EDPS

2004 2024 EDPS

EUROPEAN DATA PROTECTION SUPERVISOR

The EU's independent data protection authority

03. Juni 2024

Generative KI und die EUDPR.

Erste EDSB-Orientierungen zur Gewährleistung der Einhaltung des Datenschutzes bei der Verwendung von generativen KI-Systemen.

edps.europa.eu

Diese Leitlinien des EDSB zu generativer künstlicher Intelligenz (generative KI) und dem Schutz personenbezogener Daten sollen den Organen, Einrichtungen, Ämtern und Agenturen der EU (EUI) praktische Ratschläge und Anweisungen zur Verarbeitung personenbezogener Daten bei der Nutzung generativer KI-Systeme geben, um ihnen die Einhaltung ihrer Datenschutzverpflichtungen zu erleichtern, die insbesondere in der Verordnung (EU) 2018/1725 festgelegt sind. Diese Leitlinien wurden so formuliert, dass sie möglichst viele Szenarien und Anwendungen abdecken und keine spezifischen technischen Maßnahmen vorschreiben. Stattdessen legen sie den Schwerpunkt auf die allgemeinen Grundsätze des Datenschutzes, die den EU-Institutionen helfen sollen, die Datenschutzanforderungen gemäß der Verordnung (EU) 2018/1725 zu erfüllen.

Diese Orientierungen sind ein erster Schritt hin zu detaillierteren Leitlinien, die die Entwicklung von generativen KI-Systemen und -Technologien, ihre Nutzung durch EU-Institutionen und die Ergebnisse der Überwachungs- und Aufsichtstätigkeiten des EDSB berücksichtigen werden.

Der EDSB gibt diese Orientierungen in seiner Rolle als Datenschutzaufsichtsbehörde und nicht in seiner neuen Rolle als AI-Aufsichtsbehörde gemäß dem AI-Gesetz heraus.

Das Gesetz über künstliche Intelligenz bleibt von diesen Orientierungen unberührt.

Einleitung und Umfang

1. Diese Leitlinien sollen den Organen, Einrichtungen, Ämtern und Agenturen der EU (EUI) einige praktische Ratschläge für die Verarbeitung personenbezogener Daten bei der Nutzung generativer KI-Systeme an die Hand geben, um sicherzustellen, dass sie ihre Datenschutzverpflichtungen einhalten, die insbesondere in der Verordnung (EU) 2018/1725 ("die Verordnung" oder EUDPR) festgelegt sind. Auch wenn in der Verordnung das Konzept der künstlichen Intelligenz (KI) nicht ausdrücklich erwähnt wird, ist die richtige Auslegung und Anwendung der Datenschutzgrundsätze von entscheidender Bedeutung, um eine vorteilhafte Nutzung dieser Systeme zu erreichen, die die Grundrechte und -freiheiten des Einzelnen nicht beeinträchtigt.

2. Der EDSB gibt diese Orientierungen in seiner Rolle als Datenschutzaufsichtsbehörde und nicht in seiner neuen Rolle als AI-Aufsichtsbehörde gemäß dem AI-Gesetz heraus.

3. Diese Leitlinien zielen nicht darauf ab, alle relevanten Fragen im Zusammenhang mit der Verarbeitung personenbezogener Daten beim Einsatz von generativen KI-Systemen, die von den Datenschutzbehörden zu prüfen sind, in allen Einzelheiten zu behandeln. Einige dieser Fragen sind noch offen, und weitere werden wahrscheinlich auftauchen, wenn die Nutzung dieser Systeme zunimmt und sich die Technologie so weiterentwickelt, dass ein besseres Verständnis der Funktionsweise generativer KI möglich wird.

4. Da sich die Technologie der künstlichen Intelligenz schnell weiterentwickelt, sind die spezifischen Werkzeuge und Mittel, die zur Bereitstellung dieser Art von Diensten verwendet werden, vielfältig und können sich sehr schnell ändern. Daher wurden diese Orientierungshilfen so formuliert, dass sie möglichst viele Szenarien und Anwendungen abdecken.

5. Diese Orientierungshilfen sind wie folgt gegliedert: Schlüsselfragen, gefolgt von ersten Antworten mit einigen vorläufigen Schlussfolgerungen und weiteren Erläuterungen oder Beispielen.

6. Diese ersten Orientierungshilfen dienen als erster Schritt zur Entwicklung umfassenderer Leitlinien. Im Laufe der Zeit werden diese Orientierungen aktualisiert, verfeinert und erweitert, um weitere Elemente zu berücksichtigen, die zur Unterstützung der EU-Institutionen bei der Entwicklung und Umsetzung dieser Systeme erforderlich sind. Eine solche Aktualisierung sollte spätestens zwölf Monate nach der Veröffentlichung dieses Dokuments erfolgen.

1. Was ist generative KI?

Generative KI ist ein Teilbereich der KI, bei dem spezielle Modelle des maschinellen Lernens verwendet werden, die eine breite und allgemeine Vielfalt von Ergebnissen erzeugen, die für eine Reihe von Aufgaben und Anwendungen geeignet sind, z. B. für die Erzeugung von Text, Bildern oder Audio. Konkret stützt sie sich auf die Verwendung so genannter Grundmodelle, die als Basismodelle für andere generative KI-Systeme dienen, die auf dieser Grundlage "feinabgestimmt" werden.

Ein Basismodell dient als Kernarchitektur oder Basis, auf der andere, speziellere Modelle aufgebaut werden. Diese Modelle werden auf der Grundlage verschiedener und umfangreicher Datensätze trainiert, einschließlich solcher, die öffentlich zugängliche Informationen enthalten. Sie können komplexe Strukturen wie Bilder, Audio, Video oder Sprache abbilden und können für bestimmte Aufgaben oder Anwendungen fein abgestimmt werden.

Große Sprachmodelle sind eine spezielle Art von Basismodellen, die auf riesigen Mengen von Textdaten (von Millionen bis Milliarden von Wörtern) trainiert werden und auf der Grundlage von Mustern und Beziehungen zwischen Wörtern und Sätzen natürlichsprachliche Antworten auf eine Vielzahl von Eingaben erzeugen können. Diese riesige Textmenge, die zum Trainieren des Modells verwendet wird, kann aus dem Internet, aus Büchern und anderen verfügbaren Quellen stammen. Einige bereits genutzte Anwendungen sind Systeme zur Codegenerierung, virtuelle Assistenten, Tools zur Erstellung von Inhalten, Sprachübersetzungsmaschinen, automatische Spracherkennung, medizinische Diagnosesysteme, wissenschaftliche Forschungsinstrumente usw.

Die Beziehung zwischen diesen Konzepten ist hierarchisch. Generative KI ist die umfassende Kategorie, die Modelle zur Erstellung von Inhalten umfasst. Ein Basismodell, z. B. ein großes Sprachmodell, dient als Grundarchitektur, auf der spezialisierte Modelle aufgebaut werden. Spezialisierte Modelle, die auf dem Grundmodell aufbauen, sind auf bestimmte Aufgaben oder Anwendungen ausgerichtet und nutzen das Wissen und die Fähigkeiten der Grundarchitektur.

Der Lebenszyklus eines generativen KI-Modells umfasst verschiedene Phasen, beginnend mit der Definition des Anwendungsfalls und des Anwendungsbereichs des Modells. In einigen Fällen ist es möglich, ein geeignetes Basismodell zu ermitteln, mit dem man beginnen kann, in anderen Fällen muss ein neues Modell von Grund auf erstellt werden. In der folgenden Phase wird das Modell mit relevanten Datensätzen für den Zweck des künftigen Systems trainiert, einschließlich der Feinabstimmung des Systems mit spezifischen, benutzerdefinierten Datensätzen, die für den Anwendungsfall des Modells erforderlich sind. Um das Training abzuschließen, werden spezielle Techniken eingesetzt, die menschliches Handeln erfordern, um genauere Informationen und kontrolliertes Verhalten zu gewährleisten. Die folgende Phase dient der Bewertung des Modells und der Festlegung von Metriken zur regelmäßigen Bewertung von Faktoren wie der Genauigkeit und der Anpassung des Modells an den Anwendungsfall. Schließlich werden die Modelle eingesetzt und implementiert, einschließlich einer kontinuierlichen Überwachung und regelmäßigen Bewertung anhand der in den vorangegangenen Phasen festgelegten Messgrößen.

Relevante Anwendungsfälle der generativen KI sind allgemeine verbraucherorientierte Anwendungen (wie ChatGPT und ähnliche Systeme, die es bereits in verschiedenen Versionen und Größen gibt[1]einschließlich solcher, die auf einem Mobiltelefon ausgeführt werden können). Es gibt auch Geschäftsanwendungen in bestimmten Bereichen, vortrainierte Modelle, Anwendungen, die auf vortrainierten Modellen basieren, die für einen bestimmten Einsatz in einem Bereich abgestimmt sind

[1] Die Größe eines Large Language Model wird in der Regel an der Anzahl der Parameter (Token) gemessen, die es enthält. Die Größe eines LLM-Modells ist wichtig, da einige Fähigkeiten erst dann zum Tragen kommen, wenn das Modell über bestimmte Grenzen hinaus wächst.

und schließlich Modelle, bei denen die gesamte Entwicklung, einschließlich des Ausbildungsprozesses, von der zuständigen Stelle durchgeführt wird.

