Bonn, Bukarest, Dublin, Lissabon, Madrid, Mailand, Paris, Den Haag, Wien, Warschau

Generative KI: Die Auswirkungen auf den Datenschutz

CEDPO AI-Arbeitsgruppe
16. Oktober 2023

Kontaktinformationen:
https://cedpo.eu
info@cedpo.eu

Über diesen Leitfaden

Künstliche Intelligenz ist für Datenschutzbeauftragte und Datenschützer kein neues Konzept. Generative KI jedoch schon. Als OpenAIs ChatGPT im November 2022 auf den Markt kam, hatte die Mehrheit der Datenschutzbeauftragten noch nie etwas von generativer KI gehört und beschäftigte sich in ihrer täglichen Arbeit sicherlich nicht mit solchen Technologien.

Jetzt, da ChatGPT von über 100 Millionen Nutzern weltweit verwendet wird und viele andere Anbieter wie Google Bard und Anthropic's Claude auf den Markt kommen, ist es für Datenschützer eine betriebliche Realität und eine Notwendigkeit geworden, sich mit den Folgen der generativen KI-Tools zu befassen, die in Unternehmen rasch eingesetzt werden. Unabhängig davon, ob diese Tools einfach übernommen oder von Unternehmen anhand ihrer eigenen Datensätze feinabgestimmt werden, gibt es neuartige und bisher nicht untersuchte Auswirkungen auf den Datenschutz, mit denen sich Datenschutzbeauftragte rasch auseinandersetzen müssen.

Ziel dieses Papiers ist es, Datenschützer durch das Labyrinth der Probleme zu führen, die sich mit der raschen Einführung dieser Technologien in Unternehmen ergeben. Neben anderen Schlüsselthemen befasst sich dieses Papier mit den Risiken der Datenweitergabe, der Genauigkeit personenbezogener Daten, der Durchführung von Datenschutzfolgenabschätzungen für generative KI-Tools, der Umsetzung des Datenschutzes durch Design, der Auswahl einer rechtmäßigen Grundlage für das Training generativer KI-Systeme, der Optimierung von Organisationsstrukturen, der Anwendung von Techniken zur Verbesserung des Datenschutzes und dem Umgang mit den Rechten betroffener Personen im Zusammenhang mit diesen Technologien.

Es wird keine Zukunft ohne generative KI geben, und da Daten beim Training und Betrieb dieser Systeme eine so zentrale Rolle spielen, werden die DSB eine zentrale Rolle dabei spielen, sicherzustellen, dass sowohl Datenschutz- als auch Data-Governance-Standards im Mittelpunkt dieser Technologien stehen.

Inhaltsverzeichnis

Inhalt

1. Richtigkeit der personenbezogenen Daten

Die Richtigkeit personenbezogener Daten, die von generativen Werkzeugen der künstlichen Intelligenz (KI) verarbeitet werden, ist ein grundlegendes Datenschutzproblem bei solchen Technologien. In Artikel 5 Absatz 1 Buchstabe d der Datenschutz-Grundverordnung DS-GVO) heißt es: "Personenbezogene Daten müssen sachlich richtig und, wenn nötig, auf den neuesten Stand gebracht sein; es sind alle angemessenen Maßnahmen zu treffen, damit unrichtige personenbezogene Daten (...) unverzüglich gelöscht oder berichtigt werden".

Es liegt auf der Hand und ist eine Frage des gesunden Menschenverstands, dass die Verarbeitung ungenauer personenbezogener Daten sehr reale Auswirkungen auf die betroffene Person hinter den Daten haben kann. Generative KI-Tools wie der weit verbreitete textbasierte Chatbot ChatGPT von OpenAI weisen jedoch von Natur aus zahlreiche Ungenauigkeiten in den Daten auf, die sie verarbeiten. Es liegt in der Natur der Sache, dass diese Tools riesige Mengen an Trainingsdaten aufnehmen, die sie aus massiven Datensammlungen im Internet beziehen. Diese Daten sind zwangsläufig mit allen Ungenauigkeiten behaftet und werden zu einem Teil der Datenbank, mit der die Benutzer von ChatGPT Abfragen durchführen. Wenn ein Nutzer eine Antwort erhält, die entweder ganz oder teilweise ungenau ist, führt dies zu dem, was KI-Anbieter als "Halluzinationen" oder, in der Fachsprache, "Unwahrheiten" bezeichnen.

Sogar OpenAI selbst warnt auf seiner Website vor den damit verbundenen Gefahren und davor, dass man sich nicht automatisch auf die Richtigkeit der abgerufenen Daten verlassen kann. In einem Abschnitt mit der Überschrift "Einschränkungen" heißt es: "ChatGPT schreibt manchmal plausibel klingende, aber falsche oder unsinnige Antworten"[1] . OpenAI weist außerdem darauf hin, dass das Tool oft zu Ungenauigkeiten beiträgt, indem es im Wesentlichen errät, was ein unsicherer Benutzer meint. Sie erklärt: Idealerweise würde das Modell klärende Fragen stellen, wenn der Benutzer eine mehrdeutige Frage stellt. Stattdessen raten unsere aktuellen Modelle in der Regel, was der Nutzer gemeint hat.[2] '

In Verbindung mit der Tatsache, dass die Datenverarbeitungsbedingungen von ChatGPT deutlich machen, dass der Nutzer, der für die Datenverarbeitung Verantwortliche ist, während OpenAI lediglich der Datenverarbeiter ist, sollte es für die Nutzer klar sein, dass es sich hier um einen Markt

[1] https://openai.com/blog/chatgpt
[2] Ebd.

handelt, bei dem der Käufer aufpassen muss. Und warum? Weil jede Partei, die ungenaue personenbezogene Daten weiterverarbeitet, für die Nichteinhaltung von Artikel 5 (1) (d) haftbar gemacht werden kann. Im Zusammenhang mit ChatGPT macht also die Verwendung von ungenauen personenbezogenen Daten, die vom Tool zur Verfügung gestellt wurden, den Nutzer haftbar für die Nichteinhaltung der DS-GVO, insbesondere wenn eine solche Weiterverwendung die Grundrechte und -freiheiten der betroffenen Personen beeinträchtigt.

Die Unternehmen sollten sich darüber im Klaren sein, dass dies nicht nur ein theoretischer Punkt ist und dass die Aufsichtsbehörden die Unternehmen der generativen KI bereits zur Rechenschaft über die Genauigkeit ihrer Daten aufgefordert haben. Im März 2023 hat die italienische Datenschutzbehörde den Einsatz von ChatGPT in Italien blockiert und dabei unter anderem festgestellt, dass die Daten häufig nicht korrekt waren. Sie stellte fest, dass "die von ChatGPT zur Verfügung gestellten Informationen nach den bisher durchgeführten Tests nicht immer mit den tatsächlichen Gegebenheiten übereinstimmen, so dass ungenaue personenbezogene Daten verarbeitet werden".[3]

Die Datenschutzbeauftragten (DSB) müssen sich also weiterhin der Risiken der Verarbeitung ungenauer Daten bewusst sein. Den Nutzern innerhalb der Organisation eines Datenschutzbeauftragten sollten klare Leitlinien an die Hand gegeben werden, damit sie verstehen, dass die Ergebnisse generativer KI-Tools wie ChatGPT mit einer Gesundheitswarnung einhergehen, d. h., dass der menschliche Nutzer immer noch für die Überprüfung der Richtigkeit der erhaltenen personenbezogenen Daten verantwortlich ist. Dies ist ein entscheidender Punkt.

Ein weiteres damit verbundenes Risiko ergibt sich aus Art. 5 Abs. 1 lit. D DS-GVO, wonach personenbezogene Daten "auf dem neuesten Stand gehalten" werden müssen. ChatGPT und ähnliche Tools wie Bard von Google und Claude von Anthropic stützen sich auf das Scraping von Daten bis zu einem bestimmten Zeitpunkt, was bedeutet, dass ihre Datenbank veraltet und somit zwangsläufig nicht mehr auf aktuelle Ereignisse reagiert. Dies birgt die Gefahr, dass die Nutzer personenbezogene Daten erhalten, die nicht mehr relevant sind, denen vielleicht der Kontext fehlt oder die schlichtweg ungenau sind, wenn man bedenkt, wie sich die Ereignisse in der Zwischenzeit verändert haben oder wie sich die Informationen weiterentwickelt haben.

Die DSB sollten sich auch darüber im Klaren sein, auf welche Art und Weise unzulässige Verzerrungen und Diskriminierungen in den Trainingssätzen indirekt zu ungenauen

[3] https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9870847#english

Datenausgaben führen könnten, was wiederum den Nutzer dem Risiko einer weiteren Verarbeitung ungenauer Daten aussetzt.

Ein letztes, globales Risiko bei generativen KI-Chatbots ist der Ton, den sie anschlagen: ein orakelhaftes Maß an Gewissheit und Autorität, das man fast als dunkles Muster bezeichnen könnte, so irreführend ist es in seiner Wirkung auf die Bewertung von Suchergebnissen. Wenn generative KI-Chatbots offensichtlich falsch oder ungenau sind, dann oft auf eine sehr selbstbewusste und verwirrend endgültige Art und Weise, eine Haltung, die die Tatsache verschleiert, dass die Antwort, wie zum Beispiel OpenAI zugibt, einfach "Unsinn" sein kann. Bei allen Suchergebnissen sollte der Tonfall der Antwort ignoriert werden, und auch hier sollten sich die Nutzer darüber im Klaren sein, dass die Ergebnisse dieser Tools einer menschlichen Bewertung bedürfen, vor allem, wenn es um Fragen zur Richtigkeit der betreffenden personenbezogenen Daten geht.

2. Freigabe persönlicher Daten mit generativen KI-Tools

Künstliche Intelligenz (KI) hat sich rasch von einem Konzept der Science-Fiction zu einem relativ alltäglichen Merkmal unseres Lebens entwickelt. Ein sich rasch entwickelnder Zweig der KI ist die generative KI, die neue, zuvor nicht vorhandene Daten erzeugen kann, die den Eingabedaten sehr ähnlich sind. Generative KI-Modelle können unter den richtigen Bedingungen hochwertige Texte, Bilder, Musik und vieles mehr erzeugen. Der Komfort und das innovative Potenzial der generativen KI haben jedoch ihren Preis. Trotz ihrer vielversprechenden Fähigkeiten birgt die gemeinsame Nutzung personenbezogener Daten mit diesen Systemen erhebliche Risiken für den Datenschutz, die Vertraulichkeit sowie die Integrität und Sicherheit der Daten. Das Verständnis dieser Risiken ist für den Schutz der individuellen Datenschutzrechte und für die Aufrechterhaltung eines sicheren digitalen Umfelds von entscheidender Bedeutung.

Wie die meisten KI-Systeme ist auch die generative KI datengesteuert. Beim herkömmlichen KI-Training werden große Datensätze in KI-Modelle eingespeist, die dann aus diesen Daten Muster und Merkmale lernen können. Sobald das Training abgeschlossen ist, ist das KI-System in der Lage, auf der Grundlage der erlernten Muster und Merkmale Ergebnisse zu erzeugen. Das bedeutet, dass personenbezogene Daten, sobald sie Teil des Trainingssatzes der KI sind, zur Bildung des internen Modells der KI beitragen und unweigerlich ihr Verhalten und ihre Ergebnisse beeinflussen. Die Daten werden somit zu einem "Teil" der KI in dem Sinne, dass sie das Verständnis und das Wissen

des Systems beeinflussen. Dies wirft erhebliche Datenschutzbedenken auf, wenn personenbezogene Daten als Trainingsdaten verwendet werden.

