- Das bloße Vorkommen eines Namens oder Datums im Output ist noch kein Beleg für eine Reproduktion personenbezogener Daten.
- Nur bei sehr häufigen, gut bekannten Informationen – etwa zu Personen des öffentlichen Lebens – kann eine Ausgabe entstehen, die mit den öffentlich zugänglichen Informationen übereinstimmt. Dabei ist zu beachten, dass Personen des öffentlichen Lebens in der Regel andere Erwartungen an den Datenschutz haben als andere betroffene Personen.

Die Forschung zu Bedrohungsarten (z. B. Membership Inference, Model Inversion) ist wertvoll, aber vielfach theoretisch und schwer auf industrielle Kontexte übertragbar. Die Erforschung von Angriffen auf die Privatsphäre (z. B. Mitgliedschaftsableitung, Modellinversion) ist wertvoll. Die meisten Entwickler führen Routinetests durch, um sicherzustellen, dass die Modelle wie vorgesehen funktionieren und dass problematische Eingaben und/oder Ausgaben gemäß den Kontrollen des Modells blockiert werden.

Verarbeitung memorisierter Daten

Frage 4

Datenschutzrecht knüpft an die Verarbeitung personenbezogener Daten an. Jede Eingabe eines Prompts löst eine Berechnung im KI-Modell aus, bei der die in Form von Parametern repräsentierten (personenbezogenen) Daten Einfluss auf das Berechnungsergebnis nehmen. Stellt diese Berechnung eine Verarbeitung dieser Daten im Sinne von Artikel 4 Nr. 2 DSGVO dar, selbst wenn das Berechnungsergebnis, also die Ausgabe des KI-Modells, nicht personenbezogen ist?

Bitkom vertritt die Auffassung, dass die bloße Inferenz eines LLMs keine »Verarbeitung personenbezogener Daten« im Sinne des Art. 4 Nr. 2 DSGVO darstellt – sofern keine personenbezogenen Daten im Output enthalten sind.

Der Vergleich mit einem Taschenrechner ist hier hilfreich: Auch dieser verarbeitet Eingaben anhand interner Rechenregeln, ohne dabei personenbezogene Daten zu »verarbeiten«. Genauso berechnet ein LLM eine statistisch plausible Antwort auf Basis seiner gelernten Parameter – nicht auf Basis gespeicherter personenbezogener Informationen.

Ein zu weiter Verarbeitungsbegriff würde letztlich dazu führen, dass fast jede komplexe Software – selbst einfache Rechtschreibkorrekturen – in den Anwendungsbereich der DSGVO fielen. Ein solcher Ansatz wäre praxisfern und würde Innovation erheblich behindern.

Wir betonen, dass Modellparameter keine personenbezogenen Daten darstellen, solange sie nicht in einer Weise genutzt werden, die eine Identifizierung natürlicher Personen ermöglicht. Darüber hinaus ist die interne Berechnung eines LLM funktional mit der Verarbeitung anonymer Daten vergleichbar, sofern das Ergebnis keine personenbezogenen Daten enthält und keine Rekonstruktion möglich ist. Dies steht im Einklang mit dem fallbezogenen Ansatz des EDSA.

Problem: Verantwortlichkeit

Unternehmen nutzen häufig LLMs von Drittanbietern, die ggf. lokal weiterentwickelt oder feinjustiert werden. Dies wirft datenschutzrechtliche Fragen hinsichtlich der Rollenverteilung (z. B. Verantwortlicher, Auftragsverarbeiter) entlang der KI-Wertschöpfungskette auf. Vor dem Hintergrund der EDSA-Stellungnahme 28/2024 ist insbesondere zu klären, wie Verantwortlichkeiten für die Kontrolle der Datenverarbeitung aufgeteilt werden. Der Modellentwickler trägt die Verantwortung für den Trainingsprozess und die eingesetzten Daten, während der Bereitsteller oder Anwender für etwaige Feinabstimmungen oder eigene Verarbeitungen zuständig ist. Dabei sollte sichergestellt werden, dass Nachweise über den Umgang mit Trainingsdaten und etwaige Anonymisierungen transparent und überprüfbar sind, ohne dass es zu unverhältnismäßigen Doppelprüfungen oder unnötigen bürokratischen Belastungen entweder beim Entwickler oder beim Bereitsteller kommt.

Zudem sollte eine solche Feststellung durch die zuständige Datenschutzaufsichtsbehörde des LLM-Providers auch für andere Aufsichtsbehörden verbindlich gelten, um eine einheitliche und harmonisierte Bewertung innerhalb der EU zu gewährleisten.