Generative KI bietet, wie andere neue Technologien auch, Lösungen in verschiedenen Bereichen, die die menschlichen Fähigkeiten unterstützen und verbessern sollen. Sie bringt jedoch auch Herausforderungen mit sich, die sich auf die Grundrechte und -freiheiten auswirken können und die Gefahr bergen, dass sie unbemerkt bleiben, übersehen oder nicht angemessen berücksichtigt und bewertet werden.

Das Training eines Large Language Model (LLM) (und generell eines jeden maschinellen Lernmodells) ist ein iterativer, komplexer und ressourcenintensiver Prozess, der mehrere Stufen umfasst und

Techniken, die darauf abzielen, ein Modell zu schaffen, das in der Lage ist, als Reaktion auf Befehle (oder Aufforderungen) der Nutzer menschenähnlichen Text zu erzeugen. Der Prozess beginnt damit, dass das Modell auf riesigen Datensätzen trainiert wird, von denen die meisten in der Regel unbeschriftet sind und aus öffentlichen Quellen mithilfe von Web-Scraping-Technologien stammen (die Datenschutzbehörden haben bereits ihre Besorgnis geäußert und auf die wichtigsten Risiken für den Schutz der Privatsphäre und den Datenschutz hingewiesen, die mit der Verwendung öffentlich zugänglicher personenbezogener Daten verbunden sind). Danach werden LLMs - nicht in allen Fällen - durch überwachtes Lernen oder durch Techniken, die menschliches Handeln einbeziehen (wie Reinforcement Learning with Human Feedback (RLHF) oder Adversarial Testing via Domain-Experten), feinabgestimmt, um dem System zu helfen, Informationen und Kontext besser zu erkennen und zu verarbeiten, sowie um bevorzugte Antworten zu bestimmen, ob die Ausgabe bei der Beantwortung sensibler Fragen begrenzt werden soll und um sie mit den Werten der Entwickler in Einklang zu bringen (z. B. um zu vermeiden, dass schädliche oder giftige Ausgaben produziert werden). Sobald sie in Produktion sind, verwenden einige Systeme die durch die Interaktion mit den Nutzern gewonnenen Eingabedaten als neuen Trainingsdatensatz, um das Modell zu verfeinern.

---

2. Können EU-Institutionen generative KI nutzen?

Als EUI besteht grundsätzlich kein Hindernis für die Entwicklung, den Einsatz und die Nutzung generativer KI-Systeme bei der Erbringung öffentlicher Dienstleistungen, sofern die Vorschriften der EUI dies zulassen und alle geltenden rechtlichen Anforderungen erfüllt werden, insbesondere in Anbetracht der besonderen Verantwortung des öffentlichen Sektors für die Gewährleistung der uneingeschränkten Achtung der Grundrechte und -freiheiten des Einzelnen bei der Nutzung neuer Technologien.

In jedem Fall gilt die Verordnung in vollem Umfang, wenn der Einsatz generativer KI-Systeme die Verarbeitung personenbezogener Daten beinhaltet. Die Verordnung ist technologieneutral und gilt für alle Verarbeitungen personenbezogener Daten, unabhängig von den verwendeten Technologien und unbeschadet anderer Rechtsrahmen, insbesondere des KI-Gesetzes. Der Grundsatz der Rechenschaftspflicht erfordert, dass die Verantwortlichkeiten zwischen den verschiedenen Akteuren, die an der Lieferkette für generative KI-Modelle beteiligt sind, klar festgelegt und beachtet werden.

EU-Institutionen können ihre eigenen generativen KI-Lösungen entwickeln und einsetzen oder alternativ auf dem Markt verfügbare Lösungen für den eigenen Gebrauch nutzen. In beiden Fällen können EU-Institutionen auf Anbieter zurückgreifen, um alle oder einige der Elemente, die Teil des generativen KI-Systems sind, zu erhalten. In diesem Zusammenhang müssen die EU-Institutionen die spezifischen Rollen - für die Verarbeitung Verantwortlicher, Auftragsverarbeiter, gemeinsam für die Verarbeitung Verantwortlicher - für die spezifischen Verarbeitungsvorgänge und deren Auswirkungen in Bezug auf die Pflichten und Verantwortlichkeiten gemäß der Verordnung klar festlegen.

Angesichts des raschen Fortschritts der KI-Technologien müssen die EU-Institutionen sorgfältig abwägen, wann und wie sie generative KI verantwortungsvoll und nutzbringend für das öffentliche Wohl einsetzen. Alle Phasen des Lebenszyklus einer generativen KI-Lösung sollten im Einklang mit den geltenden rechtlichen Rahmenbedingungen, einschließlich der Verordnung, erfolgen, wenn das System die Verarbeitung personenbezogener Daten beinhaltet.

Die Begriffe vertrauenswürdige oder verantwortungsvolle KI beziehen sich auf die Notwendigkeit, sicherzustellen, dass KI-Systeme auf ethische und rechtliche Weise entwickelt werden. Dazu gehört die Berücksichtigung der unbeabsichtigten Folgen der Einsatz von KI-Technologie und die Notwendigkeit, einen risikobasierten Ansatz zu verfolgen, der alle Phasen des Lebenszyklus des Systems abdeckt. Dazu gehört auch Transparenz in Bezug auf die Verwendung von Trainingsdaten und deren Quellen, auf die Art und Weise, wie Algorithmen entworfen und implementiert werden, welche Art von Verzerrungen im System vorhanden sein könnten und wie mögliche Auswirkungen auf die Grundrechte und -freiheiten des Einzelnen behandelt werden. In diesem Zusammenhang müssen generative KI-Systeme transparent, erklärbar, konsistent, überprüfbar und zugänglich sein, um eine faire Verarbeitung personenbezogener Daten zu gewährleisten.

---

3. Woher weiß man, ob der Einsatz eines generativen KI-Systems die Verarbeitung personenbezogener Daten beinhaltet?

Die Verarbeitung personenbezogener Daten in einem generativen KI-System kann auf verschiedenen Ebenen und in verschiedenen Phasen seines Lebenszyklus erfolgen, ohne dass dies auf den ersten Blick ersichtlich sein muss. Dies gilt für die Erstellung der Trainingsdatensätze, für die Trainingsphase selbst, für die Ableitung neuer oder zusätzlicher Informationen, sobald das Modell erstellt und in Betrieb ist, oder einfach durch die Eingaben und Ausgaben des Systems, sobald es in Betrieb ist.

Wenn ein Entwickler oder ein Anbieter eines generativen KI-Systems behauptet, dass sein System keine personenbezogenen Daten verarbeitet (z. B. aus Gründen wie der angeblichen Verwendung anonymisierter Datensätze oder synthetischer Daten während des Entwurfs, der Entwicklung und der Tests), ist es von entscheidender Bedeutung, nach den spezifischen Kontrollen zu fragen, die eingerichtet wurden, um dies zu gewährleisten. Im Wesentlichen möchten die EU-Institutionen möglicherweise wissen, welche Schritte oder Verfahren der Anbieter anwendet, um sicherzustellen, dass das Modell keine personenbezogenen Daten verarbeitet.

Der EDSB hat bereits gewarnt[2] vor der Verwendung von Web-Scraping-Techniken zur Sammlung personenbezogener Daten, durch die Einzelpersonen die Kontrolle über ihre personenbezogenen Daten verlieren können, wenn diese ohne ihr Wissen, entgegen ihren Erwartungen und für Zwecke gesammelt werden, die sich von denen der ursprünglichen Sammlung unterscheiden. Der EDSB hat auch betont, dass die Verarbeitung personenbezogener Daten, die öffentlich zugänglich sind, weiterhin den EU-Datenschutzvorschriften unterliegt. In dieser Hinsicht könnte die Verwendung von Web-Scraping-Techniken zum Sammeln von Daten von Websites und deren Verwendung für Schulungszwecke nicht mit den einschlägigen Datenschutzgrundsätzen, einschließlich der Datenminimierung und des Grundsatzes der Richtigkeit, übereinstimmen, sofern keine Bewertung der Zuverlässigkeit der Quellen vorliegt.

[Vollmer: Alles klar... aber warum erfolgt hier keine praxisrelevante Einschätzung?]

Durch regelmäßige Überwachung und die Durchführung von Kontrollen in allen Phasen kann sichergestellt werden, dass keine Verarbeitung personenbezogener Daten erfolgt, wenn das Modell nicht dafür vorgesehen ist.