Generative KI-Modelle, die auf personenbezogene Daten trainiert wurden, können potenziell sensible Informationen wie Namen, Adressen, Gesundheitsinformationen oder sogar Finanzdaten extrahieren und diese Daten dann in den Suchergebnissen für andere Nutzer wieder veröffentlichen. Darüber hinaus können generative KI-Modelle die Exposition verstärken, indem sie mehr Daten erzeugen, die der ursprünglichen Eingabe ähneln. Dritte können diese Daten dann für unrechtmäßige Aktivitäten wie invasive Werbung, Phishing-Betrug oder in schwerwiegenderen Fällen für Betrug oder Identitätsdiebstahl nutzen. Dies verdeutlicht die Komplexität der Kontrolle darüber, wie personenbezogene Daten von generativen KI-Modellen verwendet werden. Sobald personenbezogene Daten an generative KI-Modelle weitergegeben wurden, wird die Verwaltung und Nachverfolgung ihrer Nutzung aufgrund der Art und Weise, wie KI-Systeme Informationen verarbeiten sowie Daten speichern und über verschiedene Systeme hinweg replizieren, zu einer komplizierten (wenn nicht gar unmöglichen) Aufgabe. Daher kann es unglaublich schwierig oder unrealistisch sein, personenbezogene Daten, die mit generativen KI-Modellen geteilt wurden, zurückzuziehen. Für DSB bedeutet dies, dass die Nutzer genau verstehen müssen, welche Arten von Informationen mit generativen KI-Tools geteilt werden können und welche nicht, denn wenn personenbezogene Daten einmal geteilt wurden, ist der Rubikon überschritten und es wird sehr schwierig sein, das Geschehene rückgängig zu machen.

Eines der bedenklichsten Risiken im Zusammenhang mit der Weitergabe persönlicher Daten an generative KI ist die Schaffung und Verbreitung von "Deepfakes". Deepfakes bezeichnen die Anwendung von KI zur Erstellung, Veränderung oder Manipulation von Inhalten wie Bildern, Audio- und Videodateien in einer Weise, dass hyperrealistische, aber völlig falsche Inhalte erzeugt werden. Durch Training mit personenbezogenen Daten kann generative KI synthetische Medien erzeugen, die sich überzeugend als natürliche oder juristische Personen ausgeben. Diese Deepfakes können dann böswillig eingesetzt werden, z. B. für Desinformationskampagnen, Betrug oder Belästigung. Damit verbunden ist die Tatsache, dass die Genauigkeit generativer KI-Entscheidungen stark von der Qualität und Vielfalt der eingegebenen Trainingsdaten abhängt. Sind diese persönlichen Daten verfälscht, können auch die Ergebnisse der KI verfälscht werden, was zu ungerechten Konsequenzen führt.

Generative KI birgt ein großes Potenzial für zahlreiche Anwendungen, aber ihre Nutzung personenbezogener Daten muss sorgfältig verwaltet werden, um potenzielle Risiken zu

minimieren. Durch strenge Datenschutzkontrollen, ethische KI-Praktiken und einen soliden Rechtsschutz kann es möglich sein, das Potenzial der generativen KI zu nutzen und gleichzeitig die Rechte des Einzelnen auf Datenschutz zu wahren und ein sicheres digitales Umfeld zu fördern.

3. Was ist eine angemessene Rechtsgrundlage?

Die Rechtsgrundlage, die für das Training von KI-Systemen mit personenbezogenen Daten gelten soll, ist eine wichtige Überlegung. Auf den ersten Blick gibt es keinen offensichtlichen Kandidaten, der diese Verarbeitungstätigkeit eindeutig legitimieren und gleichzeitig die Datenschutzrechte der betroffenen Personen wahren würde. Dies ist ein entscheidender Gesichtspunkt, da die Menge der Trainingsdaten, die für generative KI-Anwendungen verwendet werden, enorm ist und weiter zunimmt. Wenn solche Trainingsaktivitäten fortgesetzt werden sollen und wenn die KI ihr Versprechen einlösen soll, dann kann sie nicht auf einer unsicheren Rechtsgrundlage in Bezug auf personenbezogene Daten beruhen. Darüber hinaus ist das Gesetz über künstliche Intelligenz (KI-Gesetz) in diesem Punkt nicht besonders aufschlussreich, da Artikel 10 (der sich mit der Datenverwaltung und der Verwaltung von Trainingsdaten für KI-Systeme befasst) keine spezifische Rechtsgrundlage für die Verwendung personenbezogener Daten für das Training von KI-Systemen schafft. Wir müssen uns also an die Datenschutz-Grundverordnung wenden, um eine geeignete Rechtsgrundlage für diese Tätigkeit zu finden.

Zunächst werden wir kurz darauf eingehen, wie Daten für das Training generativer KI-Systeme verwendet werden. Dies geschieht auf vier grundlegende Arten:

1. Basierend auf persönlichen Daten, die aus dem Internet gesammelt (Scrapping) wurden;
2. Wenn die personenbezogenen Daten von den Nutzern des KI-Systems zur Verfügung gestellt wurden, z. B. wenn sie Eingabeaufforderungen an generative KI-Tools übermitteln;
3. wenn die personenbezogenen Daten von Dritten erhoben wurden, z. B. von Datenmaklern oder Unternehmen, die über Datenbanken verfügen, die für die KI-Trainingsphase relevant sind (z. B. eine Datenbank mit Gerichtsentscheidungen für ein prädiktives KI-Tool im juristischen Bereich); und
4. Wenn KI-Entwickler/Betreiber die in ihren eigenen Datenbanken gespeicherten personenbezogenen Daten zum Trainieren des KI-Systems verwenden.

In diesen Fällen sind gemäß Art. 6 DS-GVO drei Rechtsgrundlagen maßgeblich: Vertrag, berechtigtes Interesse und Einwilligung.

1. **Vertrag:**

In Art. 6 Abs. 1 lit. b der DS-GVO wird darauf hingewiesen, dass ein Vertrag eine Rechtsgrundlage für die Verarbeitung personenbezogener Daten sein kann, wenn die "Verarbeitung für die Erfüllung eines Vertrags, dessen Vertragspartei die betroffene Person ist, oder für Maßnahmen erforderlich ist, die auf Antrag der betroffenen Person vor Abschluss eines Vertrags getroffen werden".

Die Anwendung des ersten Zweigs der vertraglichen Rechtsgrundlage (*d. h. die* Erfüllung des Vertrags selbst) würde den Nachweis erfordern, dass das Training des KI-Systems (und nicht die Nutzung der KI nach dem Training) für die Erfüllung eines Vertrags mit der betroffenen Person unbedingt erforderlich ist.

Dieses Erfordernis der Notwendigkeit wird von den Datenschutzbehörden sehr eng ausgelegt. Nach Ansicht des Europäischen Datenschutzausschusses (EDSA) sollte es nicht möglich sein, den Hauptgegenstand des spezifischen Vertrags mit der betroffenen Person zu erfüllen, wenn die Verarbeitung der fraglichen personenbezogenen Daten nicht erfolgt. Mit anderen Worten: Die Verarbeitung der personenbezogenen Daten auf diese Weise sollte eine notwendige Voraussetzung für die Erfüllung des Vertrags sein.

In Anbetracht dieser engen Auslegung bleibt für die Vertragsbasis bei der Schulung eines KI-Systems nur sehr wenig Raum. Diese Grundlage könnte theoretisch angewandt werden, wenn die Nutzung des KI-Systems Gegenstand des zwischen dem KI-Betreiber und dem Nutzer geschlossenen Vertrags *ist* und wenn es keine andere Möglichkeit gibt, diesen Vertrag zu erfüllen, als die KI mit den Daten der Nutzer zu trainieren.

Was den zweiten Teil dieser Rechtsgrundlage, d. *h.* die vorvertraglichen Schritte, betrifft, so müsste für ihre Anwendung nachgewiesen werden, dass eine betroffene Person eine Anfrage im Zusammenhang mit dem möglichen Abschluss eines Vertrags gestellt hat und dass es keine andere Möglichkeit gibt, ihren Forderungen nachzukommen, als die KI zu trainieren (und nicht nur einmal trainiert zu verwenden). Dies ist eine noch restriktivere und begrenztere Option als der erste Teil dieser Rechtsgrundlage.

Insgesamt sind die Umstände, unter denen die Rechtsgrundlage eines Vertrags zur Rechtfertigung des Trainings von KI-Systemen mit personenbezogenen Daten herangezogen werden könnte, sehr

begrenzt, und in der Praxis wird diese Grundlage keine brauchbare Option für die Begründung solcher Verarbeitungstätigkeiten sein.

Im Fall der generativen KI ist ein Vertrag als Rechtsgrundlage ohnehin besonders ungeeignet, da in der Regel kein Vertrag zwischen den betroffenen Personen, deren Daten verwendet werden, und den für die Schulung solcher Systeme mit diesen Daten zuständigen Organisationen besteht.

2. Berechtigte Interessen

Die Grundlage der berechtigten Interessen könnte nur dann Anwendung finden, wenn der für die Verarbeitung Verantwortliche eine Bewertung der berechtigten Interessen vornimmt, um sicherzustellen, dass diese Interessen nicht durch die Interessen oder Grundrechte und -freiheiten der betroffenen Person überlagert werden.

Dies kann jedoch eine Herausforderung sein, zumal die Organisation, die hinter dem Training generativer KI-Tools wie OpenAI steht, in den meisten Fällen weder in direktem Kontakt mit den betroffenen Personen steht, noch irgendeine Form von Beziehung zu diesen Personen unterhält. In diesem Zusammenhang sind die jüngsten Maßnahmen der italienischen Datenschutzaufsichtsbehörde (Garante per la protezione dei dati) gegen ChatGPT zu erwähnen. Im März 2023 sperrte die Behörde ChatGPT im italienischen Hoheitsgebiet, bis OpenAI in der Lage war, bestimmte Fragen zufriedenstellend zu beantworten; eine davon war, dass OpenAI die Rechtsgrundlage für die Schulung von ChatGPT mit personenbezogenen Daten angeben musste. In seiner Antwort auf diesen Punkt nannte OpenAI legitime Interessen als rechtmäßige Grundlage. Dies ist eine höchst bedeutsame Verpflichtung und Erklärung von OpenAI, da es die enorme Aufgabe des Trainings generativer KI-Systeme effektiv an eine Rechtsgrundlage bindet, die von Natur aus unsicher ist, da die betroffenen Personen gemäß Art. 21 der DS-GVO ausdrücklich das Recht haben, einer solchen Verarbeitung zu widersprechen.

