Verein Unternehmens-Datenschutz

Verwendung generativer KI

Leitfaden zum Datenschutzgesetz

Entwurf "for public comment" | 29. August 2023

Einhaltung des Datenschutzgesetzes bei der Verwendung generativer KI

Dieser Leitfaden legt dar, wie das geltende Schweizer Datenschutzgesetz (DSG) in Bezug auf die Verwendung von generativer KI in der Praxis ausgelegt werden kann, so dass Unternehmen im Anwendungsbereich des DSG ihr Verhalten datenschutzkonform gestalten können. Der Leitfaden umschreibt das Thema nicht abschliessend, sondern nimmt zu einigen der sich in der Praxis aktuell stellenden Fragen Stellung.

Der Leitfaden gilt nicht für die Datenschutz-Grundverordnung der EU (DSGVO), auch wenn viele der Grundsätze dort analog gelten werden. Bei der DSGVO sind jedoch die Systemunterschiede zu beachten, insbesondere das Erfordernis einer Rechtsgrundlage für jede Bearbeitung von Personendaten. Die folgenden Ausführungen beziehen sich einzig auf das Datenschutz- bzw. Persönlichkeitsschutzrecht, nicht auf andere Rechtsgebiete. Er berücksichtigt nicht branchenspezifische Regulierung.

Der Leitfaden behandelt auch ausdrücklich nicht ethische, reputative oder sonst übergesetzliche Aspekte des Einsatzes von KI (z.B. ob über die Verwendung von generativer KI informiert werden muss, auch wenn dies rein datenschutzrechtlich nicht nötig wäre). Diese muss jedes Unternehmen für sich entscheiden.

Der Leitfaden gibt nicht zwingend die Ansichten aller Mitglieder des VUD wieder.

Hinweise zur internen Umsetzung der Vorgaben des DSG

Das DSG ist technologieneutral, was die Vorgaben zum Umgang mit Personendaten betrifft; es sagt, was zu tun ist und welche Ziele eingehalten werden sollen. Diese Vorgaben sind immer dieselben, ganz gleich, ob KI eingesetzt wird oder nicht.

Das DSG sagt hingegen nicht, mit welchen Hilfsmitteln und anderen technischen und organisatorischen Massnahmen wie z.B. Prozesse die Vorgaben des DSG sicher-gestellt werden sollen. Wie beispielsweise sichergestellt wird, dass der Output einer generativen KI für den konkreten Zweck ausreichend ist. Sie werden auch "TOM" genannt und sind vom Anwendungsfall und den Umständen abhängig. Gerade hier zeigen sich im Falle von KI die wirklichen Neuerungen, etwa weil neue Verfahren und technische Mittel nötig werden, um die "altbekannten" Vorgaben weiterhin einzuhalten.

Auch dieser Leitfaden schreibt nicht vor, welche TOM wie zum Einsatz kommen sollen. Hierbei ist jedes Unternehmen letztlich frei. Dieser Leitfaden soll nur (aber immerhin) zum Verstehen der Vorgaben des DSG beitragen, indem er Überlegungen zur möglichen Interpretation macht, so dass die für den konkreten Fall angemessenen TOM festgelegt werden können. Dabei sind nebst den Vorgaben des DSG selbst für den Datenschutz noch weitere Aspekte zu berücksichtigen, wie etwa Vorgaben und Usanzen einer Branche. Diese definiert dieser Leitfaden auch nicht.

Zu bedenken ist schliesslich, dass das beste Verständnis der Vorgaben des DSG nichts nutzt, wenn ein Unternehmen zwei Dinge nicht hat: Erstens eine Vorstellung, wohin es will, insbesondere das jeweilige "Target Operating Model". Und zweitens eine solide Governance, d.h. die Aufgaben, Kompetenzen und Verantwortlichkeiten sind intern klar geregelt und werden so gelebt. Anders klappt es in einer stark arbeitsteiligen Welt nicht.