→ EUI-X, eine fiktive EU-Institution, erwägt die Anschaffung eines Produkts für die automatische Spracherkennung und -transkription. Nach Prüfung der verfügbaren Optionen hat sie sich auf die

die Möglichkeit, ein generatives KI-System zu verwenden, um diese Funktion zu erleichtern. In diesem speziellen Fall handelt es sich um ein System, das ein vortrainiertes Modell für die Spracherkennung und -übersetzung bietet. Da dieses Modell für die Transkription von Sitzungen unter Verwendung aufgezeichneter Sprachdateien verwendet wird, wurde festgestellt, dass die Verwendung dieses Modells die Verarbeitung personenbezogener Daten erfordert und daher die Einhaltung der Verordnung gewährleisten muss.

---

[2] *Stellungnahme 41/2023 vom 2. September 2023 zu dem Vorschlag für eine Verordnung über die Arbeitsmarktstatistik der Unternehmen in der Europäischen Union*

4. Welche Rolle spielen die behördlichen Datenschutzbeauftragten bei der Entwicklung und Einführung generativer KI-Systeme?

In Artikel 45 der Verordnung sind die Aufgaben des Datenschutzbeauftragten festgelegt. Die behördlichen Datenschutzbeauftragten informieren und beraten über die einschlägigen Datenschutzverpflichtungen, unterstützen die für die Verarbeitung Verantwortlichen bei der Überwachung der internen Einhaltung der Vorschriften, beraten auf Wunsch in Bezug auf Datenschutzfolgenabschätzungen und fungieren als Kontaktstelle für betroffene Personen und den EDSB.

Im Zusammenhang mit der Einführung generativer KI-Systeme, die personenbezogene Daten verarbeiten, durch EU-Institutionen muss sichergestellt werden, dass die behördlichen Datenschutzbeauftragten im Rahmen ihrer Rolle als unabhängige Berater und Unterstützer bei der Anwendung der Verordnung den Lebenszyklus des generativen KI-Systems, das die EU-Institution zu beschaffen, zu entwickeln oder einzuführen beabsichtigt, und dessen Funktionsweise genau verstehen. Dies bedeutet, dass Informationen darüber eingeholt werden müssen, wann und wie diese Systeme personenbezogene Daten verarbeiten und wie die Eingabe- und Ausgabemechanismen sowie die durch das Modell implementierten Entscheidungsprozesse funktionieren. Wie in der Verordnung dargelegt, ist es wichtig[3]die für die Verarbeitung Verantwortlichen bei der Durchführung von Datenschutz-Folgenabschätzungen zu beraten. Die für die Verarbeitung Verantwortlichen müssen dafür sorgen, dass alle Prozesse ordnungsgemäß dokumentiert werden und die Transparenz gewährleistet ist, wozu auch die Aktualisierung der Aufzeichnungen über die Verarbeitung und - als bewährtes Verfahren - die Durchführung einer spezifischen Bestandsaufnahme der generativen KI-gesteuerten Systeme und Anwendungen gehören. Schließlich sollte der DSB an der Überprüfung von Compliance-Fragen im Zusammenhang mit Vereinbarungen über die gemeinsame Nutzung von Daten mit Modellanbietern beteiligt werden.

Aus organisatorischer Sicht sollte die Einführung von generativen KI-Systemen im Einklang mit der Verordnung nicht von einer Person allein bewältigt werden. Es sollte ein kontinuierlicher Dialog zwischen allen Beteiligten über den gesamten Lebenszyklus des Produkts hinweg stattfinden. Daher sollten die für die Verarbeitung Verantwortlichen mit allen relevanten Funktionen innerhalb der Organisation zusammenarbeiten, insbesondere mit dem Datenschutzbeauftragten, dem Rechtsdienst, dem IT-Dienst und dem lokalen Beauftragten für die Sicherheit der Datenverarbeitung (LISO), um sicherzustellen, dass die EU-KI innerhalb der Parameter einer vertrauenswürdigen generativen KI und einer guten Datenverwaltung funktioniert und der Verordnung entspricht. Die Einrichtung einer KI-Taskforce, der auch der DSB angehört, und die Ausarbeitung eines Aktionsplans, einschließlich Sensibilisierungsmaßnahmen auf allen Ebenen der Organisation und der Ausarbeitung interner Leitlinien, können zur Erreichung dieser Ziele beitragen.

Als Beispiel für Vertragsklauseln hat die Europäische Kommission im Rahmen der Initiative "Procurement of AI Community" (Beschaffung der KI-Gemeinschaft) die für die Beschaffung von KI relevanten Akteure zusammengebracht.

Lösungen zur Entwicklung umfassender Mustervertragsklauseln für die Beschaffung von künstlicher Intelligenz durch öffentliche Organisationen. Es ist auch wichtig, die Standardvertragsklauseln zwischen für die Verarbeitung Verantwortlichen und Auftragsverarbeitern gemäß der Verordnung[1] zu berücksichtigen.

[3] *Artikel 39 Absatz 2 der Verordnung*

5. Eine EUI möchte generative KI-Systeme entwickeln oder einsetzen. Wann sollte eine DPIA durchgeführt werden?

Die Grundsätze des Datenschutzes durch Technik und durch datenschutzfreundliche Voreinstellungen[4] zielen darauf ab, personenbezogene Daten während des gesamten Lebenszyklus der Datenverarbeitung zu schützen, beginnend mit der Anfangsphase. Durch die Einhaltung dieses Grundsatzes der Verordnung, der auf einem risikoorientierten Ansatz beruht, können die Bedrohungen und Risiken, die generative KI mit sich bringen kann, berücksichtigt und im Voraus ausreichend gemindert werden. Entwickler und Anwender müssen unter Umständen ihre eigenen Risikobewertungen durchführen und die getroffenen Abhilfemaßnahmen dokumentieren.

Die Verordnung schreibt vor, dass eine Datenschutzfolgenabschätzung[5] vor jeder Verarbeitung durchgeführt werden muss, die wahrscheinlich ein hohes Risiko [6] für die Grundrechte und -freiheiten natürlicher Personen beinhalten. Die Verordnung weist darauf hin, wie wichtig es ist, eine solche Bewertung durchzuführen, wenn neue Technologien eingesetzt werden sollen oder wenn es sich um eine neue Art von Technologie handelt, für die der für die Verarbeitung Verantwortliche noch keine Bewertung durchgeführt hat, wie z. B. bei generativen KI-Systemen.

Der für die Verarbeitung Verantwortliche ist verpflichtet, bei der Durchführung einer Datenschutzfolgenabschätzung den Rat des Datenschutzbeauftragten (DSB) einzuholen. Aufgrund der Bewertung müssen geeignete technische und organisatorische Maßnahmen ergriffen werden, um die ermittelten Risiken unter Berücksichtigung der Verantwortlichkeiten, des Kontexts und des verfügbaren Stands der Technik einzudämmen.

Im Zusammenhang mit dem Einsatz generativer KI kann es angebracht sein, die Meinung der von dem System Betroffenen einzuholen, entweder der betroffenen Person selbst oder ihrer Vertreter im Bereich der geplanten Verarbeitung. Zusätzlich zu den Überprüfungen, mit denen beurteilt wird, ob die Datenschutzfolgenabschätzung ordnungsgemäß durchgeführt wird, müssen die Risikobewertungen regelmäßig überwacht und überprüft werden, da die Funktionsweise des Modells festgestellte Risiken verschärfen oder neue Risiken schaffen kann. Diese Risiken beziehen sich auf den Schutz personenbezogener Daten, aber auch auf andere Grundrechte und -freiheiten.

Alle an der DPIA beteiligten Akteure müssen sicherstellen, dass alle Entscheidungen und Maßnahmen ordnungsgemäß dokumentiert werden und den gesamten Lebenszyklus des generativen KI-Systems abdecken, einschließlich der Maßnahmen, die zur Beherrschung der Risiken ergriffen werden, und der anschließenden Überprüfungen, die durchzuführen sind.

Es liegt in der Verantwortung der EUI, die mit dem Einsatz von generativen KI-Systemen verbundenen Risiken angemessen zu managen. Die Datenschutzrisiken müssen während des gesamten Lebenszyklus des generativen KI-Systems ermittelt und behandelt werden. Dazu gehört eine regelmäßige und systematische Überwachung, um bei der Weiterentwicklung des Systems festzustellen, ob sich bereits identifizierte Risiken verschlechtern oder ob neue Risiken auftreten. Das

Verständnis der mit dem Einsatz generativer KI verbundenen Risiken ist noch nicht abgeschlossen, so dass ein wachsamer Ansatz erforderlich ist, um

[4] Artikel 27 der Verordnung

[5] Artikel 39 und 89 der Verordnung.

[6] Die Einstufung eines KI-Systems als "risikoreich" aufgrund seiner Auswirkungen auf die Grundrechte gemäß dem KI-Gesetz löst eine Vermutung des "hohen Risikos" gemäß der DSGVO, der EU-DSGVO und der LED aus, soweit personenbezogene Daten verarbeitet werden.

nicht identifizierte, neu auftretende Risiken. Wenn Risiken festgestellt werden, die nicht mit angemessenen Mitteln gemindert werden können, ist es an der Zeit, den EDSB zu konsultieren.