Um sich tatsächlich auf die Grundlage der berechtigten Interessen berufen zu können, wäre es insbesondere erforderlich:

- eine fallweise Untersuchung des Kontextes der Ausbildung und der Nutzung der KI sowie der Erhebung der personenbezogenen Daten, um zu überprüfen, ob die Datenverarbeitung den berechtigten Erwartungen der betroffenen Personen entspricht;

- ein Nachweis der strikten Notwendigkeit dieser Verarbeitung und der Tatsache, dass die KI nicht effizient arbeiten kann, ohne mit den betreffenden personenbezogenen Daten trainiert worden zu sein;
- eine größere Transparenz der Datenverarbeitung gegenüber den betroffenen Personen. Alle nach der DS-GVO erforderlichen Informationen müssten den betroffenen Personen in geeigneter Weise zur Verfügung gestellt werden;
- ein wirksames Opt-out-System, das den betroffenen Personen innerhalb einer angemessenen Frist zur Kenntnis gebracht wird, bevor ihre Daten an das KI-System übermittelt werden[4] ;
- ganz allgemein ein effizientes System zur Wahrung der Rechte der betroffenen Personen, das angesichts der Besonderheiten der generativen KI schwierig umzusetzen wäre

3. **Einwilligung::**

Die Grundlage der Einwilligung könnte ebenfalls Anwendung finden, allerdings nur unter sehr genau umschriebenen Umständen. Während sie in Extremfällen die einzig mögliche Rechtsgrundlage sein kann (z. B. bei der Verarbeitung besonderer Datenkategorien oder von Daten über Minderjährige), hat sie in der Regel beim Training generativer KI-Systeme, wie es derzeit konzipiert ist, kaum eine Bedeutung. Der gesamte Apparat, der für das Training von KI-Systemen eingesetzt wird, macht es nahezu unmöglich, eine Einwilligung einzuholen. Der Grund dafür ist, dass die meisten Daten, die zum Trainieren solcher Systeme verwendet werden, von Datenmaklern gekauft werden, die diese Daten durch Scraping im Internet erhalten haben, eine Tätigkeit, die zwangsläufig nicht mit der Einholung der Einwilligung der betroffenen Personen verbunden ist. In der Tat ist die Rechtmäßigkeit von Data Scraping als kommerzielle Tätigkeit alles andere als sicher, und die jüngste gemeinsame Mitteilung von zwölf globalen Datenschutzbehörden, einschließlich des britischen Information Commissioner's Office (ICO), unterstreicht diesen Punkt.[5]

[4] Wenn das Opt-out-System ihnen erst nach der Schulung der KI zur Kenntnis gebracht wird, wäre ein Widerspruch gegen die Verarbeitung in den meisten Fällen zum einen irrelevant, da die Datenverarbeitung bereits stattgefunden hat, und zum anderen sehr schwer zu unterbinden, wenn große Mengen personenbezogener Daten zahlreicher betroffener Personen an eine KI weitergeleitet werden (*siehe Teil XXX zur Ausübung der Rechte der betroffenen Personen*).

[5] https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2023/08/joint-statement-on-data-scraping-and-data-protection/

Die Verwendung der Einwilligung als Rechtsgrundlage würde, voraussetzen, dass alle Anforderungen für eine gültige Einwilligung gemäß der DS-GVO erfüllt sind, d. h. sie müsste aus einer klaren bestätigenden Handlung resultieren, frei gegeben, spezifisch, in Kenntnis der Sachlage und unzweideutig sein. Dies ist in der Tat eine sehr hohe Messlatte, die in der Welt der Schulung von KI-Systemen zu erreichen ist.

Steht der KI-Anbieter nicht in Kontakt mit den betroffenen Personen, was in der Regel der Fall ist, müsste diese Einwilligung vom Nutzer des KI-Systems eingeholt werden, also von der Organisation, zu der die betroffene Person *eine* Beziehung hat. Dies wird jedoch in der Regel erst im Nachhinein geschehen, wenn das KI-System bereits trainiert wurde, so dass ein Widerspruch gegen die Verarbeitung in den meisten Fällen irrelevant wäre, da die Datenverarbeitung bereits stattgefunden hat. Darüber hinaus wäre es auch sehr schwierig, die Verarbeitung rückgängig zu machen, wenn das KI-System bereits große Mengen personenbezogener Daten über zahlreiche betroffene Personen erfasst hat.

Zusammenfassend lässt sich sagen, dass das berechtigte Interesse höchstwahrscheinlich die geeignetste Grundlage für das Training von KI-Systemen mit personenbezogenen Daten ist. Wie bereits erwähnt, bietet es jedoch keine sichere Grundlage, da eine Bewertung der berechtigten Interessen durchgeführt werden muss und die betroffenen Personen jederzeit gegen eine solche Verarbeitung Widerspruch einlegen können.

4. Risiken von Jailbreaking und Datenschutzgarantien

Kurz nach der Veröffentlichung von ChatGPT versuchten Hacker, den KI-Chatbot zu "knacken", indem sie versuchten, seine Sicherheitsvorkehrungen zu umgehen und ihn dazu zu bringen, unangemessene oder irrationale Dinge zu sagen. Diese kompliziert formulierten Aufforderungen, die darauf abzielen, die für KI-Programme geltenden Beschränkungen zu umgehen, sind als "Jailbreaks" bekannt geworden. Dieser Begriff wurde ursprünglich im Zusammenhang mit digitaler Technologie verwendet, um sich Zugang zum Betriebssystem eines Smartphones oder Tablets zu verschaffen, insbesondere eines von Apple hergestellten, um modifizierte oder nicht autorisierte Software auszuführen.

Im Zusammenhang mit generativen KI-Modellen bezieht sich der Begriff nun auf die Gestaltung von Aufforderungen, die die Chatbots dazu bringen, die Regeln für die Produktion von hasserfüllten Inhalten oder das Schreiben über illegale Handlungen zu umgehen. Bei diesen

Angriffen werden die generativen KI-Systeme so manipuliert, dass sie Inhalte produzieren, die gegen die vorgesehenen Regeln verstoßen, z. B. hasserfülltes oder illegales Material. Eine weitere Anwendung dieser Angriffe könnte die Verleumdung und der persönliche Angriff auf eine Person sein, sobald persönliche Daten durchsickern.

Ein auf KI spezialisiertes Sicherheitsunternehmen war in der Lage, GPT-4, den neuesten textgenerierenden Chatbot von OpenAI, nur wenige Stunden nach der ersten Veröffentlichung des Systems zu knacken. Mithilfe sorgfältig ausgearbeiteter Eingabeaufforderungen umging der Geschäftsführer des Sicherheitsunternehmens die Sicherheitssysteme von OpenAI und brachte GPT-4 in kürzester Zeit dazu, homophobe Aussagen zu generieren, Phishing-E-Mails zu erstellen und Gewalt zu befürworten. Dieses abweichende Verhalten stellt ein ernsthaftes Risiko dar, da es das Potenzial hat, persönliche Daten preiszugeben, die versehentlich oder vielleicht sogar absichtlich in das System eingegeben wurden, und somit das Potenzial hat, von bösen Akteuren manipuliert zu werden.

Ein eng damit verbundener Angriff ist der Prompt-Injection-Angriff, mit dem unbemerkt bösartige Daten oder Anweisungen in KI-Modelle eingefügt werden können. Ein Prompt-Injection-Angriff zielt darauf ab, LLM-basierten Tools eine unbeabsichtigte Reaktion zu entlocken. Dadurch kann ein unbefugter Zugriff erreicht, Antworten manipuliert oder Sicherheitsmaßnahmen umgangen werden. Die spezifischen Techniken und Folgen von Prompt-Injection-Angriffen variieren je nach System.

Jailbreaks und Prompt-Injection-Angriffe sind eine Form von unkonventionellem Hacking, bei den gut formulierten Sätzen anstelle von Code verwendet werden, um Schwachstellen in KI-Systemen auszunutzen. Während sich diese Angriffe derzeit auf die Umgehung von Inhaltsfiltern konzentrieren, warnen Sicherheitsforscher vor dem Potenzial für Datendiebstahl und weit verbreitete cyberkriminelle Aktivitäten, da generative KI-Systeme immer weiterverbreitet sind.

Zahlreiche beliebte Online-Dienste und -Produkte stützen sich in hohem Maße auf große Datensätze, um ihre KI-Algorithmen zu trainieren und zu verbessern. Datenströme aus Netzwerken, Social-Media-Plattformen, mobilen Geräten und verschiedenen anderen Quellen tragen zu der riesigen Menge an Informationen bei, die Unternehmen zum Trainieren ihrer maschinellen Lernsysteme verwenden. Daher ist es wichtig zu wissen, dass einige der in diesen Datensätzen enthaltenen Daten wahrscheinlich als personenbezogene Daten betrachtet, werden könnten, selbst von Nutzern, die sich weniger um den Datenschutz sorgen. Leider ist der

Datenschutz aufgrund des Missbrauchs und des falschen Umgangs mit personenbezogenen Daten durch bestimmte Unternehmen zu einem dringenden globalen politischen Thema geworden.

In ähnlicher Weise werden auch viele unserer sensiblen Daten gesammelt, um KI-gestützte Prozesse zu verbessern. Diese Daten spielen eine entscheidende Rolle bei der Einführung des maschinellen Lernens, da ausgefeilte Algorithmen für die Entscheidungsfindung in Echtzeit auf solche Daten angewiesen sind. Suchalgorithmen, Sprachassistenten, Empfehlungsmaschinen und andere KI-Lösungen nutzen umfangreiche Datensätze mit realen Nutzerdaten, um personalisierte und relevante Ergebnisse zu liefern.

Anfang 2023 wurde eine Website namens Jailbreak Chat eingerichtet, auf der Aufforderungen für KI-Chatbots wie ChatGPT aus Online-Foren gesammelt und geteilt werden. Besucher der Website können ihre eigenen Jailbreaks beisteuern, von anderen eingereichten Prompts ausprobieren und über deren Wirksamkeit abstimmen. Böswillige Nutzer könnten diese Jailbreaks ausnutzen, um an die in den Systemen enthaltenen persönlichen Daten zu gelangen, um Straftaten wie Identitätsdiebstahl zu begehen und Deepfakes zu erstellen, um sich als lebende Personen auszugeben.

Die Auswirkungen von Jailbreaks und Prompt-Injection-Angriffen werden noch bedeutsamer, wenn diese Systeme Zugriff auf persönliche und sensible Daten erhalten. Wenn beispielsweise ein erfolgreicher Prompt-Injection-Angriff eine persönliche Assistenten-KI anweist, frühere Anweisungen zu ignorieren und eine E-Mail an alle Kontakte zu senden, könnte dies nicht nur zu einer peinlichen Situation für die betroffene Person führen, sondern auch zu weitreichenden Problemen für die betroffenen Personen und zur schnellen Verbreitung schädlicher Inhalte in den persönlichen und beruflichen Netzwerken der Person.

Die Gewährleistung der Sicherheit von Stiftungsmodellen wie ChatGPT ist von größter Bedeutung, da ihre Nutzung immer weiterverbreitet wird. Die Hacker werden jedoch nicht so schnell aufgeben. Mit der Weiterentwicklung von KI-Systemen sind auch die Jailbreaks komplexer geworden. Einige beinhalten mehrere Charaktere, komplizierte Hintergrundgeschichten, Übersetzungen und sogar Elemente der Codierung, um bestimmte Ergebnisse zu erzeugen.

Einige autorisierte "rote Teams" veranlassen Angriffe auf KI-Modelle, um Schwachstellen aufzudecken. Ein rotes Team in der Cybersicherheit ist das offensive Sicherheitsteam, das für die Aufdeckung von Sicherheitslücken durch Penetrationstests zuständig ist. Mit GAI suchen diese Teams nach Exploits, die tatsächlichen Schwachstellen beinhalten, das Verhalten des Systems

beeinflussen oder Benutzer täuschen, um die Sicherheit des Systems zu umgehen. Andere Versuche stammen von Hobbyisten, die ihre lustigen oder beunruhigenden Ergebnisse in den sozialen Medien präsentieren wollen. Dieser Sicherheitsansatz ist suboptimal, da er fragmentiert ist und sich auf die virale Verbreitung und einflussreiche Personen verlässt, um Abhilfe zu schaffen.

Während Unternehmen wie OpenAI, Google und Microsoft Maßnahmen gegen Jailbreaking und Prompt-Injection-Angriffe ergriffen haben, finden die Forscher, die hinter diesen Angriffen stehen, immer wieder neue Wege, um Schwachstellen auszunutzen. Die Entwicklung generativer KI-Systeme erfordert Ansätze, die über die traditionellen Red-Teaming-Methoden hinausgehen, z. B. die Verwendung eines zweiten KI-Modells zur Analyse von Prompts oder die klare Trennung von System- und Benutzer-Prompts.