Verarbeitet ein Anwender bei der Weiterentwicklung eines LLMs selbst personenbezogene Daten, trägt er die Verantwortung für die Anonymität des angepassten Modells. Eine erneute Prüfung des ursprünglichen Modells wäre in diesem Fall nicht erforderlich.

Eingriffsintensität

Bei der datenschutzrechtlichen Bewertung, z.B. bei der Wahl einer Rechtsgrundlage, kann die Eingriffsintensität einer Datenverarbeitung zu beurteilen sein.

Frage 5

Haben Sie bereits Erfahrung gemacht mit Methoden, die die Menge und Art der personenbezogenen memorisierten Daten abschätzen, bzw. ob das verwendete KI-Modell personenbezogene Daten einer bestimmten Person enthält (z.B. Privacy Attacks/PII Extraction Attacks, etc.)? Wenn ja, wie bewerten Sie deren Aussagekraft und mögliche Einschränkungen?

Unsere Mitglieder nutzen unterschiedliche Methoden zur Risikobewertung, unter anderem:

- Red Teaming mit gezielten Prompts zur Testung auf PII-Ausgaben,
- Empirische Speicheranalysen nach Training und Fine-Tuning,
- Simulation von Angriffsszenarien,
- Probabilistische Methoden zur Abschätzung extrahierbarer Informationen (z. B. discoverable extraction)
- Loss Masking und synthetische Daten in der Feintuning-Phase, um zu verhindern, dass Modelle sensible Eingaben lernen oder reproduzieren.
- Kein Zugriff auf Original-Trainingsdaten während der Inferenz, um das Risiko von Memorisation-basierten Datenschutzverletzungen zu minimieren.

Diese Verfahren sind hilfreich, aber keineswegs vollständig. Sie sollten immer im Zusammenspiel mit technischen und organisatorischen Maßnahmen betrachtet werden.

Frage 6

Wie hoch ist die Menge personenbezogener memorisierter Daten in Ihnen bekannten KI-Modellen (in Prozent sowie Gesamtmenge Trainingsdaten)?

Eine pauschale Quantifizierung des Anteils personenbezogener Daten im Modell ist aus methodischen und technischen Gründen nicht möglich und auch nicht zielführend. KI-Modelle speichern keine Kopien von Trainingsdaten, sondern passen ihre internen Parameter so an, dass allgemeine Muster und Strukturen erlernt werden.

Ziel des Trainings ist es, Sprach- und Wissensmuster abzubilden, nicht jedoch einzelne betroffene Personen zu identifizieren, zu profilieren oder personenbezogene Daten anderweitig zu verwenden.

Daraus folgt: Weder ist es realisitisch noch sachgerecht, nachträglich Prozentwerte für einen vermeintlichen »Datenanteil« im Modell zu bestimmen. Eine solche Vorgehensweise würde im Gegenteil eine strukturierte Indizierung der Trainingsdaten erfordern, die ausschließlich der Identifikation, personenbezogener Daten dient und damit einen stärkeren Eingriff in die Privatsphäre darstellen würde, als es das Modelltraining selbst tut.

Stattdessen setzen unsere Mitglieder auf etablierte und wirksame Schutzmechanismen, um das Risiko einer personenbezogenen Wiedergabe zu minimieren:

- Transparenz über verwendete Datenquellen,
- Qualitätssicherung in der Datenvorverarbeitung,
- Dokumentation und Auditierung von Trainingsprozessen.
- Technische Kontrollen wie Vorverarbeitungs- sowie Ein- und Ausgabekontrollen

Betroffenenrechte

Die Black-Box-Architektur von KI-Modellen stellt eine Herausforderung für die wirksame Gewährleistung von Betroffenenrechten dar, insbesondere hinsichtlich der Ansprüche auf Auskunft, Berichtigung und Löschung gemäß Artikel 15 – 17 DSGVO.

Frage 7

Wie gehen Sie vor, wenn eine Person ihren Anspruch auf Auskunft über personenbezogene Daten, Berichtigung oder Löschung ihrer personenbezogenen Daten im KI-Modell geltend macht?

Bitkom erkennt die Bedeutung von Auskunfts-, Berichtigungs- und Löschrechten ausdrücklich an. Zugleich ist aber festzustellen, dass die klassische Umsetzung dieser Rechte auf die LLMs nicht ohne Weiteres selbst übertragbar ist. Die Wahrnehmung von Betroffenenrechten sollte vor allem auf Ebene der Trainingsdaten (vorher) und der Modellausgaben (nachher) erfolgen. Nachträgliche Schutzmechanismen wie Red Teaming und Filterung sind zentrale Bestandteile eines praxisnahen Ansatzes. Eine gezielte Änderung oder Löschung ist nur durch vollständiges Neu-Training denkbar und daher in der Regel nicht zumutbar.