Der EDSB hat eine Vorlage erstellt, die es den für die Verarbeitung Verantwortlichen ermöglicht, zu beurteilen, ob sie eine Datenschutzfolgenabschätzung durchführen müssen [Anhang sechs zu Teil I des Toolkits zur Rechenschaftspflicht]. Darüber hinaus hat der EDSB
eine offene Liste von Verarbeitungen erstellt, die der Pflicht zur Durchführung einer Datenschutz-Folgenabschätzung unterliegen. Erforderlichenfalls führt der für die Verarbeitung Verantwortliche eine Überprüfung durch, um zu beurteilen, ob die Datenverarbeitung im Einklang mit der Datenschutz-Folgenabschätzung durchgeführt wird, zumindest wenn sich die mit den Verarbeitungen verbundenen Risiken ändern. Wenn die für die Verarbeitung Verantwortlichen nach der Datenschutz-Folgenabschätzung nicht sicher sind, ob die Risiken angemessen gemindert werden, sollten sie eine vorherige Konsultation mit dem EDSB durchführen.

---

6. Wann ist die Verarbeitung personenbezogener Daten bei der Konzeption, Entwicklung und Validierung von generativen KI-Systemen rechtmäßig?

Die Verarbeitung personenbezogener Daten in generativen KI-Systemen kann sich über den gesamten Lebenszyklus des Systems erstrecken und umfasst alle Verarbeitungstätigkeiten im Zusammenhang mit der Erhebung von Daten, dem Training, der Interaktion mit dem System und der Generierung von Inhalten durch das System. Zu den sammlungs- und trainingsbezogenen Verarbeitungstätigkeiten gehört die Beschaffung von Daten aus öffentlich zugänglichen Quellen im Internet, direkt, von Dritten oder aus den eigenen Dateien der EUKI. Personenbezogene Daten können von dem generativen KI-Modell auch direkt von den Nutzern über die Eingaben in das System oder durch Ableitung neuer Informationen gewonnen werden. Im Zusammenhang mit generativen KI-Systemen beruhen das Training und die Nutzung der Systeme in der Regel auf einer systematischen und groß angelegten Verarbeitung personenbezogener Daten, in vielen Fällen ohne das Bewusstsein der Personen, deren Daten verarbeitet werden.

Die Verarbeitung personenbezogener Daten durch die EU-Institutionen ist rechtmäßig, wenn mindestens einer der in der Verordnung aufgeführten Rechtmäßigkeitsgründe[7] in der Verordnung aufgeführt ist. Damit die Verarbeitung besonderer Kategorien personenbezogener Daten rechtmäßig ist, muss darüber hinaus eine der in der Verordnung aufgeführten Ausnahmen[8] eine der in der Verordnung aufgeführten Ausnahmen gelten. Wenn die Verarbeitung zur Wahrnehmung einer Aufgabe erfolgt, die im öffentlichen Interesse liegt[9] oder für die Erfüllung einer rechtlichen Verpflichtung erforderlich ist[10] der der für die Verarbeitung Verantwortliche unterliegt, muss die Rechtsgrundlage für die Verarbeitung im EU-Recht verankert sein. Außerdem muss das EU-Recht, auf das verwiesen wird, klar und eindeutig sein, und seine Anwendung muss für die betroffenen Personen vorhersehbar sein, entsprechend den Anforderungen der Charta der Grundrechte der Europäischen Union und der Europäischen Konvention zum Schutze der Menschenrechte und Grundfreiheiten.

Wenn eine Rechtsgrundlage zu einem schwerwiegenden Eingriff in die Grundrechte auf Datenschutz und Schutz der Privatsphäre führt, besteht außerdem ein größerer Bedarf an klaren und präzisen Regeln für den Anwendungsbereich und die Anwendung der Maßnahme sowie für die begleitenden Garantien. Je stärker der Eingriff ist, desto robuster und detaillierter sollten die Vorschriften und Garantien sein. Wenn man sich auf interne Vorschriften stützt, sollten diese internen Vorschriften den Umfang des Eingriffs in das Recht auf den Schutz personenbezogener Daten genau definieren, indem der Zweck der Verarbeitung, die Kategorien der betroffenen Personen, die Kategorien der zu verarbeitenden personenbezogenen Daten, der für die Verarbeitung Verantwortliche und die Auftragsverarbeiter sowie die Aufbewahrungsfristen angegeben werden, zusammen mit einer Beschreibung der konkreten Mindestgarantien und Maßnahmen zum Schutz der Rechte natürlicher Personen.

Die Verwendung der Einwilligung[11] als Rechtsgrundlage kann unter bestimmten Umständen im Zusammenhang mit dem Einsatz generativer KI-Systeme gelten. Einholung der Einwilligung[12] Damit die Einwilligung nach der Verordnung gültig ist, muss sie alle rechtlichen Anforderungen erfüllen, einschließlich des Erfordernisses einer klaren bestätigenden Handlung der Person, sie muss frei, spezifisch, in Kenntnis der Sachlage und unzweideutig erteilt werden. In Anbetracht der Art und Weise, wie generative KI-Systeme trainiert werden, und der Quellen von Trainingsdaten, einschließlich öffentlich zugänglicher Informationen, muss die Verwendung der Einwilligung als

solche sorgfältig geprüft werden, auch im Zusammenhang mit ihrer Verwendung durch öffentliche Stellen.

---

[7] Artikel 5 der Verordnung.
[8] Artikel 10 Absatz 2 der Verordnung.
[9] Artikel 5(1)(a) der Verordnung.
[10] Artikel 5 Absatz 1 Buchstabe b der Verordnung
[11] Artikel 5 Absatz 1 Buchstabe d) und Artikel 7 der Verordnung.
[12] EDPB-Leitlinien 05/2020 zur Zustimmung gemäß der Verordnung 2016/679, abrufbar unter https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf

Einrichtungen, wie z. B. EU-Institutionen. Wird die Einwilligung widerrufen, bleiben außerdem alle Datenverarbeitungsvorgänge, die auf dieser Einwilligung beruhen und vor dem Widerruf - und im Einklang mit der Verordnung - stattgefunden haben, rechtmäßig. In diesem Fall muss der für die Verarbeitung Verantwortliche jedoch die betreffenden Verarbeitungen einstellen. Gibt es keine andere rechtmäßige Grundlage, die die Verarbeitung von Daten rechtfertigt, muss der für die Verarbeitung Verantwortliche die betreffenden Daten löschen.

Anbieter von generativen KI-Modellen können ein berechtigtes Interesse gemäß der EU-Datenschutzgrundverordnung[13] (GDPR) als Rechtsgrundlage für die Datenverarbeitung heranziehen, insbesondere im Hinblick auf die Erhebung von Daten, die zur Entwicklung des Systems verwendet werden, einschließlich der Schulungs- und Validierungsprozesse. Der Gerichtshof der Europäischen Union (EuGH) hat entschieden[14] dass der Rückgriff auf das berechtigte Interesse drei kumulative Voraussetzungen für die Rechtmäßigkeit der Verarbeitung personenbezogener Daten auf dieser Rechtsgrundlage enthält. Erstens, die Verfolgung eines berechtigten Interesses durch den für die Verarbeitung Verantwortlichen oder einen Dritten; zweitens, die Notwendigkeit der Verarbeitung personenbezogener Daten für die Zwecke des verfolgten berechtigten Interesses; und drittens, dass die Interessen oder Grundfreiheiten und Rechte der vom Datenschutz betroffenen Person nicht Vorrang vor dem berechtigten Interesse des für die Verarbeitung Verantwortlichen oder eines Dritten haben. In dieser Hinsicht tragen die EU-Institutionen eine besondere Verantwortung dafür, zu überprüfen, ob die Anbieter generativer KI-Systeme die Bedingungen für die Anwendung dieser Rechtsgrundlage eingehalten haben, wobei die spezifischen Bedingungen der Verarbeitung durch diese Systeme zu berücksichtigen sind.

Als für die Verarbeitung personenbezogener Daten Verantwortliche sind die EU-Institutionen für die von ihnen veranlassten Übermittlungen personenbezogener Daten und für die in ihrem Namen innerhalb und außerhalb des Europäischen Wirtschaftsraums durchgeführten Übermittlungen verantwortlich. Diese Übermittlungen können nur erfolgen, wenn die betreffende EU-Institution sie angeordnet oder zugelassen hat oder wenn solche Übermittlungen nach EU-Recht oder nach dem Recht der EU-Mitgliedstaaten erforderlich sind. Übermittlungen können im Zusammenhang mit der Entwicklung oder dem Einsatz generativer KI-Systeme auf verschiedenen Ebenen stattfinden, unter anderem wenn EUI auf Systeme zurückgreifen, die auf Cloud-Diensten basieren, oder wenn sie in bestimmten Fällen personenbezogene Daten bereitstellen müssen, die zum Trainieren, Testen oder Validieren eines Modells verwendet werden. In jedem Fall müssen diese Datenübermittlungen mit den Bestimmungen in Kapitel V[15] der Verordnung entsprechen, wobei auch die anderen Bestimmungen der Verordnung zu beachten sind, und mit dem ursprünglichen Zweck der Datenverarbeitung vereinbar sein.