Automatisierung und fortschrittliche Techniken sind notwendig, um Jailbreaks und Injektionsangriffe in großem Maßstab zu erkennen und zu entschärfen. Durch die Automatisierung des Prozesses der Identifizierung von Schwachstellen und unbeabsichtigtem Verhalten wollen die Forscher eine größere Anzahl dieser Sicherheitsrisiken entdecken und beseitigen.

Diese Arten von automatisierten Techniken können als Ausgangspunkt für ein tiefergehendes Engagement der KI-Entwickler bei der Beurteilung und Bewertung der Sicherheit ihrer Systeme gesehen werden. Durch die Einbeziehung eines breiten Spektrums von Akteuren und die Betonung von Transparenz und Rechenschaftspflicht sollen die Sicherheit, Zuverlässigkeit und ethische Nutzung generativer KI-Technologie verbessert werden. Bewertungen durch Dritte, die automatische Entschärfung von Jailbreaks und der Einsatz von Red-Teaming werden eine entscheidende Rolle bei der Erreichung dieses Ziels und der Verbesserung der Praktiken im Zusammenhang mit der KI-Entwicklung spielen, um die Anforderungen der DS-GVO und des kommenden KI-Gesetzes zu erfüllen.

5. Wie werden die Rechte der Betroffenen mit generativen KI-Tools umgesetzt?

Generative KI oder GenAI sind KI-Systeme, die in der Lage sind, Texte, Bilder oder andere Medien als Reaktion auf Eingabeaufforderungen zu generieren. Generative Modelle lernen die Muster und die Struktur der Eingabedaten und generieren anschließend neue Inhalte, die den Trainingsdaten ähnlich sind, aber einen gewissen Grad an Neuartigkeit aufweisen, im Gegensatz zur bloßen

Klassifizierung oder Vorhersage von Daten. Diese KI-Systeme basieren häufig auf Generative Pretrained Transformers (GPT), künstlichen neuronalen Netzen, die auf der Transformator-Architektur aufbauen, mit großen Mengen an unmarkierten Textdaten trainiert wurden und in der Lage sind, menschenähnlichen Text zu erzeugen. Sie verwenden große Sprachmodelle (LLMs), um Daten auf der Grundlage des Trainingsdatensatzes zu erzeugen, der zu ihrer Erstellung verwendet wurde.

Das Verständnis der Technologie, die hinter der generativen KI steht, ist von entscheidender Bedeutung, um zu erkennen, dass diese Werkzeuge verschiedene Phasen umfassen und personenbezogene Daten in jeder Phase verarbeitet werden können. Die Verarbeitung personenbezogener Daten in einer Phase impliziert jedoch nicht zwangsläufig die Verarbeitung von Daten in einer anderen Phase.

Zu den datenschutzrechtlichen Phasen, in denen die Rechte der betroffenen Personen in Bezug auf personenbezogene Daten im Zusammenhang mit generativer KI Anwendung finden könnten, gehören:

1. Die Phase der Trainingsdaten, in der die persönlichen Daten einbezogen werden.
2. Die Bereitstellungsphase, in der personenbezogene Daten für die Erstellung von Inhalten und das Inhaltsergebnis selbst verwendet werden.
3. Das Modell selbst, das personenbezogene Daten enthalten kann.

Es ist auch wichtig darauf hinzuweisen, dass generative KI-Software indirekt Daten verarbeiten kann, die sich insbesondere auf den Nutzer der Lösung beziehen, wie z. B. Kontodaten oder Metadaten im Zusammenhang mit der Nutzung der Lösung.

Bei gängigen Modellen des maschinellen Lernens ist die Identifizierung der Personen, auf die sich die Trainingsdaten beziehen, eine potenzielle Herausforderung für die Wahrung ihrer Rechte. In der Regel enthalten diese Daten nur die für die Vorhersagen relevanten Informationen ohne eindeutige Identifikatoren der betroffenen Personen. Sie werden verschiedenen Vorverarbeitungsmaßnahmen unterzogen, um sie für maschinelle Lernalgorithmen geeignet zu machen, wobei personenbezogene Daten oft in eine Form umgewandelt werden, die es schwieriger (aber nicht unmöglich) macht, sie mit bestimmten Personen in Verbindung zu bringen. Die Datenschutzgesetze können daher auch auf diese umgewandelten Daten Anwendung finden, da sie immer noch zur Identifizierung von Personen verwendet werden könnten. Dieser Prozess muss bei der Beantwortung von Anfragen zu den Rechten des Einzelnen berücksichtigt werden.

Dieser Prozess ist bei generativen KI-Modellen anders als bei herkömmlichen Modellen des maschinellen Lernens, wie im vorigen Abschnitt erläutert. Generative KI-Modelle werden oft mit Daten trainiert, die im Internet zugänglich sind, und ihr Wert liegt auch oft in der Generierung von Ergebnissen, die sich auf natürliche Personen beziehen, was eine erhebliche Menge an personenbezogenen Daten in den Trainingsdaten für diese Modelle impliziert. Infolgedessen könnten diese Datensätze das Ziel von Anfragen der betroffenen Personen sein.

Bei generativen KI-Modellen stellt das "kontinuierliche Lernen" auch eine besondere Herausforderung für die Einhaltung der Datenschutzgrundverordnung dar. Diese Modelle werden regelmäßig auf der Grundlage von Nutzerinteraktionen aktualisiert, was bedeutet, dass kontinuierlich personenbezogene Daten verarbeitet werden. Diese Daten stammen meist aus den Interaktionen und Aufforderungen der Nutzer des Tools, und es ist zu beachten, dass die betroffenen Personen und die Datenlieferanten im Zusammenhang mit kontinuierlich lernenden KI-Modellen nicht notwendigerweise dieselbe Einheit sind.

In Anbetracht dieser Überlegungen stellt der Umgang mit den Datenrechten im Rahmen der DS-GVO im Zusammenhang mit generativen KI-Modellen eine besondere Herausforderung dar, insbesondere für die Rechte auf Löschung, Berichtigung, Auskunft und Widerspruch.

Das erste gemeinsame Problem ist die Nicht-Wiederauffindbarkeit von Daten in generativen KI-Modellen. Wie bereits erwähnt, beziehen diese Modelle Daten aus einem breiten Spektrum von Quellen, wie z. B. Web-Scraping und Benutzerinteraktionen. Dieser vielschichtige Ansatz der Datenerfassung erschwert die Rückverfolgung einzelner Beiträge. Im Gegensatz zu herkömmlichen Datenspeichersystemen sind personenbezogene Daten in Generativ AI-Systemen außerdem tief in komplexe Algorithmen eingebettet, was die Isolierung bestimmter Daten erschwert. Dies macht es schwierig, die DS-GVO-Rechte zu erfüllen, da die Identifizierung, ob und wo personenbezogene Daten innerhalb des Systems verarbeitet werden.

Eine weitere Ebene der Komplexität ist die Frage der "abgeleiteten personenbezogenen Daten". Dabei handelt es sich um Schlussfolgerungen, die das Modell auf der Grundlage seines Trainings ziehen kann. So könnte ein generatives KI-Modell beispielsweise die politische Zugehörigkeit eines Nutzers auf der Grundlage früherer Dateninteraktionen ableiten. Die vorherrschende Meinung tendiert dazu, diese Rückschlüsse bei der Beantwortung von Anfragen zu Rechten zu berücksichtigen, da sie indirekt persönliche Informationen preisgeben könnten. Das Konzept der "abgeleiteten Gruppendaten" verdient ebenfalls Aufmerksamkeit. Diese Art von Daten wird auf

der Grundlage breiterer Muster erzeugt, die während des Trainings erkannt werden. Ob diese Gruppendaten als personenbezogen betrachtet werden, hängt von ihrer späteren Verarbeitung und Nutzung ab.

Neben den allgemeinen Herausforderungen gibt es auch spezifische Herausforderungen im Zusammenhang mit den Betroffenenrechten, die eine Änderung oder Löschung von Daten erfordern. Insbesondere das Ändern oder Entfernen von Daten aus dem Trainingssatz nach einer Anfrage der betroffenen Person könnte die Validierung und Korrektheit des Modells beeinträchtigen. Die ursprünglichen Daten dienen oft als Grundlage für solche Validierungsprozesse. Darüber hinaus würde das Löschen oder Ändern von Daten, die bereits in das Modell eingebettet sind, oft bedeuten, dass diese Daten entfernt oder geändert werden müssen, um das Modell neu zu trainieren, was sowohl kostspielig als auch zeitaufwändig ist.

Zusammenfassend lässt sich sagen, dass die Überschneidung von DS-GVO-Rechten und generativen KI-Modellen ein Labyrinth von Herausforderungen darstellt, die jeweils ihre eigenen Feinheiten und Komplikationen haben. Die Natur dieser Modelle, von der Art und Weise, wie sie Daten einbetten und verarbeiten, bis hin zu den Schwierigkeiten bei der Nachverfolgung einzelner Beiträge, macht die Einhaltung der DS-GVO noch komplexer. Es gibt zwar kein Patentrezept, um diese Herausforderungen nahtlos zu meistern, aber die sich entwickelnde Landschaft bietet einige neue Lösungen, die als Ausgangspunkt für die Einhaltung der Vorschriften dienen könnten.

Zunächst einmal können trotz des Fehlens einer Einheitslösung proaktive Schritte unternommen werden. Die Umsetzung des Grundsatzes des "eingebauten und standardmäßigen Datenschutzes" während der Entwicklungs- und Implementierungsphase des generativ AI-Modells bietet eine grundlegende Ebene des Datenschutzes, die von Anfang an integriert ist.

Bei der Navigation durch das komplexe Terrain des Datenschutzes könnte man eine präventive Strategie in Betracht ziehen, die den Umfang der Daten und ihre Erkennungsmerkmale eingrenzt. Auf diese Weise könnten viele der Schwierigkeiten, die sich später im Datenverarbeitungszyklus ergeben könnten, möglicherweise gemildert werden. Die Datenminimierung könnte ein wesentlicher Bestandteil dieser frühzeitigen Planung sein und den für die Datenverarbeitung Verantwortlichen anleiten, nur das wirklich Notwendige zu erfassen. Darauf aufbauend könnten Techniken zur Anonymisierung personenbezogener Daten oder der Einsatz von Technologien zum Schutz der Privatsphäre (PETs), wie z. B. synthetische Daten, eine weitere Verringerung des Umfangs ermöglichen, der potenziell von den Rechten der betroffenen Person betroffen ist.

Darüber hinaus ist die Investition in proaktive Maßnahmen wie Datenzuordnung und Datenkennzeichnung von entscheidender Bedeutung. Solche Maßnahmen schaffen Klarheit über die Herkunft und die Merkmale von Schulungsdaten und erleichtern die Bearbeitung von Rechteanfragen in späteren Phasen.

Wenn generative KI-Modelle von der Entwicklung zur Bereitstellung übergehen, verlagert sich der Schwerpunkt auf die Optimierung der Anpassungsfähigkeit und Rückverfolgbarkeit. In dieser Phase ist das Führen sorgfältiger Datenverarbeitungsprotokolle nicht nur eine gute Praxis, sondern unerlässlich, um die Beantwortung richtiger Anfragen zu erleichtern. Dies ist umso wichtiger, als die Daten in dieser Phase zunehmend anpassungsfähig sind. Darüber hinaus können die Herausforderungen des kontinuierlichen Lernens bei eingesetzten Modellen durch Methoden der Versionierung wirksam angegangen werden. Dies ermöglicht ein effizientes Rollback zu einem früheren Modellzustand, ohne dass ein mühsames Neulernen von Grund auf erforderlich ist. Diese Verknüpfung stellt sicher, dass sowohl die Anpassungsfähigkeit als auch die Rückverfolgbarkeit berücksichtigt werden, und bietet einen robusten Rahmen für die Einhaltung der Vorschriften.