Verantwortlichkeit

Bearbeitungsgrundsätze und Informationspflicht

Betroffenenrechte und Compliance-Massnahmen

Unterschiedliche Verantwortlichkeiten für Erstellung, Besitz und Verwendung

Es ist betreffend die datenschutzrechtliche Verantwortlichkeit zwischen der Erstellung eines KI-Modells (z.B. durch Training; wie ein Autohersteller), dessen Besitz (z.B. durch eine lokale Installation oder Anbieten als Service auf eigenen Systemen; wie ein Taxifahrer) und dessen Verwendung (für die Erzeugung von Output auf bestimmten Input hin; wie ein Taxigast) zu trennen. Wenn dabei Personendaten zum Einsatz kommen bzw. das Modell solche enthält, handelt es sich um drei separate Datenbearbeitungen, für welche jeweils verschiedene Stellen (allein) verantwortlich sein können.

Die drei Datenbearbeitungen sind abgeschlossene Vorgänge, die jeweils einem eigenen Zweck dienen und ihre eigenen Mittel aufweisen. Der Verwender eines KI-Modells hat oft keinen Einfluss auf die Erstellung des Modells. Für ihn ist das KI-Modell dann nur ein Werkzeug, das er zur Bearbeitung seiner Daten benutzt. Wie und wozu er es einsetzt, bestimmt alleine er.

Wer wiederum lediglich Hersteller eines Werkzeugs ist, entscheidet nicht über dessen Zweck und Mittel. Enthält das KI-Modell selbst Personendaten (bei einem Large Language Model typischerweise der Fall), so stellt sein Besitz (verstanden als die tatsächliche "Gewalt" über das Modell bzw. ein Exemplar davon und damit dessen Aufbewahrung) ebenfalls eine Bearbeitung von Personendaten dar. Wer ein solches Modell mit Personendaten auf seinen eigenen Systemen speichert, bearbeitet daher Personendaten, auch wenn er das Modell nicht benutzt. Wer ein solches Modell hingegen nur als Service nutzt (z.B. in der Cloud), trifft keine Entscheide betreffend das Modell an sich und dessen Aufbewahrung bzw. Haltung, sondern nur bezüglich seines Einsatzes für eigene Zwecke – er stellt dem Modell (im Besitz eines anderen, z.B. eines Providers) gewissermassen eine Frage und erhält eine Antwort darauf. Er ist blosser Verwender. Die Praxis zeigt freilich, dass die Einordnung nicht immer einfach oder klar ist.

Verantwortlichkeit des Verwenders generativer KI

Der Verwender eines Modells bzw. einer damit betriebenen Anwendung muss sicherstellen, dass die Verwendung des erzeugten Outputs in Bezug auf die darin allfällig enthaltenen Personendaten den datenschutzrechtlichen Grundsätzen genügt bzw. eine allfällige Verletzung gerechtfertigt ist. Für die Rechtmässigkeit der Erzeugung bzw. der Erstellung des Modells oder des damit potenziell möglichen Outputs ist er als Verwender hingegen nicht verantwortlich.

Es handelt sich wie gezeigt um unterschiedliche Datenbearbeitungen. Der Verwender eines Werkzeugs kann für dessen allfällig unrechtmässige Herstellung nicht verantwortlich sein, erst recht nicht, wenn das Werkzeug nicht in seinem Besitz ist, sondern er es lediglich einsetzt. Erst wenn sich eine im Modell angelegte Unrechtmässigkeit im konkret erzeugten Output fortsetzt (z.B. wenn dieser fehlerhafte Personendaten enthält), kann eine Verantwortlichkeit des Verwenders greifen (die wiederum durch eine Kontrolle des Outputs mitigiert werden kann). Auch dann ist jeweils unabhängig vom Modell zu beurteilen, ob eine unrechtmässige Datenbearbeitung vorliegt (z.B. ob die Personendaten im Hinblick auf den Zweck tatsächlich fehlerhaft sind und ihre Bearbeitung nicht gerechtfertigt ist).

Verantwortlichkeit bei Bereitstellung als Service

Wer ein Modell besitzt und es im Rahmen eines Service Dritten zur Nutzung anbietet, ist als Besitzer des Modells ein Verantwortlicher. Er muss daher z.B. die Einhaltung von Betroffenenrechten in Bezug auf das Modell als solches gewährleisten. Als dessen Anbieter kann er Verantwortlicher oder Auftragsbearbeiter sein. Ist er Auftragsbearbeiter, so muss er die Rechtmässigkeit der Verwendung seines Modells bzw. seiner Anwendung und den generierten Output nicht von sich aus überprüfen, es sei denn, sein Geschäftsmodell sei auf eine unrechtmässige Verwendung ausgerichtet. Dies ist bei einem "General Purpose"-Service in der Regel nicht der Fall. Kann er Betroffenenrechte nicht selbst erfüllen (z.B. Entfernung von bestimmten Personendaten aus den vom Modell generierten Output), kann er diese Aufgabe an den Verwender des Modells delegieren.