Die Verarbeitung personenbezogener Daten im Zusammenhang mit generativen KI-Systemen erfordert eine Rechtsgrundlage im Einklang mit der Verordnung. Beruht die Datenverarbeitung auf einer rechtlichen Verpflichtung oder der Ausübung öffentlicher Gewalt, muss diese Rechtsgrundlage klar und präzise im EU-Recht festgelegt sein. Die Verwendung der Einwilligung als Rechtsgrundlage muss sorgfältig geprüft werden, um sicherzustellen, dass sie die Anforderungen der Verordnung erfüllt, damit sie gültig ist.

[13] Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung)

[14] Urteil vom 4. Juli 2023, Meta Platforms u. a. (Allgemeine Nutzungsbedingungen eines sozialen Netzwerks), C-252/21, EU:C:2023:537, Randnr. 106 und die dort zitierte Rechtsprechung

[15] Artikel 46 bis 51 der Verordnung

So heißt es beispielsweise in der GPA-Entschließung zu generativen Systemen der künstlichen Intelligenz, dass die Entwickler, Anbieter und Betreiber generativer KI, sofern dies nach den einschlägigen Rechtsvorschriften erforderlich ist

Systeme müssen zu Beginn die Rechtsgrundlage für die Verarbeitung personenbezogener Daten im Zusammenhang mit a) der Erhebung von Daten zur Entwicklung generativer KI-Systeme, b) Trainings-, Validierungs- und Testdatensätzen zur Entwicklung oder Verbesserung generativer KI-Systeme, c) der Interaktion von Personen mit generativen KI-Systemen und d) von generativen KI-Systemen erzeugten Inhalten angeben.

7. Wie kann der Grundsatz der Datenminimierung beim Einsatz generativer KI-Systeme gewährleistet werden?

Der Grundsatz der Datenminimierung bedeutet, dass die für die Verarbeitung Verantwortlichen dafür sorgen müssen, dass die verarbeiteten personenbezogenen Daten den Zwecken entsprechen, für die sie verarbeitet werden, dafür erheblich sind und auf das für die Zwecke, für die sie verarbeitet werden, erforderliche Maß beschränkt sind. Es besteht der Irrglaube, dass der Grundsatz der Datenminimierung[16]im Zusammenhang mit der künstlichen Intelligenz keinen Platz hat. Die für die Datenverarbeitung Verantwortlichen sind jedoch verpflichtet, die Erhebung und sonstige Verarbeitung personenbezogener Daten auf das für die Zwecke der Verarbeitung erforderliche Maß zu beschränken und eine wahllose Verarbeitung personenbezogener Daten zu vermeiden. Diese Verpflichtung gilt für den gesamten Lebenszyklus des Systems, einschließlich der Test-, Abnahme- und Produktionsphase. Personenbezogene Daten sollten nicht wahllos gesammelt und verarbeitet werden. Die EU-Institutionen müssen sicherstellen, dass das an der Entwicklung generativer KI-Modelle beteiligte Personal die verschiedenen technischen Verfahren kennt, die zur Minimierung der Verwendung personenbezogener Daten zur Verfügung stehen, und dass diese in allen Phasen der Entwicklung gebührend berücksichtigt werden.

Die EU-Institutionen sollten Modelle entwickeln und verwenden, die mit hochwertigen Datensätzen trainiert wurden, die auf die personenbezogenen Daten beschränkt sind, die zur Erfüllung des Zwecks der Verarbeitung erforderlich sind. Auf diese Weise sollten diese Datensätze im Rahmen geeigneter Data-Governance-Verfahren, einschließlich einer regelmäßigen und systematischen Überprüfung des Inhalts, gut beschriftet und kuratiert werden. Datensätze und Modelle müssen von einer Dokumentation über ihre Struktur, Pflege und beabsichtigte Verwendung begleitet werden. Bei der Nutzung von Systemen, die von Drittanbietern entwickelt oder betrieben werden, sollten die EU-Institutionen in ihren Bewertungen Überlegungen zum Grundsatz der Datenminimierung anstellen.

Die Verwendung großer Datenmengen zum Trainieren eines generativen KI-Systems bedeutet nicht zwangsläufig eine größere Effizienz oder bessere Ergebnisse. Der sorgfältige Entwurf gut strukturierter Datensätze, die in Systemen verwendet werden, bei denen Qualität vor Quantität geht, die einem ordnungsgemäß überwachten Trainingsprozess folgen und regelmäßig überwacht werden, ist von entscheidender Bedeutung, um die erwarteten Ergebnisse zu erzielen, nicht nur im Hinblick auf die Datenminimierung, sondern auch in Bezug auf die Qualität der Ergebnisse und die Datensicherheit.

EUI-X beabsichtigt, ein KI-System zu trainieren, das in der Lage ist, bei Aufgaben im Zusammenhang mit der Softwareentwicklung und Programmierung zu helfen. Zu diesem Zweck möchten sie ein Werkzeug zur Generierung von Inhalten verwenden, das

werden über die Konten der einzelnen IT-Mitarbeiter verfügbar sein. Die EUI-X muss vor der Schulung des Algorithmus Überlegungen anstellen, um sicherzustellen, dass keine personenbezogenen Daten verarbeitet werden, die für den beabsichtigten Zweck nicht nützlich sind. Sie können beispielsweise eine statistische Analyse durchführen, um nachzuweisen, dass eine minimale Datenmenge erforderlich ist, um das gewünschte Ergebnis zu erzielen. Außerdem müssen sie prüfen und begründen, ob sie besondere Kategorien personenbezogener

Daten verarbeiten werden. Außerdem müssen sie die Art der Daten prüfen (d. h. synthetisierte, anonymisierte oder pseudonymisierte Daten). Schließlich müssen sie alle relevanten technischen und rechtlichen Elemente der verwendeten Datenquellen überprüfen, einschließlich ihrer Rechtmäßigkeit, Transparenz und Genauigkeit.

[16]Gemäß Artikel 4 Absatz 1 Buchstabe c der Verordnung müssen personenbezogene Daten, die verarbeitet werden, den Zwecken entsprechen, für die sie verarbeitet werden, dafür erheblich sein und auf das erforderliche Maß beschränkt werden.

8. Respektieren generative KI-Systeme den Grundsatz der Datengenauigkeit?

Generative KI-Systeme können in allen Phasen ihres Lebenszyklus, insbesondere in der Trainingsphase, große Mengen an Informationen, einschließlich personenbezogener Daten, verwenden.

Der Grundsatz der Datenrichtigkeit [17] verlangt, dass die Daten richtig und aktuell sind, während der für die Datenverarbeitung Verantwortliche verpflichtet ist, unrichtige Daten zu aktualisieren oder zu löschen. Die für die Verarbeitung Verantwortlichen müssen die Datengenauigkeit in allen Phasen der Entwicklung und Nutzung eines generativen KI-Systems sicherstellen. Sie müssen nämlich die notwendigen Maßnahmen zur Integration des "eingebauten Datenschutzes" ergreifen, die dazu beitragen, die Datengenauigkeit in allen Phasen zu erhöhen.

Dazu gehört die Überprüfung der Struktur und des Inhalts der für das Training der Modelle verwendeten Datensätze, einschließlich derjenigen, die von Dritten stammen oder erhalten wurden. Ebenso wichtig ist die Kontrolle über die Ausgabedaten, einschließlich der vom Modell gezogenen Schlüsse, was eine regelmäßige Überwachung dieser Informationen erfordert, einschließlich menschlicher Aufsicht. Die Entwickler sollten Validierungssätze[18] während des Trainings und separate Testsätze für die abschließende Bewertung verwenden, um eine Einschätzung der Leistung des Systems zu erhalten. Obwohl im Allgemeinen nicht datenschutzorientiert, können Metriken zur statistischen Genauigkeit (die Fähigkeit von Modellen, auf der Grundlage der Daten, auf denen sie trainiert wurden, korrekte Ergebnisse oder Vorhersagen zu produzieren), sofern verfügbar, einen Indikator für die Genauigkeit der vom Modell verwendeten Daten sowie für die erwartete Leistung bieten.

Wenn EU-Institutionen ein generatives KI-System oder von Dritten bereitgestellte Trainings-, Test- oder Validierungsdatensätze verwenden, müssen vertragliche Zusicherungen und Unterlagen über die Verfahren eingeholt werden, mit denen die Genauigkeit der für die Entwicklung des Systems verwendeten Daten sichergestellt wird. Dazu gehören Verfahren zur Datenerfassung und -aufbereitung, wie z. B. Beschriftung, Kennzeichnung, Bereinigung, Anreicherung und Aggregation, sowie die Ermittlung möglicher Lücken und Probleme, die die Genauigkeit beeinträchtigen können. Die technische Dokumentation und die Benutzerdokumentation des Systems, einschließlich der Modellkarten, sollten den für das System Verantwortlichen in die Lage versetzen, regelmäßig geeignete Kontrollen und Maßnahmen durchzuführen, um den Grundsatz der Genauigkeit zu gewährleisten. Dies ist umso wichtiger, als Modelle, selbst wenn sie mit repräsentativen Daten von hoher Qualität trainiert wurden, Ergebnisse erzeugen können, die ungenaue oder falsche Informationen enthalten, einschließlich personenbezogener Daten, den so genannten "Halluzinationen".