6. Datenschutz durch Technikgestaltung: Wie man generative KI-Tools in Übereinstimmung mit der DS-GVO entwickelt

Der Datenschutz durch Technik spielt eine zentrale Rolle bei der Einhaltung der Datenschutz-Grundverordnung (DSGVO). Er beinhaltet den Schutz personenbezogener Daten bereits in den frühen Phasen des Entwurfs und während des gesamten Lebenszyklus des Systems. Die Idee des Datenschutzes durch Design stammt aus einer allgemeineren Reihe von Datenschutzgrundsätzen mit dem Titel Privacy by Design, die in den frühen 2000er Jahren in Kanada entwickelt wurden. Privacy by Design ist ein Ansatz für die Systemtechnik, der ursprünglich von Ann Cavoukian entwickelt und 1995 von einem gemeinsamen Team des Information and Privacy Commissioner of Ontario (Kanada), der niederländischen Datenschutzbehörde und der niederländischen Organisation für angewandte wissenschaftliche Forschung in einem Bericht über Technologien zur Verbesserung des Datenschutzes formalisiert wurde. Der Rahmen für "Privacy by Design" wurde 2009 veröffentlicht und 2010 von der Internationalen Versammlung der Datenschutzbeauftragten und Datenschutzbehörden angenommen. Im selben Jahr verabschiedete die Internationale Konferenz der Datenschutzbehörden und Datenschutzbeauftragten einstimmig eine

Entschließung, in der Privacy by Design als wesentlicher Bestandteil des grundlegenden Schutzes der Privatsphäre anerkannt wurde. Daraufhin nahm die US Federal Trade Commission Privacy by Design als eine von drei empfohlenen Praktiken zum Schutz der Privatsphäre im Internet auf.

Kurz nach 2010 begann Europa mit der Überarbeitung seiner Datenschutzgesetze. Inspiriert von "Privacy by Design" und seinen Grundsätzen hat Europa Grundsätze für den Datenschutz durch Design aufgestellt, die 2018 über Artikel 25 der Allgemeinen Datenschutzverordnung (DS-GVO) in das Gesetz aufgenommen wurden.

In den letzten Jahren hat die rasche Entwicklung der generativen KI zu einem verstärkten Bewusstsein für potenzielle Risiken und ethische Überlegungen bei der Entwicklung von Systemen geführt, die personenbezogene Daten verarbeiten. Diese Bedenken umfassen nicht nur komplexe Datenschutzrisiken wie das Durchsickern sensibler Informationen und Chatverläufe, sondern auch eine Reihe von Bedrohungen für die Rechte der betroffenen EU-Bürger, darunter das "Recht auf Vergessenwerden". Dieses Recht ermöglicht es Einzelpersonen, die Löschung ihrer personenbezogenen Daten durch ein Unternehmen zu verlangen. Während das Löschen von Daten aus Datenbanken relativ einfach ist, ist das Entfernen von Daten aus maschinellen Lernmodellen eine komplexere Aufgabe. Anonymisierungstechniken und Praktiken zur Datenminimierung können dazu beitragen, ein Gleichgewicht zwischen der Wahrung der Rechte des Einzelnen und der Erhaltung des Gesamtnutzens des generativen KI-Modells herzustellen.

Aus menschlicher Sicht ist zu bedenken, dass aufgrund der Komplexität moderner KI-Systeme die an der Entwicklung und dem Einsatz von KI-Systemen beteiligten Personen häufig über ein breiteres Spektrum an Fähigkeiten und Hintergründen verfügen als die üblichen Systementwickler, darunter herkömmliche Softwareentwickler, Systemadministratoren, Datenwissenschaftler, Statistiker und Fachleute.

Aufgrund dieses breiten Spektrums an Fachwissen besteht möglicherweise ein geringeres Verständnis für die allgemeinen Anforderungen an die Einhaltung von Sicherheitsvorschriften sowie für die Anforderungen des Datenschutzrechts im Besonderen. Für diese Personen war die Sicherheit personenbezogener Daten möglicherweise nicht immer eine Hauptpriorität, insbesondere wenn sie zuvor KI-Anwendungen mit nicht personenbezogenen Daten oder in einer Forschungsfunktion entwickelt haben, bei der personenbezogene Daten in Sandboxen geschützt waren.

Ein weiteres großes Datenschutzproblem sind verzerrte Algorithmen. Generative KI-Systeme lernen aus riesigen Datenmengen, und wenn diese Daten voreingenommen sind, können die Algorithmen diese Voreingenommenheit in ihren Ergebnissen aufrechterhalten und verstärken. Dies wirft ethische Fragen über Fairness, Diskriminierung und den potenziellen Schaden auf, der durch voreingenommene KI-Inhalte verursacht wird, wenn sie dazu verwendet werden, wichtige, lebensverändernde Entscheidungen über betroffene Personen zu treffen.

KI-Halluzinationen beziehen sich auf Fälle, in denen generative KI-Systeme Ergebnisse produzieren, die nicht auf realen oder genauen Informationen beruhen. Diese Halluzinationen können die Nutzer in die Irre führen und haben potenzielle Auswirkungen auf die Sicherheit der betroffenen Personen. Generative KI-Systeme müssen zuverlässige und vertrauenswürdige Ergebnisse liefern, insbesondere in Bezug auf europäische Bürgerinnen und Bürger, deren personenbezogene Daten und deren Richtigkeit durch die DS-GVO geschützt sind.

Der Anstieg von Deepfakes, d. h. realistischen, aber manipulierten Audio- oder Videoinhalten, wird ebenfalls mit generativer KI-Technologie in Verbindung gebracht. Deepfakes haben das Potenzial, die öffentliche Meinung zu manipulieren, Fehlinformationen zu verbreiten und die öffentliche Sicherheit zu gefährden. Die ethischen Implikationen von Deepfakes machen deutlich, dass robuste Maßnahmen erforderlich sind, um ihre Erstellung zu verhindern und ihre Verbreitung zu erkennen und zu bekämpfen.

Ein grundlegender Aspekt des "eingebauten Datenschutzes" ist die Transparenz. Sie spielt eine entscheidende Rolle beim "eingebauten Datenschutz" und gewährleistet die Rechenschaftspflicht innerhalb von KI-Systemen. Organisationen müssen ihre Datenpraktiken transparent machen und klar erklären, wie KI-Systeme funktionieren und welche Entscheidungen sie treffen. Transparenz in KI-Systemen zu erreichen, kann jedoch aufgrund ihrer Komplexität eine Herausforderung sein. Es ist von entscheidender Bedeutung, Methoden und Werkzeuge zu entwickeln, die es ermöglichen, den Endnutzern algorithmische Vorhersagen auf sinnvolle und verständliche Weise zu erklären.

Weitere Komplikationen ergeben sich daraus, dass die gängigen Praktiken für die sichere Verarbeitung personenbezogener Daten in der Datenwissenschaft und der KI-Technik noch in der Entwicklung sind. Im Rahmen der Einhaltung des Sicherheitsgrundsatzes der DS-GVO sollten Organisationen sicherstellen, dass sie bei der Entwicklung von KI-Systemen und der Verwendung

personenbezogener Daten in einem KI-Kontext die neuesten Sicherheitspraktiken aktiv überwachen und berücksichtigen.

Es ist nicht möglich, alle bekannten Sicherheitsrisiken aufzulisten, die durch den Einsatz von KI zur Verarbeitung personenbezogener Daten verschärft werden könnten. Unabhängig vom Risiko sollten Unternehmen jedoch sicherstellen, dass ihre Mitarbeiter über angemessene Fähigkeiten und Kenntnisse verfügen, um nicht nur Sicherheitsrisiken, sondern auch Datenschutzrisiken zu bewältigen. Hier kommt die Bedeutung von DS-GVO-Schulungen ins Spiel.

Die Wirksamkeit von KI-Modellen hängt in hohem Maße von der Qualität der Daten ab, die sie erhalten, so dass der Datenschutz ein wesentlicher Aspekt ihrer Entwicklung ist. Die Verwendung sensibler Daten beim Training generativer KI-Algorithmen kann dazu führen, dass persönliche Informationen in Chatbot-Ausgaben auftauchen oder die Datensicherheit bei Cyberangriffen gefährdet wird.

Bei der Entwicklung von KI-Produkten ist es daher von größter Bedeutung, personenbezogene Daten von einzelnen Nutzern zu entkoppeln, indem unter anderem synthetische Datensätze mit vollständiger Anonymisierung und nicht umkehrbaren Identifikatoren für das Algorithmus Training, die Prüfung und die Qualitätssicherung verwendet werden. Die Einführung strenger Kontrollen des Datenzugriffs innerhalb des Unternehmens und die Durchführung regelmäßiger Audits können dazu beitragen, Datenschutzverletzungen zu verhindern.

Wichtig ist auch die Erkenntnis, dass mehr Daten nicht unbedingt mit besseren Lösungen gleichzusetzen sind. Das Testen von Algorithmen unter Verwendung von Datenminimierung kann helfen, die geringste Menge an Daten zu ermitteln, die für einen praktikablen Anwendungsfall erforderlich ist. Darüber hinaus ist es von entscheidender Bedeutung, den Nutzern ein rationales Verfahren zur Verfügung zu stellen, mit dem sie die Löschung ihrer personenbezogenen Daten beantragen können.

Die Anwendung von Techniken des kontradiktorischen Lernens, bei denen widersprüchliche Datensätze während des maschinellen Lernprozesses kombiniert werden, kann dazu beitragen, Fehler und Verzerrungen in den Ergebnissen von KI-Algorithmen zu erkennen. Darüber hinaus ist die Verwendung synthetischer Datensätze, die keine tatsächlichen personenbezogenen Daten enthalten, ein möglicher Ansatz, auch wenn weitere Forschung erforderlich ist, um ihre Wirksamkeit zu bewerten.

Organisationen müssen den verantwortungsvollen Einsatz von KI mit den bestehenden Datenschutzgrundsätzen der DS-GVO in Einklang bringen. Diese Leitlinien sollten verschiedene Aspekte wie Rechenschaftspflicht, menschliches Eingreifen, Genauigkeit, Sicherheit, Vermeidung von Verzerrungen und Erklärbarkeit der automatisierten Entscheidungsfindung umfassen. Kontinuierliche Investitionen in Maßnahmen zum Schutz der Privatsphäre, Fortbildung in der Prüfung von Algorithmen und die Einführung von Methoden für Ethik, Sicherheit und Datenschutz durch Design sind notwendig, um die mit generativer KI verbundenen Chancen und Risiken effektiv zu steuern. Technologien wie Differential Privacy bieten Techniken zum Schutz der Privatsphäre, die in generative KI-Systeme integriert werden können. Skalierbare Methoden zur Bereinigung von Datensätzen, einschließlich Deduplizierung und Offenlegungspflichten für Trainingsdaten, tragen zur Bewältigung datenschutzbezogener Herausforderungen bei.