Die unterschiedlichen Pflichten ergeben sich daraus, dass das Bereithalten eines Modells (mit Personendaten) eine eigenständige Datenbearbeitung ist. Dies bedeutet auch, dass betroffene Personen in Bezug auf das Modell gegenüber dessen Besitzer Rechte geltend machen können. Ob der Verwender des Modells im Service ein Verantwortlicher oder ein Auftragsbearbeiter ist, ergibt sich aus der Gestaltung des Service und der Kundenbeziehung. Im Falle von Services für Unternehmen wird der Besitzer i.d.R. Auftragsbearbeiter sein. Als solcher hat er die Weisungen des Verantwortlichen auszuführen und muss sie nicht auf ihre Rechtskonformität überprüfen. Er kann für eine Persönlichkeitsverletzung aufgrund seiner Mitwirkung in Anspruch genommen werden, haftet jedoch nur nach Massgabe seines Verschuldens. Trifft ihn keine Überprüfungspflicht, kommt ein Verschulden nur und erst in Frage, wenn er genügende Hinweise auf eine erkennbar unrechtmässige Verwendung im Einzelfall hat. Es kommen hier die in der Schweiz im Rahmen der Hosting-Providerhaftung entwickelten Grundsätze zur Anwendung: Die blosse Möglichkeit oder statische Gewissheit, dass es bei einigen Nutzern zu einer rechtswidrigen Verwendung kommen wird, genügt nicht, soweit er sein Geschäftsmodell nicht auf solche rechtswidrige Verwendungen ausgerichtet hat. Hat er das nicht, trifft ihn daher keine Pflicht zur Überprüfung der einzelnen konkreten Verwendungen seines Modells, d.h. die Sorgfaltspflicht verlangt von ihm nicht, dass er den von seinen Kunden erzeugten Output auf rechtswidrige Inhalte oder rechtswidrige Verwendung prüft.

Nutzung von Personendaten als Input

Personendaten sollten nur als Input für eine Anwendung generativer KI verwendet werden, wenn der betreffende Verantwortliche sich vergewissert hat, was mit diesen Personendaten geschieht und wenn ihm diese Bearbeitung (falls es sich um eine Auftragsbearbeitung handelt) bzw. Bekanntgabe von Personendaten (falls die Anwendung von einem anderen Verantwortlichen betrieben wird) erlaubt ist. Aus Sicht des Datenschutzrechts gelten hier keine anderen Anforderungen als bei einer Bearbeitung der Personendaten von einem anderen Computersystem oder Menschen.

Es gibt keinen Grund, aus Sicht des Datenschutzrechts andere oder höhere Anforderungen an die Nutzung von Personendaten als Input eines KI-Modells zu stellen, als sie für andere Datenbearbeitungen gelten.

Wenn überhaupt, liegen die systemischen Risiken solcher Anwendungen primär im von generativer KI erzeugten Inhalt (dazu hinten und in den Vorbemerkungen), d.h. dem Output, nicht im Input. Unsicherheiten bestehen in der Praxis primär in der Frage, ob der Anbieter einer KI-Anwendung die erforderliche Datensicherheit gewährleistet und ob er die ihm übermittelten Personendaten und generierten Outputs möglicherweise auch für eigene Zwecke verwendet (z.B. das Training des KI-Modells). Das kann ihm der Verwender der KI-Anwendungen aufgrund des Grundsatzes der Zweckbindung oder der konkreten Umstände möglicherweise nicht erlauben. Hier stellt sich jedoch nicht die Frage, ob und welche Regeln anwendbar sind, sondern ob und wie (d.h. mit welchen technischen und organisatorischen Massnahmen) sie im konkreten Fall eingehalten werden.

Nutzung des Outputs generativer KI

Von generativer KI erzeugter Output unterliegt hinsichtlich der darin enthaltenen Personendaten aus Sicht des Datenschutzrechts keinen anderen Anforderungen als von einem Menschen erzeugten Inhalte. Hingegen werden sich die Massnahmen zur Gewährleistung des Datenschutzrechts und der vom Verantwortlichen für eine Datenbearbeitung festgelegten Parameter unterscheiden, weil jede Bearbeitungsmethode eigene Besonderheiten (mit allfälligen Defiziten) aufweist, die zu einer Nonkonformität zum DSG führen können. Dies gilt für den menschlichen Geist wie für Maschinen.