Trotz der Bemühungen, die Datengenauigkeit zu gewährleisten, sind generative KI-Systeme immer noch anfällig für ungenaue Ergebnisse, die sich auf die Grundrechte und -freiheiten des Einzelnen auswirken können.

Während die Anbieter fortschrittliche Trainingssysteme implementieren, um sicherzustellen, dass die Modelle genaue Daten verwenden und generieren, sollten die EU-Institutionen die Datengenauigkeit während des gesamten Lebenszyklus der generativen KI-Systeme sorgfältig bewerten und den Einsatz solcher Systeme in Betracht ziehen, wenn die Genauigkeit nicht aufrechterhalten werden kann.

[17] Artikel 4 Absatz 1 Buchstabe d der Verordnung.
[18] Validierungssätze werden zur Feinabstimmung der Parameter eines Modells und zur Bewertung seiner Leistung verwendet.

→ EUI-X hat auf Anraten des DSB beschlossen, dass die Ergebnisse des ASR-Modells bei der Transkription von offiziellen Sitzungen und Anhörungen einer Validierung durch qualifizierte Mitarbeiter des EUI. In Fällen, in denen das Modell für andere, weniger sensible Sitzungen verwendet wird, wird die Transkription immer mit einem klaren Hinweis versehen, dass es sich um ein von einem KI-System erstelltes Dokument handelt. EUI-X hat auf der obersten Führungsebene eine Richtlinie für die Verwendung des Modells sowie Datenschutzhinweise im Einklang mit der Verordnung ausgearbeitet und genehmigt, in denen die Zustimmung der Personen sowohl für die Aufzeichnung ihrer Stimme während der Sitzungen als auch für die Verarbeitung durch das Transkriptionssystem eingeholt wird. Vor der Einführung des KI-Systems durch das EUI wurde auch eine Datenschutzprüfung durchgeführt.

9. Wie können Personen über die Verarbeitung personenbezogener Daten informiert werden, wenn EU-Institutionen generative KI-Systeme verwenden?

Eine angemessene Informations- und Transparenzpolitik kann dazu beitragen, die Risiken für den Einzelnen zu mindern und die Einhaltung der Anforderungen der Verordnung zu gewährleisten, insbesondere durch die Bereitstellung detaillierter Informationen darüber, wie, wann und warum EU-Institutionen personenbezogene Daten in generativen KI-Systemen verarbeiten. Dies erfordert umfassende Informationen - die von den Entwicklern bzw. Zulieferern bereitgestellt werden müssen - über die in den verschiedenen Entwicklungsphasen durchgeführten Verarbeitungstätigkeiten, einschließlich der Herkunft der Datensätze, des Kuratierungs-/Kennzeichnungsverfahrens sowie aller damit verbundenen Verarbeitungen. Insbesondere sollten die EU-Institutionen sicherstellen, dass sie angemessene und relevante Informationen über die von ihren Anbietern oder Lieferanten verwendeten Datensätze erhalten und dass diese Informationen zuverlässig sind und regelmäßig aktualisiert werden. Bestimmte Systeme (z. B. Chatbots) können besondere Transparenzanforderungen erfordern, einschließlich der Information von Personen, dass sie mit einem KI-System ohne menschliches Eingreifen interagieren.

Da das Recht auf Information[19] die Verpflichtung beinhaltet, Personen in Fällen von Profilerstellung und automatisierten Entscheidungen aussagekräftige Informationen über die Logik solcher Entscheidungen sowie über deren Bedeutung und mögliche Folgen für die Personen zu geben, ist es wichtig, dass die EUI aktuelle Informationen nicht nur über die Funktionsweise der verwendeten Algorithmen, sondern auch über die verarbeiteten Datensätze bereithält. Diese Verpflichtung sollte generell auf Fälle ausgedehnt werden, in denen das Entscheidungsverfahren zwar nicht vollständig automatisiert ist, aber vorbereitende Handlungen auf der Grundlage einer automatisierten Verarbeitung umfasst.

Die EU-Institutionen müssen den Personen alle in der Verordnung geforderten Informationen zur Verfügung stellen, wenn sie generative KI-Systeme einsetzen, die personenbezogene Daten verarbeiten. Die Informationen, die den Personen zur Verfügung gestellt werden, müssen bei Bedarf aktualisiert werden, damit sie angemessen informiert sind und die Kontrolle über ihre eigenen Daten behalten.

→ EU-X bereitet einen Chatbot vor, der Personen beim Zugriff auf bestimmte Bereiche seiner Website unterstützen soll. Die betroffenen für die Verarbeitung Verantwortlichen haben mit Beratung durch den DSB ein Datenschutzkonzept erstellt
Hinweis, der auf der EU-X-Website verfügbar ist. Der Hinweis enthält Informationen über den Zweck der Verarbeitung, die Rechtsgrundlage, die Identität des für die Verarbeitung Verantwortlichen und die Kontaktdaten des DSB, die Empfänger der Daten, die Kategorien der erhobenen personenbezogenen Daten, die Aufbewahrung der Daten sowie darüber, wie die Rechte des Einzelnen ausgeübt werden können. Der Hinweis enthält auch Informationen über die Funktionsweise des Systems und über die mögliche Verwendung der Eingaben des Nutzers zur Verfeinerung der Chat-Funktion. EU-X verwendet die Einwilligung als Rechtsgrundlage, aber die Nutzer können ihre Einwilligung jederzeit widerrufen. In der Mitteilung wird auch klargestellt, dass Minderjährige den Chatbot nicht nutzen dürfen. Vor der Nutzung des Chatbots der EUI können Einzelpersonen nach dem Lesen des

Datenschutzhinweises ihre Zustimmung geben.

[19] Artikel 14 der Verordnung.

10. Was ist mit automatisierten Entscheidungen im Sinne von Artikel 24 der Verordnung?

Der Einsatz eines generativen KI-Systems bedeutet nicht unbedingt eine automatisierte Entscheidungsfindung[20] im Sinne der Verordnung. Es gibt jedoch generative KI-Systeme, die Informationen für die Entscheidungsfindung liefern, die mit automatisierten Mitteln gewonnen wurden und Profiling und/oder individuelle Bewertungen beinhalten. Je nach der Verwendung solcher Informationen bei der endgültigen Entscheidung durch eine öffentliche Stelle können EU-KI in den Anwendungsbereich von Artikel 24 der Verordnung fallen, so dass sie sicherstellen müssen, dass individuelle Garantien gewährleistet sind, einschließlich zumindest des Rechts, ein menschliches Eingreifen seitens des für die Verarbeitung Verantwortlichen zu erwirken, seinen oder ihren Standpunkt darzulegen und die Entscheidung anzufechten.

Bei der Verwaltung von KI-Entscheidungsinstrumenten müssen die EU-Institutionen sorgfältig prüfen, wie sie sicherstellen können, dass das Recht auf menschliches Eingreifen ordnungsgemäß umgesetzt wird. Dies ist von größter Bedeutung, wenn EU-Institutionen autonome KI-Agenten einsetzen, die ohne menschliches Eingreifen oder Anleitung Aufgaben erfüllen und Entscheidungen treffen können.

Die EU-Institutionen müssen sehr genau darauf achten, welches Gewicht die vom System bereitgestellten Informationen in den letzten Schritten des Entscheidungsverfahrens haben und ob sie einen entscheidenden Einfluss auf die endgültige Entscheidung des für die Verarbeitung Verantwortlichen haben. Es ist wichtig, die einzigartigen Risiken und potenziellen Schäden generativer KI-Systeme im Zusammenhang mit der automatisierten Entscheidungsfindung anzuerkennen, insbesondere für schutzbedürftige Bevölkerungsgruppen und Kinder[21].

Wenn generative KI-Systeme zur Unterstützung von Entscheidungsverfahren geplant sind, müssen die EU-Institutionen sorgfältig abwägen, ob sie sie einsetzen, wenn ihre Verwendung Fragen hinsichtlich ihrer Rechtmäßigkeit oder ihres Potenzials für unfaire, unethische oder diskriminierende Entscheidungen aufwirft.

[20] Artikel 24 der Verordnung.
[21] Globale Datenschutzversammlung (GPA) (2023). Entschließung über generative Systeme der künstlichen Intelligenz.

→ EUI-X erwägt den Einsatz eines KI-Systems für das erste Screening und die Filterung von Bewerbungen. Dienstanbieter C hat ein generatives KI-System angeboten, das eine Analyse der
die formalen Anforderungen und eine automatische Bewertung der Bewerbungen, die Punkte und Vorschläge für die Auswahl der Bewerber für die nächste Phase liefert. Nach Einsichtnahme in die Unterlagen zu dem Modell, einschließlich der verfügbaren Messwerte zur statistischen Genauigkeit (Messwerte zur Präzision und Sensitivität des Modells), und in Anbetracht des möglichen Vorhandenseins von Verzerrungen in dem Modell hat EUI-X beschlossen, das System zumindest so lange nicht zu verwenden, bis es eindeutige Hinweise darauf gibt, dass das Risiko von Verzerrungen beseitigt wurde und sich die Messwerte zur Präzision bei der Analyse der formalen Anforderungen verbessern.