Die kollektiven Bemühungen der Datenschutz- und Technik-Community, gepaart mit dem Engagement einzelner Organisationen und Datenschutzexperten, spielen eine unverzichtbare Rolle bei der Bewältigung der Datenschutzprobleme im Zusammenhang mit generativer KI. Durch die Einhaltung der Grundsätze des "Data Protection by Design" und die Integration umfassender Datenschutz- und Grundrechtsbewertungen können Organisationen eine vertrauenswürdige Implementierung generativer KI anstreben und gleichzeitig die DSGVO einhalten. Es ist wichtig, weiterhin in Datenschutzschulungen zu investieren, Algorithmen zu prüfen und Methoden für Ethik, Sicherheit und Datenschutz durch Design zu integrieren, um die verantwortungsvolle und ethische Nutzung generativer KI sicherzustellen.

7. Techniken zur Verbesserung der Privatsphäre und synthetische Daten

Generative KI-Tools sind komplexe Werkzeuge, und wie alle diese Technologien stellen sie viele bedeutende rechtliche Herausforderungen dar. Generative KI ist datenhungrig, aber solche Daten (insbesondere Qualitätsdaten) können schwer zu beschaffen oder rechtlich geschützt sein, entweder aus Sicht des geistigen Eigentums oder der Datenschutzgesetze.

Aus Sicht des Datenschutzes können Technologien zum Schutz der Privatsphäre (PETs) eine gute Lösung sein, um Datenschutzprobleme im Hinblick auf Datenminimierung, Integrität, Vertraulichkeit und Datenschutz durch Technik anzugehen. Die Agentur der Europäischen Union für Cybersicherheit (ENISA) definiert Technologien zum Schutz der Privatsphäre als "*Software- und*

Hardwarelösungen (z. B. Systeme, die technische Prozesse, Methoden oder Wissen umfassen), um bestimmte Funktionen zum Schutz der Privatsphäre oder des Datenschutzes zu erreichen oder um sich gegen Risiken für die Privatsphäre einer Person oder einer Gruppe natürlicher Personen zu schützen",

Unter den verschiedenen Technologien zum Schutz der Privatsphäre, die im Rahmen der generativen KI eingesetzt werden können, können Datensynthesealgorithmen, die "künstliche" Daten, besser bekannt als synthetische Daten, erzeugen, eine zentrale Rolle spielen.

Laut dem Europäischen Datenschutzbeauftragten (EDSB) *sind "synthetische Daten künstliche Daten, die aus Originaldaten und einem Modell erzeugt werden, das so trainiert wird, dass es die Merkmale und die Struktur der Originaldaten reproduziert (...). Der Generierungsprozess, der auch als Synthese bezeichnet wird, kann mit verschiedenen Techniken durchgeführt werden, wie z. B. Entscheidungsbäumen oder Deep-Learning-Algorithmen. Synthetische Daten können je nach Art der Originaldaten klassifiziert werden: Der erste Typ verwendet reale Datensätze, der zweite verwendet stattdessen von den Analysten gesammeltes Wissen, und der dritte Typ ist eine Kombination aus diesen beiden.*"

Im Wesentlichen handelt es sich bei synthetischen Daten um computergenerierte Daten, die aus vorhandenen realen Daten oder aus Algorithmen und Modellen abgeleitet werden, die Merkmale, Muster und Eigenschaften von realen Daten ganz oder teilweise nachbilden.

Die Verwendung synthetischer Daten kann daher viele Vorteile für das Training generativer KI-Tools bringen, vor allem, wenn es darum geht:

a) verringert die Notwendigkeit, große Mengen an echten personenbezogenen Daten zu sammeln. In der Phase des Trainings von KI-Modellen ist dies besonders wichtig, da Ingenieure auf diese Weise aus relativ kleinen Mengen personenbezogener Daten viel größere Datensätze erzeugen können;
b) ermöglicht eine nahezu perfekte Kennzeichnung (z. B. genau definiert für die Entwicklung eines bestimmten KI-Modells) und qualitativ hochwertigere Daten, wodurch reale Datensätze ergänzt oder ersetzt werden können. Eine Studie von Gartner hat vorausgesagt, dass "*bis 2024 60 % der für die Entwicklung von KI- und Analyseprojekten verwendeten Daten synthetisch erzeugt werden*";
c) bei ordnungsgemäßer Erkennung und Korrektur die Verzerrung oder statistische Unausgewogenheit der Originaldatensätze verringern und damit die Fairness der auf den Daten basierenden Entscheidungsfindung erhöhen;

d) stärkt die Privatsphäre und verringert die Angriffsfläche für Cybersicherheit, indem es das Risiko des Verlusts der Vertraulichkeit, Integrität oder Verfügbarkeit echter personenbezogener Daten begrenzt;
e) senkt die Kosten in allen Phasen der Datenwertschöpfungskette, indem es den Bedarf an übermäßiger Datenerfassung, -bereinigung, -aufbereitung und -speicherung begrenzt.

Das bedeutet jedoch nicht, dass synthetische Daten die vollständige Lösung für alle Datenschutzprobleme sind. Es gibt immer noch einige rechtliche Bedenken, die von den Datenschutzbeauftragten berücksichtigt werden müssen.

Erstens entsprechen synthetische Daten nicht notwendigerweise anonymen Daten, was bedeutet, dass das Risiko der Re-Identifizierung in dem einen oder anderen Maße bestehen bleibt. In der Praxis zielen synthetische Daten darauf ab, reale Daten zu replizieren, und je genauer sie ein Proxy sind, der alle Merkmale und Muster der Originaldaten beibehält, desto effizienter sind sie für das generative KI-Modell, das auf diesen Daten trainiert wird; der Nachteil ist jedoch, dass diese Effizienz in direktem Verhältnis das Risiko der **Re-Identifizierung** erhöht. Das bedeutet, dass das Risiko, aus dem synthetischen Datensatz oder dem KI-Modell selbst auf Daten zu schließen, die sich auf eine bestimmte Person beziehen, nicht beseitigt wird.

Die britische Datenschutzbehörde ICO weist darauf hin. "*Sie sollten sich darauf konzentrieren, inwieweit Personen in den synthetischen Daten identifiziert oder identifizierbar sind und welche Informationen über sie offengelegt würden, wenn die Identifizierung erfolgreich ist. Einige Methoden zur Generierung synthetischer Daten haben sich als anfällig für Modellinversionsangriffe, Angriffe zur Ableitung von Zugehörigkeiten und das Risiko der Offenlegung von Attributen erwiesen. Diese können das Risiko erhöhen, auf die Identität einer Person zu schließen....*

Die Verwendung anderer Technologien zum Schutz der Privatsphäre (z. B. differentieller Datenschutz) oder die Unterdrückung von Ausreißern (Datenpunkte mit eindeutigen Identifizierungsmerkmalen) kann das Risiko der Re-Identifizierung personenbezogener Daten verringern, aber nicht vollständig ausschließen.

Darüber hinaus kann die Generierungsphase synthetischer Daten die Verarbeitung personenbezogener Daten beinhalten, insbesondere bei der Sammlung und Analyse realer Datensätze, was die Einhaltung der DS-GVO und der damit verbundenen Verpflichtungen erforderlich macht.

Besonders zu erwähnen ist auch die Pflicht zur umfassenden Information der betroffenen Personen gemäß Art. 13 DSGVO über die betroffenen Personen, deren Daten erhoben und dann für KI-Schulungszwecke verwendet werden, sowie über die Ermittlung einer rechtmäßigen Grundlage für die Verarbeitung gemäß Art. 6 der GDPR.

Schließlich besteht die Verpflichtung zur strikten Einhaltung der Grundsätze gemäß Art. 5 der GDPR immer dann, wenn personenbezogene Daten betroffen sind. Insbesondere einige der folgenden Grundsätze aus Art. 5 sind im Falle der generativen KI erwähnenswert:

a) Transparenz: Dies beschränkt sich nicht auf die Informationen, die den betroffenen Personen gemäß Art. 13 DSGVO, sondern auch gegenüber den Nutzern in Bezug auf die von KI-Modellen erzeugten synthetischen Ergebnisse, um das Risiko von Fälschungen und/oder sozialen Manipulationen zu vermeiden;
b) Zweckbeschränkung: Da synthetische Daten aus realen Daten abgeleitet werden können, die personenbezogene Informationen enthalten können, muss dargelegt werden, dass diese Daten für festgelegte, eindeutige und rechtmäßige Zwecke erhoben wurden und dass die weitere Verarbeitung (z. B. für die Synthetisierung von Daten und das anschließende KI-Modell-Training) mit den ursprünglichen Zwecken nicht unvereinbar ist.

   Ein ähnlicher Grundsatz wurde in Bezug auf den Anonymisierungsprozess durch WP Art. 29 (Stellungnahme 5/2014) festgelegt, wonach: "*Der Anonymisierungsprozess, d. h. die Verarbeitung (...) personenbezogener Daten zum Zwecke ihrer Anonymisierung, ist ein Fall von "Weiterverarbeitung". Als solche muss diese Verarbeitung der Vereinbarkeitsprüfung gemäß den Leitlinien der Datenschutzgruppe in ihrer Stellungnahme 03/2013 zur Zweckbindung genügen*".

   Insbesondere im Hinblick auf die Trainingsphase von KI-Modellen ist der Verweis auf die "statistischen Zwecke" nicht grundsätzlich unvereinbar mit den ursprünglichen Zwecken gemäß Buchstabe b) von Art. 5 Abs. 1 nicht grundsätzlich unvereinbar ist, könnte diesem Zweck dienen[6] .

c) Genauigkeit und Fairness: Hier muss darauf geachtet werden, dass das Risiko einer "Halluzination" oder einer Duplizierung von Verzerrungen, Fehlern oder Ungenauigkeiten,

[6] Siehe zu diesem Thema die Studie im Auftrag des Gremiums für die Zukunft von Wissenschaft und Technologie (Europäischer Parlamentarischer Forschungsdienst) "*The Impact of the GDPR on artificial intelligence*", Juni 2020

die im Originaldatensatz enthalten sind, vermieden wird. Dies ist besonders wichtig, wenn das mit den synthetischen Daten trainierte KI-Modell anschließend dazu verwendet wird, Entscheidungen zu treffen, die die Rechte oder Interessen der Menschen beeinträchtigen könnten.

Von größter Bedeutung für diesen Zweck ist die Entwicklung von Techniken, die es ermöglichen, die Ergebnisse von KI-Systemen zu erklären, die mit Hilfe von synthetischen Daten trainiert wurden.

8. Spezifische Probleme der bild- und audiobasierten generativen KI

Bei nicht-textbasierten generativen KI-Anwendungen, wie z. B. Tools zur Erzeugung von Bildern, Audios und Videos, gibt es eindeutige Auswirkungen auf den Datenschutz. Beliebte Anwendungen wie beispielsweise Midjourney und Stable Diffusion, die es Nutzern ermöglichen, durch die Eingabe von Textaufforderungen schnell Bilder und Videos zu erzeugen, basieren auf großen Mengen von Bild- und Videoinhalten. Diese zugrundeliegenden Daten enthalten zahlreiche Kategorien personenbezogener Daten, die ausreichen, um die betroffenen Personen zu identifizieren, wobei die wichtigste Kategorie das Bild und die Ähnlichkeit einer betroffenen Person ist, die oft in den Ergebnissen dargestellt wird.

Im Einzelnen können die behördlichen Datenschutzbeauftragten davon ausgehen, dass die folgenden Kategorien personenbezogener Daten in solche Instrumente einbezogen werden:

- Fotos von betroffenen Personen;
- künstlerische Darstellungen von betroffenen Personen;
- Videomaterial von betroffenen Personen; und
- Audio, sprachbasierte Daten

Die Unternehmen müssen sich darüber im Klaren sein, dass die Weiterverarbeitung solcher Daten in den Anwendungsbereich der DS-GVO fällt. Wenn beispielsweise eine Marketingabteilung Werbematerial erstellen möchte und Bilder von betroffenen Personen verwendet, die durch generative KI gewonnen wurden, muss sie diese Bilder im Einklang mit den Datenschutzgesetzen verarbeiten und grundlegende Prinzipien wie Transparenz, Rechtmäßigkeit und Fairness beachten.