Das DSG ist prinzipienbasiert und technologieneutral ausgestaltet und regelt Datenbearbeitungen, teilweise auch deren Automatisierung, nicht aber die Technik, mit der sie umgesetzt werden. Daher kann das DSG ohne Weiteres auch auf den Einsatz generativer KI angewandt werden. Zwar gibt es Unterschiede zwischen den verschiedenen Formen maschineller und menschlicher Bearbeitung von Personendaten. Sie betreffen jedoch naturgemäss nicht den Zweck und die weiteren datenschutzrechtlichen Parameter einer Datenbearbeitung (Datenkategorien, Empfänger, Dauer etc.), sondern deren konkrete Umsetzung. Ob eine Datenbearbeitung von Menschen oder Maschinen ausgeführt wird, hat daher primär Auswirkungen auf die Auswahl der Massnahmen, um die Konformität zum DSG möglichst sicherzustellen bzw. negative Auswirkungen zu minimieren. Nur wenn dies angesichts der Bearbeitungsform nicht hinreichend möglich ist, werden auch die Parameter der Datenbearbeitung angepasst werden müssen oder muss die Datenbearbeitung aufgegeben werden. Was als hinreichend gilt, ist dabei unabhängig davon zu beurteilen, ob Mensch oder Maschine zum Einsatz kommt; entscheidend ist das Ergebnis, da nur dieses die Persönlichkeit der betroffenen Person tangiert.

Keine Hinweispflicht betreffend Erzeugung durch KI

Es muss grundsätzlich nicht besonders darauf hingewiesen werden, dass bestimmte Inhalte im Falle von einer KI erzeugt worden sind, auch dann nicht, wenn sie Personendaten enthalten – soweit nicht die Bestimmungen über automatisierte Einzelentscheide greifen. Unabhängig von der Art und Weise der Erzeugung eines Inhalts kann jedoch z.B. eine Massnahme zur Richtigkeit sinnvoll sein, um den Inhalt in den richtigen Kontext zu stellen (z.B. dass ein Inhalt erfunden ist und nicht den Tatsachen entspricht).

Dies ergibt sich aus dem Grundsatz "Nutzung von Personendaten als Input". Es stellt sich zwar die Frage, ob KI-generierte Inhalte als solche gekennzeichnet werden müssen. Aus Sicht des Datenschutzes ist dies jedoch nicht der Fall, denn das DSG regelt nicht die technische Methode bzw. Massnahme im Einzelfall, mit welcher Personendaten bearbeitet werden und schränkt diese – mit Ausnahme im Falle von automatisierten Einzelentscheiden – nicht ein. Die Frage der Deklaration kann dort angebracht sein, wo die von einer KI erschaffenen Darstellung eines erfundenen Sachverhalts für das Publikum besonders überzeugend als wahr erscheinen kann (z.B. im Falle sog. Deep Fakes). Der datenschutzrechtliche Verstoss liegt hierbei aber nicht in der mangelnden Transparenz über die Erstellungsmethode, sondern in der Täuschung bzw. Irreführung des Publikums über den Wahrheitsgehalt der Information. Die Verbreitung von im Hinblick auf den jeweiligen Zweck falschen Personendaten ist dabei immer eine Verletzung des Bearbeitungsgrundsatzes der Richtigkeit, der – falls nicht gerechtfertigt – letztlich die Klarstellung verlangt, dass die Personendaten falsch sind (Kennzeichnung als "Fotomontage"), soweit sich das nicht aus den Umständen ergibt (z.B. im Falle einer Karikatur). Ob KI im Spiel war oder nicht ändert daran nichts.

Rechtfertigung für Training des Modells

Für das Training eines KI-Modells einschliesslich der erforderlichen Beschaffung von Personendaten kann grundsätzlich auf den Rechtfertigungsgrund der Bearbeitung für nicht personenbezogene Zwecke zurückgegriffen werden, falls überhaupt ein Rechtfertigungsgrund nötig ist. Dies gilt, falls die Bearbeitung nicht auf eine bestimmte Person ausgerichtet ist und die weiteren Voraussetzungen nach DSG erfüllt sind, also unter anderem die Verwendung des Modells nicht zu einer Veröffentlichung von Personendaten führt.