Wenn ein solches System als "zweckdienlich" (d. h. für die Überprüfung der Bewerber) und mit allen für das EUI geltenden Vorschriften vereinbar angesehen wird, sollte das EUI in jedem Fall nachweisen können, dass es sich auf eine der Ausnahmen gemäß Artikel 24 Absatz 2 der Verordnung berufen kann und dass es geeignete Maßnahmen zum Schutz der Rechte des Einzelnen ergriffen hat, einschließlich des Rechts, von dem EUI ein menschliches Eingreifen zu erhalten, seinen Standpunkt darzulegen und die Entscheidung anzufechten (z. B. die Nichtwählbarkeit).

Das EUI muss gemäß Artikel 15 Absatz 2 Buchstabe f der Verordnung Informationen über die Logik des KI-Systems sowie über die voraussichtlichen Folgen einer solchen Verarbeitung für die betroffene Person bereitstellen, wenn die Daten bei der betroffenen Person erhoben werden. Vor dem Einsatz des KI-Systems durch das EUI muss ebenfalls eine Datenschutzfolgenabschätzung durchgeführt werden.

Die EUI-X kann beschließen, anstelle eines generativen KI-Systems ein "einfacheres" automatisiertes Online-Tool für das Screening von Stellenbewerbungen zu verwenden (z. B. ein IT-Tool, das automatisch die Anzahl der Jahre der Berufserfahrung oder der Ausbildung überprüft).

11. Wie kann bei der Verwendung generativer KI-Systeme eine faire Verarbeitung sichergestellt und Verzerrungen vermieden werden?

Im Allgemeinen neigen Lösungen der künstlichen Intelligenz dazu, bestehende menschliche Voreingenommenheiten zu verstärken und möglicherweise neue einzubauen, was zu neuen ethischen Herausforderungen und Risiken für die Einhaltung von Rechtsvorschriften führen kann. Voreingenommenheit kann in jeder Phase der Entwicklung eines generativen KI-Systems durch das Training von Datensätzen, die Algorithmen oder durch die Menschen, die das System entwickeln oder nutzen, entstehen. Voreingenommenheit in generativen KI-Systemen kann zu erheblichen nachteiligen Folgen für die Grundrechte und -freiheiten des Einzelnen führen, einschließlich unfairer Verarbeitung und Diskriminierung, insbesondere in Bereichen wie der Personalverwaltung, der medizinischen Versorgung im öffentlichen Gesundheitswesen und der Erbringung sozialer Dienstleistungen, wissenschaftlichen und technischen Verfahren, politischen und kulturellen Prozessen, dem Finanzsektor, der Umwelt und Ökosystemen sowie der öffentlichen Verwaltung.

Hauptquellen für Verzerrungen können unter anderem bestehende Muster in den Trainingsdaten, fehlende Informationen (vollständig oder teilweise) über die betroffene Bevölkerung, die Einbeziehung oder Auslassung von Variablen und Daten, die nicht in den Datensätzen enthalten sein sollten, methodische Fehler oder sogar Verzerrungen sein, die durch die Überwachung eingeführt werden.

Es ist von entscheidender Bedeutung, dass die Datensätze, die zur Erstellung und zum Training von Modellen verwendet werden, eine angemessene und faire Darstellung der realen Welt gewährleisten - ohne Verzerrungen, die den potenziellen Schaden für Einzelpersonen oder Kollektive, die in den Trainingsdatensätzen nicht gut repräsentiert sind, erhöhen können - und dass gleichzeitig Mechanismen zur Rechenschaftspflicht und Überwachung eingeführt werden, die eine kontinuierliche Überwachung ermöglichen, um das Auftreten von Verzerrungen zu verhindern, die sich auf Einzelpersonen auswirken, und um diese Verhaltensweisen zu korrigieren. Dazu gehört auch, dass die Verarbeitungstätigkeiten nachvollziehbar und überprüfbar sind[22] und dass die EU-Institutionen unterstützende Unterlagen aufbewahren. In dieser Hinsicht ist es wichtig, dass die EU-Institutionen technische Dokumentationsmodelle annehmen und umsetzen, was besonders wichtig sein kann, wenn die Modelle mehrere Datensätze verwenden und/oder verschiedene Datenquellen kombinieren.

Die Anbieter generativer KI-Systeme versuchen, Verzerrungen in ihren Systemen zu erkennen und abzuschwächen. Die EU-Institutionen kennen jedoch ihren Geschäftsfall am besten und sollten prüfen und regelmäßig überwachen, ob die Systemausgabe verzerrt ist, indem sie auf ihre Geschäftsanforderungen zugeschnittene Eingabedaten verwenden.

Als Behörden sollten die EU-Institutionen Sicherheitsvorkehrungen treffen, um zu vermeiden, dass sie sich zu sehr auf die von den Systemen gelieferten Ergebnisse verlassen, was zu Automatisierungs- und Bestätigungsfehlern führen kann.

Die Anwendung von Verfahren und bewährten Praktiken zur Minimierung und

Abschwächung von Verzerrungen sollte in allen Phasen des Lebenszyklus von generativen KI-Systemen Priorität haben, um eine faire Verarbeitung zu gewährleisten und diskriminierende Praktiken zu vermeiden. Dazu bedarf es der Überwachung und des Verständnisses der Funktionsweise der Algorithmen und der für das Training des Modells verwendeten Daten.

---

[22] Die Prüfung von Trainingsdaten kann dazu beitragen, Verzerrungen und andere problematische Aspekte aufzudecken, indem untersucht wird, wie die Trainingsdaten gesammelt, gekennzeichnet, kuratiert und kommentiert werden. Die Qualität des Audits und seiner Ergebnisse hängt vom Zugang zu den relevanten Informationen ab, einschließlich der Trainingsdatensätze, der Dokumentation und der Implementierungsdetails.

→ EU-X prüft das Vorhandensein von Stichprobenverzerrungen bei automatischen Spracherkennungssystemen. Übersetzungsdienste haben deutlich höhere Wortfehlerraten für einige Sprecher als für andere. Es scheint, dass das System Schwierigkeiten hat, mit einigen englischen Akzenten zurechtzukommen. Nach Rücksprache mit dem Entwickler wurde festgestellt, dass die Trainingsdaten für bestimmte Akzente unzureichend sind, vor allem wenn es sich nicht um Muttersprachler handelt. Da dies systematisch geschieht, erwägt EU-X eine Verfeinerung des Modells mit selbst erstellten Datensätzen.

12. Wie steht es mit der Ausübung der Rechte des Einzelnen?

Die besonderen Merkmale der generativen KI-Systeme bedeuten, dass die Ausübung der Rechte des Einzelnen[23] besondere Herausforderungen mit sich bringen kann, nicht nur im Bereich des Auskunftsrechts, sondern auch in Bezug auf das Recht auf Berichtigung, Löschung und Widerspruch gegen die Datenverarbeitung. Eines der wichtigsten Elemente ist zum Beispiel die Schwierigkeit, die im System gespeicherten personenbezogenen Daten zu identifizieren und Zugang zu ihnen zu erhalten. In großen Sprachmodellen werden z. B. einzelne Wörter wie "Katze" oder "Hund" nicht als Textstrings gespeichert. Stattdessen werden sie als numerische Vektoren durch einen Prozess namens "word embedding" dargestellt. Diese Vektoren stammen aus dem Training des Modells mit großen Mengen von Textdaten. Dies hat zur Folge, dass der Zugriff auf die in diesen Modellen gespeicherten Daten sowie deren Aktualisierung oder Löschung, sofern möglich, sehr schwierig ist. In diesem Sinne kann eine ordnungsgemäße Verwaltung der Datensätze den Zugang zu Informationen erleichtern, was im Falle eines unüberwachten Trainings auf der Grundlage öffentlich zugänglicher Quellen, die personenbezogene Daten enthalten, schwierig ist. Ebenso komplex ist die Verwaltung der Produktion personenbezogener Daten, die durch Rückschlüsse gewonnen wurden. Schließlich kann die Ausübung bestimmter Rechte, wie das Recht auf Löschung, Auswirkungen auf die Wirksamkeit des Modells haben.

Das Führen von nachvollziehbaren Aufzeichnungen über die Verarbeitung personenbezogener Daten sowie die Verwaltung von Datensätzen in einer Weise, die die Rückverfolgbarkeit ihrer Verwendung ermöglicht, kann die Ausübung der Rechte des Einzelnen unterstützen. Techniken zur Datenminimierung können auch dazu beitragen, die Risiken zu mindern, die damit verbunden sind, dass die ordnungsgemäße Ausübung der Rechte des Einzelnen im Einklang mit der Verordnung nicht gewährleistet werden kann.