Außerdem sollte die Frage der Kombination von Daten aus generativen KI-Quellen mit Daten aus anderen Quellen in Betracht gezogen werden. Während die vom generativen KI-Tool erhaltenen Daten die betroffene Person möglicherweise nicht identifizieren, kann dies bei der Kombination mit anderen Daten der Fall sein, was wiederum die Anforderungen der DS-GVO ins Blickfeld rückt. Dies könnte besonders relevant sein, wenn zum Beispiel das Zusammenfügen von Bildern aus verschiedenen Quellen zur Identifizierung von Personen führt.

In den kreativeren Verwendungsfällen, in denen Organisationen die Darstellung von Bildern, Videos oder Audioinhalten ändern, umgestalten oder wesentlich verändern möchten, sollte dies unter Beachtung der Grundrechte und -freiheiten der betroffenen Personen erfolgen. Risiken, z. B. die Diffamierung oder Schädigung betroffener Personen, sollten immer berücksichtigt werden, und wenn davon ausgegangen wird, dass die Verarbeitung mit einem hohen Risiko verbunden sein könnte, sollte eine Datenschutzfolgenabschätzung durchgeführt werden.

Wenn Organisationen legitime "Deepfake"-Inhalte erstellen wollen, wie z. B. offizielle Unternehmensvideos, sollten Fragen der Zustimmung der betroffenen Personen und der Transparenz der Verarbeitung im Vordergrund stehen.

9. Umgang mit Datenschutzrisiken

Die Durchführung einer Datenschutz-Folgenabschätzung (DPIA) bei der Einführung oder Nutzung eines generativen KI-Systems ist umso wichtiger, wenn, wie es oft der Fall ist, diese Werkzeuge noch nicht richtig verstanden wurden, sowohl aus der Perspektive der Geschäftsstrategie als auch des Risikomanagements. Das Verständnis der Risiken für personenbezogene Daten, die sich aus der generativen KI-Verarbeitung ergeben, ist noch in der Entwicklung begriffen, und alle Datenschutzbeauftragten müssen versuchen, sich auf noch nicht vorhersehbare Bedrohungen und Herausforderungen einzustellen. Um diese aufkommenden Risiken zu bewältigen, sollten die folgenden Faktoren in Betracht gezogen werden.

a) Risiken für die betroffenen Personen

Die Beziehung zwischen dem Nutzer und der KI sowie die Auswirkungen, die die Verarbeitung auf den Einzelnen haben wird, sollten im Mittelpunkt der Analyse stehen. Zu den potenziellen Risiken für die betroffenen Personen gehören:

- Auswirkungen einer teilweise oder vollständig automatisierten Entscheidung durch generative KI. Die Folgen solcher Entscheidungen können in finanziellen Opportunitätsverlusten oder sogar in der Einschränkung von Grundrechten bestehen.

- Risiken der Verstärkung von Diskriminierung und Vorurteilen gegenüber bestimmten Nutzern.

- Risiken, die sich aus der Verarbeitung von Daten besonderer Kategorien gemäß Art. 9 DS-GVO. Beispielsweise könnte ein generatives KI-Tool aus bestimmten personenbezogenen Daten der betroffenen Person (aus ihren Ausdrucksweisen oder der Verwendung bestimmter Wörter) auf ihre ethnische Herkunft, ihre politische oder philosophische Einstellung oder sogar auf ihre sexuelle Ausrichtung schließen und auf dieser Grundlage eine unterschiedliche Behandlung anwenden. Um solche Risiken zu erkennen, sollte das Unternehmen, das das generative KI-Tool einsetzt, eine regelmäßige Überprüfung der Qualität der generierten Ergebnisse durchführen.

- Im Hinblick auf die IT-Sicherheit können die Informationen, die dem Angreifer im KI-System zur Verfügung stehen, einen Bedrohungsvektor darstellen. Ein sogenanntes "White Box"-Szenario, bei dem der Angreifer viele technische Informationen ableiten/finden kann, um seinen Angriff vorzubereiten, schafft eine größere Gefährdung im Vergleich zu einem "Black Box"-System, bei dem der Angreifer nur auf die vom System als Ausgabe produzierten Informationen zugreifen kann. Insbesondere die folgenden Angriffe sind spezifisch für bestimmte KI-Projektschritte:

Lernphase	Angriffstyp	Infektion	Backdooring-Angriffe
			Vergiftungsanschläge
		Exfiltration	Angriffe durch Mitgliedschaftsinferenz
			Modellinversionsangriffe
			Angriffe auf die Modellextraktion

Produktionsphase	Angriffstyp	Manipulation	Ausweichmanöver
			Angriffe durch Umprogrammierung
			Verweigerung von Dienstleistungen
		Exfiltration	Angriffe durch Mitgliedschaftsinferenz
			Modellinversionsangriffe
			Angriffe auf die Modellextraktion

b) Identifizierung von Abhilfemaßnahmen

Die Datenschutz-Folgenabschätzung sollte wie immer vor Projektbeginn durchgeführt werden und sollte dann über den "eingebauten Datenschutz" die Entwurfsphase für jedes generative KI-Tool informieren und leiten. Im Fall von generativer KI sollten die folgenden Abhilfemaßnahmen berücksichtigt werden, um die ermittelten Risiken zu bewältigen:

- Überwachtes Feintuning mit Gesprächsbeispielen, bei dem ein LLM darauf trainiert wird, einen Gesprächskorpus zu reproduzieren, der ein gewünschtes Verhalten veranschaulicht.

- Feinabstimmung mit einem menschlichen Wertemodell, bei dem menschliche Bediener die zufriedenstellendsten Ergebnisse erzielen.

- Darüber hinaus sollten organisatorische Maßnahmen darauf abzielen, eine ständige Evaluierung der vom generativen KI-Tool gelieferten Ergebnisse zu gewährleisten, und zwar sowohl auf der Ebene des menschlichen Bedieners, der das Tool einsetzt, als auch auf der Ebene einer organisatorischen Einheit, die die Ergebnisse in großem Maßstab analysiert, um eine hohe Ergebnisqualität über einen längeren Zeitraum hinweg sicherzustellen.

- Ebenso sollten wir uns so weit wie möglich um eine Begründung der von dem generativen KI-Modell getroffenen Entscheidungen bemühen, um eine echte menschliche Kontrolle zu ermöglichen. In dieser Hinsicht bleibt die menschliche Kontrolle letztlich die beste

> Methode, um die von generativen KI-Systemen ausgehenden Risiken zu mindern. Auf diese Weise kann ein übermäßiges Vertrauen in die von generativen KI-Tools erzielten Ergebnisse vermieden werden. Ein solches übermäßiges Vertrauen würde in Ermangelung wirksamer menschlicher Kontrollen dazu führen, dass völlig automatisierte Entscheidungen getroffen werden.

Eine zusätzliche Überlegung für die DSB ist die sich abzeichnende Verpflichtung zur Durchführung von Folgenabschätzungen für die Grundrechte (Fundamental Rights Impact Assessments, FRIAs). Der Entwurf des KI-Gesetzes, der sich zum Zeitpunkt der Veröffentlichung dieses Papiers noch in der Trilog-Phase der Diskussionen innerhalb der EU-Legislative befindet, enthält eine Anforderung zur Durchführung von FRIAs. Die Absicht ist, dass eine solche Bewertung entweder von einem Anbieter oder einem Nutzer eines KI-Systems durchgeführt werden muss, wenn Risiken für die Grundrechte und -freiheiten von Personen bestehen, die von der Ausgabe betroffen sind.

Angesichts der Tatsache, dass FRIAs in der Welt der KI faktisch DSFA ähneln, mit besonderen Überschneidungen beim Verständnis, wie sich Verarbeitungstätigkeiten auf Grundrechte auswirken, sollten die DSB damit rechnen, dass ihnen diese Arbeit übertragen wird, sobald das KI-Gesetz in Kraft tritt. Obwohl die DSB in gewisser Hinsicht in der Lage und qualifiziert sind, diese Arbeit zu leisten, sind sie nicht unbedingt von Natur aus mit den neuartigen technologischen Risiken vertraut, die durch die KI-Technologien rasch geschaffen werden. Aus diesem Grund sollten die DSB bereits jetzt die KI-spezifischen Risiken für personenbezogene Daten erforschen und verstehen.

Aus praktischer Sicht kann es möglich sein, FRIAs und DSFA in einem Arbeitsgang durchzuführen, aber unabhängig davon, welche Methode letztendlich gewählt wird, müssen die behördlichen Datenschutzbeauftragten bereits jetzt, im Vorgriff auf das AI-Gesetz, damit beginnen, Kenntnisse über AI-Risiken zu entwickeln.

10. Transparenz und generative KI

Bei der Erfassung und Übermittlung von Daten, einschließlich personenbezogener Daten, an eine KI zu Schulungszwecken und wenn diese Datenverarbeitung unter die DS-GVO fällt, muss die Einrichtung, die diese Schulung durchführt (der KI-Betreiber), die Transparenz dieser Datenverarbeitung gemäß Artikel 5 Abs. 1 lit. a) und Artikel 12 ff. der Verordnung gewährleisten.

Es lassen sich drei verschiedene Datenquellen unterscheiden:

- Das Scraping von Daten aus Websites mit Hilfe von Robotern oder KI-Systemen (Anwendungsfall 1);
- Die Bereitstellung von Daten durch Nutzer des Systems oder Datenlieferanten über andere Personen (Anwendungsfall 2);
- Die Bereitstellung von Daten über sich selbst durch die Nutzer der KI (Anwendungsfall 3).

Für jeden dieser Anwendungsfälle gibt es je nach Art der erforderlichen Trainings-KI unterschiedliche Möglichkeiten, die Transparenz der Datenverarbeitung zu gewährleisten.

Anwendungsfall 1

Transparenz ist ein heikles und vielleicht schwieriges Thema, wenn es um Online-Data-Scraping geht, vor allem aufgrund der Tatsache, dass alle auf diese Weise gesammelten personenbezogenen Daten nicht direkt bei der betroffenen Person erhoben werden. Daher sollte Art. 14 der DS-GVO für solche Daten gelten, d. h. personenbezogene Daten, die nicht direkt bei der betroffenen Person erhoben wurden, berechtigen die betroffene Person dazu, von dem für die Verarbeitung Verantwortlichen eine Bestätigung darüber zu erhalten, ob ihre personenbezogenen Daten verarbeitet werden, und wenn dies der Fall ist, sollte ihr Zugang zu ihren personenbezogenen Daten gewährt werden, zusammen mit anderen wichtigen Informationen wie dem Zweck der Verarbeitung und den Datenkategorien, die verarbeitet werden, usw.

Darüber hinaus sollte Art. 15 der DS-GVO über das Recht der betroffenen Person auf Auskunft über ihre personenbezogenen Daten gelten.

In einem solchen Szenario ergeben sich für den KI-Betreiber jedoch mehrere Schwierigkeiten. Insbesondere die folgenden:

- Identifizierung personenbezogener Daten unter den von der KI automatisch abgerufenen Daten, die in der Regel aus großen Datenmengen bestehen;
- Direkte Identifizierung jeder einzelnen betroffenen Person;
- Beschaffung ausreichender Kontaktinformationen, um jede betroffene Person über die Verarbeitung ihrer Daten zu informieren.