Dies ergibt sich aus dem Grundsatz nach "Nutzung von Personendaten als Input". Der Rechtfertigungsgrund der Bearbeitung für nicht personenbezogene Zwecke (Art. 31 Abs. 2 Bst. e DSG) ist technologieneutral und daher auch für Anwendungen im Bereich generativer KI zugänglich. Da das Training eines KI-Modells eine Datenbearbeitung darstellt (und von der späteren Verwendung des KI-Modells abgegrenzt werden muss und darf), stellt sich nebst der Prüfung der drei weiteren, in Art. 31 Abs. 2 Bst. e DSG aufgezählten Voraussetzungen nur die Frage, ob das Training eines KI-Modells für sich gesehen einen nicht personenbezogenen Zweck darstellt. Dies wird grundsätzlich der Fall sein, es sei denn, das Training des KI-Modells wäre auf bestimmte Personen ausgerichtet (z.B. um zu simulieren, wie eine bestimmte Person sich äussert, indem es nur mit Texten dieser Person trainiert wird oder wenn ein Anbieter einer App den Benutzern ermöglicht, diese auf ihre eigene Stimme zu trainieren, in welchem Fall der Anbieter als Verantwortlicher allerdings andere Rechtfertigungsgründe hat). Nicht personenbezogen ist hingegen das Training eines Modells z.B. dann, wenn es später in der Lage ist, eine Vielzahl von Personen zu imitieren, weil es mit entsprechenden Inputs instruiert wird. Diesfalls ist die Verwendung des KI-Modells personenbezogen, nicht dessen initiales Training.

Generierung neuer Personendaten als Beschaffung

Benutzt ein Verantwortlicher ein fremdes KI-Modell, um es über Informationen zu einer bestimmten Person zu befragen, so stellt dies eine Beschaffung von Personendaten dar. Dies kann eine entsprechende Informationspflicht nach Art. 19 DSG auslösen, und es sind die Bearbeitungsgrundsätze, wie namentlich die Zweckbindung, die Richtigkeit und die Datenminimierung zu beachten (soweit kein Rechtfertigungsgrund bzw. ein Ausnahmetatbestand vorliegt, der davon dispensiert). Keine Beschaffung liegt hingegen vor, wenn ein KI-Modell benutzt wird, um bestehende Personendaten gemäss Instruktionen des Verantwortlichen umzuformen oder (mit Sachdaten oder bereits bestehenden Personendaten) zu ergänzen, oder wenn sich im Output eines KI-Modells Personendaten befinden, die nicht beabsichtigt waren (z.B. zufällige Erwähnung einer Person). Auch hier sind die Bearbeitungsgrundsätze zu berücksichtigen, aber es entsteht keine neue Informationspflicht. Nur muss die bisherige Informationspflicht auch diese Bearbeitung abdecken, soweit sie zum Zeitpunkt der Beschaffung geplant war.

Als Beschaffung (namentlich im Sinne von Art. 19 DSG) gilt nur die planmässige Erhebung von Personendaten. Personendaten können nicht nur von der betroffenen Person selbst beschafft werden, sondern auch aus anderen Quellen. Das können auch interne Quellen sein, wenn bereits vorhandene Daten umgenutzt werden. Beim Einsatz von KI-Modellen ist entscheidend, ob diese nur dazu genutzt werden, bestehende Personendaten neu oder anders zu gestalten oder ob sie als eigentliche, allenfalls neue Informationsquellen für nicht vorhandene Personendaten genutzt werden.
Im letzteren Falle stellen sie eine Quelle und nicht nur ein Bearbeitungswerkzeug dar. Werden sie als solche planmässig eingesetzt, kann daher eine Beschaffung vorliegen, was namentlich eine neue Informationspflicht auslösen kann, d.h. eine Ergänzung der Datenschutzerklärung nötig macht und geprüft werden muss, ob und inwiefern die Erklärung den betroffenen Personen mitgeteilt werden muss (falls die bestehende Erklärung die Quelle z.B. nicht schon abdeckt). Werden Personendaten lediglich umgeformt (z.B. wenn ein KI-Modell benutzt wird, um einen passenden Antwortbrief an eine Person zu formulieren), stellt dies aber keine Beschaffung von Personendaten dar. Für die Datenschutzerklärung ist dieser Vorgang nur insofern relevant, als die angegebenen Bearbeitungszwecke und weiteren Pflichtinhalte passen müssen. Die Bearbeitungsgrundsätze sind vorbehaltlich etwaiger Rechtfertigungsgründe so oder so einzuhalten.