Als für die Verarbeitung Verantwortliche sind die EU-Institutionen für die Durchführung geeigneter technischer, organisatorischer und verfahrenstechnischer Maßnahmen verantwortlich und rechenschaftspflichtig, um die wirksame Ausübung der Rechte des Einzelnen zu gewährleisten. Diese Maßnahmen sollten bereits in den frühen Phasen des Lebenszyklus des Systems konzipiert und umgesetzt werden und eine detaillierte Aufzeichnung und Rückverfolgbarkeit der Verarbeitungstätigkeiten ermöglichen.

→ EU-X hat in den Datenschutzhinweis für den Chatbot einen Hinweis auf die Ausübung der Rechte des Einzelnen aufgenommen, einschließlich der Rechte auf Auskunft, Berichtigung, Löschung, Widerspruch und Einschränkung der Verarbeitung im Einklang mit der EU-DSGVO. Die Mitteilung enthält die Kontaktdaten des für die Verarbeitung Verantwortlichen und des behördlichen Datenschutzbeauftragten von EU-X sowie einen Hinweis auf die Möglichkeit, eine Beschwerde beim EDSB einzureichen. Auf den Antrag einer Person auf Auskunft über den Inhalt ihrer Unterhaltungen mit dem Chatbot antwortete EU-X nach Durchführung der entsprechenden Prüfungen, dass kein Inhalt der besagten Unterhaltungen über die festgelegte Aufbewahrungsfrist von 30 Tagen hinaus aufbewahrt wird. Die Unterhaltungen wurden, wie der Person mitgeteilt wurde, nicht zum Trainieren des Chatbot-Modells verwendet.

[23] Kapitel III der Verordnung.

13. Wie steht es um die Datensicherheit?

Der Einsatz von generativen KI-Systemen kann bestehende Sicherheitsrisiken verstärken oder neue schaffen und im Falle weit verbreiteter Modelle auch neue Quellen und Übertragungswege für systemische Risiken schaffen. Im Vergleich zu herkömmlichen Systemen können sich generative KI-spezifische Sicherheitsrisiken aus unzuverlässigen Trainingsdaten, der Komplexität der Systeme, Intransparenz, Problemen bei der Durchführung ordnungsgemäßer Tests, Schwachstellen in den Sicherheitsvorkehrungen der Systeme usw. ergeben. Das begrenzte Angebot an Modellen in kritischen Sektoren für die Erbringung öffentlicher Dienstleistungen wie dem Gesundheitswesen kann die Auswirkungen von Schwachstellen in diesen Systemen verstärken. Die Verordnung verpflichtet die EU-Institute, geeignete technische und organisatorische Maßnahmen zu ergreifen, um ein Sicherheitsniveau zu gewährleisten [24] das dem Risiko für die Rechte und Freiheiten natürlicher Personen angemessen ist.

Die für die Verarbeitung Verantwortlichen sollten zusätzlich zu den herkömmlichen Sicherheitskontrollen für IT-Systeme spezifische Kontrollen integrieren, die auf die bereits bekannten Schwachstellen dieser Systeme zugeschnitten sind - Modellumkehrungsangriffe[25]Soforteinschleusung[26], Jailbreaks[27] - auf eine Weise integrieren, die eine kontinuierliche Überwachung und Bewertung ihrer Wirksamkeit ermöglicht. Den für die Verarbeitung Verantwortlichen wird empfohlen, nur Datensätze zu verwenden, die von vertrauenswürdigen Quellen stammen, und regelmäßig Überprüfungs- und Validierungsverfahren durchzuführen, auch für interne Datensätze. Die EU-Institutionen sollten ihre Mitarbeiter darin schulen, wie sie Sicherheitsrisiken im Zusammenhang mit dem Einsatz generativer KI-Systeme erkennen und bewältigen können. Da sich die Risiken schnell weiterentwickeln, ist eine regelmäßige Überwachung und Aktualisierung der Risikobewertung erforderlich. Da sich auch die Modalitäten von Angriffen ändern können, muss ein angemessener Zugang zu fortgeschrittenem Wissen und Fachkenntnissen gewährleistet sein. Ein möglicher Weg, mit unbekannten Risiken umzugehen, ist der Einsatz von "Red Teaming[28]"Techniken, um Schwachstellen zu finden und aufzudecken.

Beim Einsatz von Retrieval Augmented Generation[29] mit generativen KI-Systemen muss geprüft werden, dass das generative KI-System keine persönlichen Daten preisgibt, die in der Wissensbasis des Systems enthalten sein könnten.

Der Mangel an Informationen über die Sicherheitsrisiken, die mit dem Einsatz generativer KI-Systeme verbunden sind, und darüber, wie sie sich entwickeln können, erfordert von den EU-Institutionen äußerste Vorsicht und eine detaillierte Planung aller Aspekte im Zusammenhang mit der IT-Sicherheit, einschließlich kontinuierlicher Überwachung und spezialisierter technischer Unterstützung. Die EU-Institutionen müssen sich der Risiken bewusst sein, die sich aus Angriffen böswilliger Dritter ergeben, und die verfügbaren Instrumente kennen, um diese zu mindern.

[24] Artikel 33 der Verordnung.
[25] Eine Modellumkehrung findet statt, wenn ein Angreifer durch Reverse-Engineering Informationen daraus extrahiert.
[26] Böswillige Akteure nutzen Prompt-Injection-Angriffe, um bösartige Anweisungen so einzuschleusen, als wären sie harmlos.
[27] Böswillige Akteure nutzen Jailbreaking-Techniken, um die Schutzmechanismen des Modells zu missachten.
[28] Ein rotes Team setzt Angriffstechniken ein, die darauf abzielen, Schwachstellen im System zu finden.
[29] KI-Systeme, bei denen ein Large Language Model seine Antworten auf eine Wissensbasis stützt, die vom Eigentümer des generativen KI-Systems (z. B. einer EUI) mit internen Quellen vorbereitet wurde, und nicht auf das vom LLM selbst gespeicherte Wissen.

EU-X hat nach einer Sicherheitsbewertung beschlossen, das ASR-System vor Ort zu implementieren, anstatt die API-Dienste zu nutzen, die dem Entwickler des Modells zur Verfügung gestellt werden. EU-X wird
seine IT-Mitarbeiter in enger Zusammenarbeit mit dem Anbieter in der Nutzung und Weiterentwicklung des Systems schulen. Dies kann auch Schulungen zur Verfeinerung des Modells umfassen. Darüber hinaus wird EU-X die Dienste eines externen Prüfers in Anspruch nehmen, um die ordnungsgemäße Umsetzung des Systems, auch in Bezug auf die Sicherheit, zu überprüfen.

14. Möchten Sie mehr wissen?

- **Die Arbeit des EDSB zur KI**

  - 45. geschlossene Sitzung der Weltdatenschutzversammlung - Entschließung zu generativen Systemen der künstlichen Intelligenz - 20. Oktober 2023

  - EDSB TechDispatch #2/2023 - Erklärbare künstliche Intelligenz

  - Der EDSB bei der Arbeit: Datenschutz und KI (enthält Links zu verschiedenen Dokumenten, die vom EDSB allein oder in Zusammenarbeit mit anderen Behörden veröffentlicht wurden)

  - Gemeinsame Stellungnahme 5/2021 des EDSB und des EDSB zu dem Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates mit harmonisierten Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz)

  - Stellungnahme 44/2023 des EDSB zu dem Vorschlag für ein Gesetz über künstliche Intelligenz im Lichte der legislativen Entwicklungen

    Große Sprachmodelle (EDSB-Website, Teil des EDSB-Berichts "TechSonar" 2023-2024)

- **Andere relevante Dokumente**

  - Leitlinien zur automatisierten individuellen Entscheidungsfindung und zum Profiling für die Zwecke der Verordnung 2016/679 (wp251rev.01)

  - CNIL: AI-Anleitungsblätter

  - Spanische Datenschutzbehörde: Künstliche Intelligenz: Grundsatz der Genauigkeit bei der Verarbeitungstätigkeit

  - Italienische Datenschutzbehörde: Decalogo per la realizzazione di servizi sanitari nazionali attraverso sistemi di Intelligenza Artificiale - September 2023 (Italienisch)

  - Der Hamburgische Beauftragte für Datenschutz und Informationsfreiheit - Checkliste für den Einsatz von LLM-basierten Chatbots - 15.11.2023

  - AI Security Concerns in a nutshell (DE Bundesamt für Sicherheit in der Informationstechnik, März 2023)

  - Mehrschichtiger Rahmen für gute Cybersicherheitspraktiken für KI (ENISA, Juni 2023)

  - Ethik-Leitlinien für vertrauenswürdige KI (Hochrangige Expertengruppe der Europäischen Kommission für KI, 2019)

  - Lebendige Leitlinien für den verantwortungsvollen Einsatz von generativer KI in der Forschung (Dokument des ERA-Forums für Interessenvertreter, März 2024)

  - OECD AI Incidents Monitor (AIM)

- OECD-Katalog oder Werkzeuge und Metriken für vertrauenswürdige KI