- **Auskunft:** Es ist praktisch unmöglich, Informationen zu einer bestimmten Person im Modell zu lokalisieren. Bei Vorliegen zusätzlicher Identifikationsinformationen werden »Best Efforts« unternommen
- **Berichtigung:** Eine gezielte Änderung einzelner Inhalte ist nur durch vollständiges »Neu-Trainieren« denkbar. Dies ist aus Gründen der Verhältnismäßigkeit und Datenminimierung nicht zumutbar. Fokus auf Datenqualität vor dem Training und Einsatz nachträglicher Schutzmechanismen wie Red Teaming und Filterung.
- **Löschung:** Modelle speichern keine personenbezogenen Inhalte im klassischen Sinne. Eine funktionale Entsprechung kann jedoch durch Blockade bestimmter Prompts oder Filterung von Ausgaben erfolgen. Zwar ist eine Entfernung aus dem trainierten Modell technisch nicht möglich, jedoch können entsprechende Daten aus künftigen Trainingsläufen ausgeschlossen und deren Wiedergabe durch Filter verhindert werden.

Bitkom plädiert für einen praxisnahen, mehrstufigen Ansatz. Die Betroffenenrechte sollten auf den Ebenen der Trainingsdaten (vorher) und der Modellausgaben (nachher) wahrgenommen werden, nicht aber im Inneren des Modells.

Weitere Aspekte

Frage 8

Gibt es andere Aspekte, die aus Ihrer Perspektive beim Schutz der personenbezogenen Daten in KI-Modellen eine Rolle spielen?

Bitkom spricht sich für ein datenschutzrechtliches Rahmenwerk aus, das:

- risikobasiert ist und Kontexte berücksichtigt,
- technologieneutral bleibt und Innovation ermöglicht,
- zwischen Entwicklern und Anwendern differenziert (shared responsibility),
- sich auf die Anwendungsebene konzentriert, wo Datenschutzrisiken tatsächlich entstehen
- und interoperabel mit europäischen und internationalen Regeln ausgestaltet ist.

Darüber hinaus sind wichtig:

- **Sicherheits- und Vertraulichkeitsmaßnahmen** wie Verschlüsselung, Zugriffskontrollen und sichere Infrastrukturen (z. B. Trusted Research Environments).
- **Opt-out-Mechanismen** für die Nutzung personenbezogener Daten im Training sowie Feedback-Kanäle für die Meldung unzulässiger Ausgaben.
- **Detaillierte Dokumentation** von Datenquellen, Verarbeitungsschritten, Risikobewertungen und Datenschutz-Folgenabschätzungen.

Dabei ist zu beachten, dass die Verantwortlichkeiten für die Kontrolle je nach Rolle zwischen dem Entwickler und dem Betreiber des Modells aufgeteilt werden müssen – z. B. ist der Modellentwickler für das Training des Modells verantwortlich, während der Betreiber für etwaige Feinabstimmungen verantwortlich ist. Mit anderen Worten: Die Stelle, die die Kontrolle über die Verarbeitung personenbezogener Daten ausübt, trägt auch die Verantwortung für die Einhaltung der Datenschutzbestimmungen.

Wir regen darüber hinaus die Einrichtung einer dauerhaften, interdisziplinären Arbeitsgruppe an, in der Datenschutzaufsicht, Unternehmen, Wissenschaft und Zivilgesellschaft fortlaufend zu technischen Entwicklungen, Good Practices und offenen Rechtsfragen im Bereich KI und Datenschutz zusammenarbeiten.

Bitkom vertritt mehr als 2.200 Mitgliedsunternehmen aus der digitalen Wirtschaft. Sie generieren in Deutschland gut 200 Milliarden Euro Umsatz mit digitalen Technologien und Lösungen und beschäftigen mehr als 2 Millionen Menschen. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 500 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig, kreieren Content, bieten Plattformen an oder sind in anderer Weise Teil der digitalen Wirtschaft. 82 Prozent der im Bitkom engagierten Unternehmen haben ihren Hauptsitz in Deutschland, weitere 8 Prozent kommen aus dem restlichen Europa und 7 Prozent aus den USA. 3 Prozent stammen aus anderen Regionen der Welt. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem leistungsfähigen und souveränen Digitalstandort zu machen.

Herausgeber

Bitkom e.V.
Albrechtstr. 10 | 10117 Berlin

Ansprechpartner

Isabelle Stroot | Referentin Datenschutz
T 030 27576-228 | i.stroot@bitkom.org

Elena Kouremenou | Referentin Datenschutz
T 030 27576-425 | e.kouremenou@bitkom.org

Verantwortliches Bitkom-Gremium

AK Datenschutz