Keine Beschaffung durch blosse Servicenutzung

Die Nutzung eines Service, der Zugang zu einem KI-Modell bietet, stellt für sich genommen keine Datenbeschaffung dar, selbst wenn das KI-Modell Personendaten enthält und danach abgefragt werden könnte. Eine solche liegt nur dann vor, wenn der Verwender des Services über diesen Service an Personendaten gelangt, über die er bisher nicht verfügt hat und die auch nicht nach seinen Vorgaben erzeugt wurden.

Es müssen hier dieselben Regeln wie für jede Informationsquelle gelten, die von einem Anbieter in eigener Verantwortlichkeit betrieben wird. Der Umstand, dass ein Verantwortlicher Zugang zur Datenbank eines Dritten mit Personendaten hat, bedeutet nicht, dass er über alle darin enthaltenen Personendaten verfügt. Das ist auch dann so, wenn die konkrete Abfrage eines KI-Modells (analog die Datenbanksuche) als Auftragsbearbeitung ausgestaltet ist. Auch wenn dieser das gesamte KI-Modell zur Erfüllung der Aufgabe (Formulierung des Outputs) zur Verfügung steht, hat der Verwender des KI-Modells damit nicht automatisch dessen gesamten Inhalt erhoben, sondern nur den Output. Nur für diesen ist er verantwortlich.

Output-Generierung als Bekanntgabe

Ermöglicht ein Verantwortlicher Dritten, mit eigenem Input von einem KI-Modell erzeugten Output zu generieren und enthält dieser Output dem Dritten noch nicht bekannte Personendaten, so stellt dies in der Regel eine Bekanntgabe von Personendaten an den Dritten dar. Im Falle besonders schützenswerter Personendaten ist hierfür eine Rechtfertigung erforderlich. Der Abschluss einer Vereinbarung ist hingegen keine genügende Rechtfertigung. Keine Bekanntgabe von Personendaten liegt hingegen vor, wenn vom Dritten erhaltene Personendaten gemäss den Vorgaben des Dritten lediglich umgeformt oder (mit Sachdaten) ergänzt werden.

Wird ein KI-Modell benutzt, um Inhalte zu formulieren (z.B. einen Brief) und sind diese personenbezogen, so liegen Personendaten vor.

Gab es diese bisher nicht und gilt der Betreiber des KI-Modells als Verantwortlicher, so könnte die Bereitstellung des Outputs als Bekanntgabe von Personendaten qualifiziert werden, weil diese für ihn neu sind. Dies ist jedoch dann nicht passend, wenn die Inhalte nach der Instruktion des Dritten formuliert worden sind. In diesem Fall war er es, von welchem der Sinngehalt der Personendaten stammt – das KI-Modell hat ihn lediglich formuliert. In solchen Fällen kann daher keine Bekanntgabe von Personendaten vorliegen, weil nur bekanntgegeben werden kann, was nicht schon bekannt ist. Doch selbst wenn eine Bekanntgabe von Personendaten vorliegt, ist diesbezüglich normalerweise keine Vereinbarung erforderlich, jedenfalls wenn der Empfänger ebenfalls einem angemessenen Datenschutz unterliegt. Eine Pflicht zum Abschluss einer Vereinbarung besteht nur bei Auftragsbearbeitern.

Richtigkeit des Outputs

Die Richtigkeit des Outputs muss wie immer im Hinblick auf den Verwendungszweck vorab definiert und beurteilt werden. Dieser hängt von den Umständen ab und kann als mögliche Massnahme durch entsprechende Hinweise gesetzt werden. Da heute notorisch ist, dass der mit Hilfe von Modellen generativer KI erzeugte Output keinen Anspruch auf Richtigkeit haben kann, genügt ein angemessener Hinweis darauf, dass ein konkreter Output von einem solchen Modell erzeugt worden ist, um dem Grundsatz der Richtigkeit zu genügen. Dies gilt jedenfalls, solange keine Äusserungen gemacht werden, die die Richtigkeit der im Output enthaltenen Personendaten suggerieren.