In Anbetracht dieser Schwierigkeiten könnte Art. 14 Abs. 5 lit. b der DS-GVO angewandt werden. In diesem Abschnitt des Artikels heißt es, dass ein für die Verarbeitung Verantwortlicher nicht verpflichtet ist, jeder betroffenen Person die angegebenen Informationen zu erteilen, wenn "*sich die Erteilung dieser Informationen als unmöglich erweist oder mit einem unverhältnismäßigen Aufwand verbunden wäre*". Die Rechtsprechung verschiedener Datenschutzbehörden zeigt, dass diese Ausnahme sehr streng ausgelegt werden sollte. In Anbetracht der oben genannten Schwierigkeiten bei generativen KI-Modellen könnte sie jedoch hier angewandt werden. In diesem Fall wäre der KI-Betreiber jedoch weiterhin an die Transparenzanforderungen gegenüber der betroffenen Person gebunden.

Gemäß dem genannten Art. 14 Abs. 5 lit b sollte der für die Verarbeitung Verantwortliche geeignete Maßnahmen ergreifen, um die Rechte und Freiheiten sowie die berechtigten Interessen der betroffenen Person zu schützen. Zu diesen Maßnahmen gehört die Veröffentlichung der Datenschutzrichtlinien des für die Verarbeitung Verantwortlichen auf seiner Website, aber auch möglicherweise strengere Maßnahmen wie das Beispiel der italienischen Datenschutzbehörde bei der Regulierung von ChatGPT im Jahr 2023. Schließlich erklärte sich OpenAI bereit, eine Informationskampagne ohne Werbecharakter in allen wichtigen italienischen Massenmedien (Radio, Fernsehen, Zeitungen und Internet) durchzuführen, um die Menschen über die wahrscheinliche Erhebung ihrer personenbezogenen Daten zum Zweck der Schulung von ChatGPT zu informieren. Sie erklärten sich auch bereit, auf der Website des für die Verarbeitung Verantwortlichen ein Tool zur Verfügung zu stellen, mit dem alle interessierten Personen ihr Recht auf Zugang zu ihren personenbezogenen Daten wahrnehmen können.

Andererseits kann für ein solches Auskunftsrecht auch Art. 11 der DS-GVO gelten, der Folgendes vorsieht

"*(1) Erfordern die Zwecke, für die ein für die Verarbeitung Verantwortlicher personenbezogene Daten verarbeitet, nicht oder nicht mehr die Identifizierung einer betroffenen Person durch den für die Verarbeitung Verantwortlichen, so ist der für die Verarbeitung Verantwortliche nicht verpflichtet, zusätzliche Informationen aufzubewahren, zu beschaffen oder zu verarbeiten, um die betroffene Person zu identifizieren, und zwar ausschließlich zum Zweck der Einhaltung dieser Verordnung.*

(2) Kann der für die Verarbeitung Verantwortliche in den in Absatz 1 des vorliegenden Artikels genannten Fällen nachweisen, dass er nicht in der Lage ist, die betroffene Person zu identifizieren, so teilt er dies der betroffenen Person, soweit möglich, mit. In diesen Fällen finden die Artikel 15 bis 20

keine Anwendung, es sei denn, die betroffene Person stellt zur Ausübung ihrer Rechte nach diesen Artikeln zusätzliche Informationen zur Verfügung, die ihre Identifizierung ermöglichen".

Darüber hinaus werden wir in Erwägungsgrund 4 der Datenschutz-Grundverordnung daran erinnert, dass "*das Recht auf den Schutz personenbezogener Daten kein absolutes Recht ist; es muss im Verhältnis zu seiner Funktion in der Gesellschaft betrachtet und gemäß dem Grundsatz der Verhältnismäßigkeit gegen andere Grundrechte abgewogen werden*". Infolgedessen könnte argumentiert werden, dass dem KI-Betreiber keine unverhältnismäßigen Anstrengungen auferlegt werden können, um den Antragsteller zu identifizieren und seine personenbezogenen Daten in den Trainingsdaten der KI zu erkennen.

In Anbetracht der obigen Ausführungen sollte der KI-Betreiber, der mit einem Zugangsantrag konfrontiert wird,:

1. Überprüfen Sie, ob die personenbezogenen Daten des Antragstellers identifiziert werden können;
2. Geben Sie dem Antragsteller alle identifizierten persönlichen Daten;
3. Unterrichtung der betroffenen Person darüber, dass es personenbezogene Daten über sie geben kann, die der KI-Betreiber angesichts der Merkmale der durchgeführten Datenverarbeitung nicht erkennen/zur Verfügung stellen kann.

Um Art. 25 der DS-GVO und dem Grundsatz des "eingebauten Datenschutzes" gerecht zu werden, kann der KI-Betreiber auch verpflichtet werden, nachzuweisen, dass er solche Auskunftsersuchen vorhersehen kann und dass er alle technischen Möglichkeiten geprüft hat, die er vernünftigerweise einsetzen könnte, um die personenbezogenen Daten jedes Antragstellers zu ermitteln (und dass er diese Möglichkeiten regelmäßig neu bewertet).

Anwendungsfall 2

Da die Daten den KI-Betreibern entlang der Lieferkette in der Regel von anderen Dritten geliefert werden, die in der Lieferkette weiter oben stehen (ein Nutzer oder ein Datenlieferant). Diese Dritten könnten den KI-Betreiber bei der Gewährleistung von Transparenz bei der Datenverarbeitung unterstützen, indem sie Instrumente und Anleitungen zur Verfügung stellen, wie personenbezogene Daten am besten aus dem Datensatz extrahiert werden können, da es diese Dritten sind, die die Datensätze überhaupt erst bereitstellen. Diese Dritten könnten dem KI-

Betreiber aus denselben Gründen auch bei der Bearbeitung von Auskunftsersuchen betroffener Personen helfen.

Anwendungsfall 3

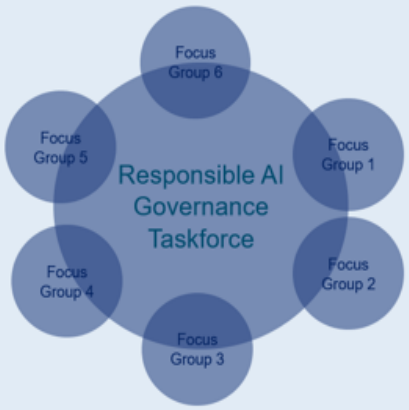
Wenn personenbezogene Daten direkt bei den Nutzern erhoben werden, gilt Art. 13 der DS-GVO. Der für die Verarbeitung Verantwortliche muss der betroffenen Person zum Zeitpunkt der Erhebung spezifische Informationen zur Verfügung stellen, z. B. die Identität und die Kontaktdaten des für die Verarbeitung Verantwortlichen, die Kontaktdaten seines Datenschutzbeauftragten, die Zwecke der Verarbeitung, für die die personenbezogenen Daten bestimmt sind, sowie die Rechtsgrundlage für die Verarbeitung und andere spezifische Informationen.

11. Optimierung der organisatorischen Strukturen

In jeder Organisation muss das Thema Generative KI aus Sicht der Managementstruktur mehrdimensional angegangen werden, um der Komplexität der Technologie und ihrer Auswirkungen Rechnung zu tragen. Es wird für Unternehmen nicht praktikabel sein, jede Funktion für sich arbeiten zu lassen und nicht mit den anderen zu interagieren.

Die Auswirkungen der künstlichen Intelligenz sind ein unternehmensweites Problem und erfordern daher ein gemeinsames unternehmensweites Konzept. Ein solcher integrierter Ansatz ist unerlässlich, um Doppelarbeit zu vermeiden, vor allem aber, um sicherzustellen, dass wichtige Entscheidungen von mehreren Disziplinen getragen werden.

Um dies zu erreichen, sollten Organisationen eine KI-Taskforce einrichten, die sich auf verantwortungsvolle KI und deren Steuerung konzentriert. Die Einrichtung einer solchen Taskforce könnte eine Initiative des DSB sein, da er eine der Funktionen ist, die am stärksten mit diesem Thema konfrontiert ist, da er einige KI-Überlegungen in einem Kontext verwalten muss, in dem personenbezogene Daten betroffen sind. Alternativ könnte sie auch von einer IT-Funktion initiiert und geleitet werden, etwa von einem Chief Data Officer oder Chief Technology Officer.



Focus Group 1: Relevant AI Tools
Focus Group 2: Relevant AI Use Cases
Focus Group 3: IP Protection, Business Secrets & Liability
Focus Group 4: Data Protection & Data Ethics
Focus Group 5: Security & Safety
Focus Group 6: AI Guidance & Communications

An dieser Task Force werden die Rechtsabteilung, die Compliance-Funktionen und insbesondere der Datenschutz maßgeblich beteiligt sein. Für die technischen Aspekte sollte die Abteilung IT-Sicherheit vertreten sein. Die Taskforce kann Mitarbeiter der Kommunikations- und Presseabteilung einbeziehen, da es notwendig sein wird, intern und möglicherweise auch extern über die von der Taskforce getroffenen Entscheidungen zu kommunizieren. Der Leiter der Taskforce kann Schwerpunktgruppen einrichten, in denen sich ausgewählte Mitglieder der Taskforce auf bestimmte Fragen konzentrieren und der Taskforce über ihre Ergebnisse berichten. Das obige Diagramm vermittelt einen Eindruck von der Zusammensetzung dieser Fokusgruppen und ihrer Beziehung zur Taskforce "Responsible AI Governance".

Die Aufgabe der Task Force besteht darin, auf die unmittelbare Notwendigkeit einer verantwortungsvollen KI-Governance innerhalb der Organisation zu reagieren und die Risiken beim Einsatz generativer KI zu untersuchen und zu bewältigen, insbesondere im Hinblick auf personenbezogene Daten, Voreingenommenheit, ethische Bedenken, aufkommende KI-Regelungen und zahlreiche rechtliche Fragen wie geistige Eigentumsrechte und Haftungsrisiken.

Das Hauptziel dieser Task Force ist die Festlegung eines Aktionsplans. Ein entscheidender Aspekt dieses Aktionsplans wird die Durchführung einer Bestandsaufnahme der im Unternehmen eingesetzten KI-Systeme sein, zu denen auch die generative KI gehört. Ein weiterer wichtiger Aspekt ist die Festlegung von Rollen und Verantwortlichkeiten für alle Funktionen innerhalb der Gruppe.

Die Aufgabe dieser KI-Taskforce besteht auch darin, das Bewusstsein für KI-Fragen auf allen Ebenen des Unternehmens zu schärfen. Dieser Punkt ist wichtig, da das Risiko natürlich von den Mitarbeitern ausgeht, die die Technologie tagtäglich nutzen, aber es muss mit der höchsten Ebene der Entscheidungsfindung verknüpft werden, denn die Entscheidung über die Art und Weise der Nutzung (oder Nichtnutzung) generativer KI ist eine Unternehmensstrategie.

Als erste Aufgabe sollte die Task Force einen vorläufigen Leitfaden für die Organisation hinsichtlich der verantwortungsvollen Nutzung generativer KI erstellen, der beispielsweise die Empfehlung enthält, keine personenbezogenen Daten in Eingabeaufforderungen einschlägiger Tools wie ChatGPT einzugeben oder Bilder mit identifizierbaren Personen hochzuladen.

Unabhängig von der Komplexität der Technologie und ihrer Umsetzung besteht die Rolle des DSB in dieser Taskforce letztlich darin, sicherzustellen, dass alle mit KI-Technologien verarbeiteten personenbezogenen Daten mit der DS-GVO konform sind.