Liefert eine KI falsche Personendaten als Output, stellt dies nur dann eine Verletzung des Grundsatzes der Datenrichtigkeit dar, wenn die Personendaten im Hinblick auf den verfolgten Zweck unrichtig sind.

Erfordert der Zweck keine richtigen Daten bzw. ist der Zweck dergestalt, dass der Adressat des Outputs nicht objektiv richtige Inhalte erwartet, so stellt unrichtiger Output keine Verletzung des Grundsatzes der Datenrichtigkeit dar. Dies ist nicht neu: Wer eine Internet-Suchmaschine einsetzt, erwartet auch nicht, dass die gefundenen Informationen richtig sind. Dies muss der Benutzer beurteilen, und jeder Benutzer einer Internet-Suchmaschine weiss dies. Dieselbe Notorietät hat aufgrund der intensiven Presseberichterstattung inzwischen auch die eingeschränkte Zuverlässigkeit der Large Language Models diverser Anbieter und Pendants zur Erzeugung von Bildern. Wer einen KI-Chat-Roboter benutzt, weiss, dass seine Antworten nicht unbedingt richtig sind – selbst wenn sie sehr selbstsicher formuliert sind. Dies ist auch im datenschutzrechtlichen Kontext zu berücksichtigen.

Umsetzung von Widersprüchen

Die Rechte von Betroffenen müssen auch beim Einsatz generativer KI im gesetzlichen Umfang vom Verantwortlichen gewährleistet bleiben, selbst wenn ihre Umsetzung technisch schwierig sein mag. Allerdings gelten diese Rechte nie absolut, und Eigenschaften einer Technik, die bereits eine "normale" Bearbeitung von Daten einschränken, können auch bei der Beschränkung von Betroffenenrechten berücksichtigt werden, etwa im Rahmen der gesetzlich vorgesehenen Interessenabwägung. Eine solche ist auch im Bereich generativer KI möglich: Wird auf die Widerspruchserklärung einer Person nur der sie betreffende Output z.B. geschwärzt und die Verwendung des KI-Modells in Bezug auf Personendaten der Person hinreichend kontrolliert, müssen etwaige im Modell vorhandene Personendaten nicht auch gelöscht werden; der damit verbundene Eingriff wird in der Regel nicht gerechtfertigt sein, da die Massnahme die legitimen Interessen der betroffenen Person bereits hinlänglich wahrt.

Dies leitet sich wiederum daraus ab, dass das Datenschutzrecht technologieneutral ausgestaltet ist – auch in Bezug auf die Rechte der Betroffenen. Auch im Bereich KI gelten somit nur aber immerhin die im DSG vorgesehenen Anspruchsgrundlagen und Ausnahmebestimmungen.

Das bedeutet, dass Schwierigkeiten bei der Umsetzung von Betroffenenrechten mittels technischer und organisatorischer Massnahmen (z.B. gezielte Extraktion von Personendaten aus einem Modell) im Rahmen einer etwaigen Interessenabwägung berücksichtigt werden können, soweit sie den vorliegend relevanten technischen Verfahren immanent sind bzw. mit denen selbst jene zu "kämpfen" hätten, die vorausschauend versuchen würden Wege zur Erfüllung der relevanten Betroffenenrechte zu finden. Von einem Verantwortlichen kann dabei nur der Einsatz des bewährten Stands der Technik erwartet werden (und z.B. nicht von Verfahren, die sich erst im Stadium der Erforschung befinden).

Umsetzung des Auskunftsrechts

Soll Auskunft über Personendaten in einem KI-Modell erteilt werden, ist die Anfrage an diejenigen zu richten, die eine Kopie des KI-Modells besitzen. Wer es lediglich als Service für seine Zwecke verwendet (d.h. es mit Input zur Erzeugung von Output bespielt, darüber aber selbst nicht verfügt), ist in Bezug auf das Modell nicht auskunftspflichtig, sondern nur in Bezug auf seinen eigenen Input und Output, falls er über diesen überhaupt noch verfügt (ihn für die Zwecke des Auskunftsrechts aufbewahren muss er ihn nicht).

Auskunftspflichtig ist nur der Verantwortliche der betreffenden Datenbearbeitung. Das Speichern bzw. Bereithalten des KI-Modells ist eine von dessen Verwendung unterschiedliche Datenbearbeitung (vgl. Grundsatz "Unterschiedliche Verantwortlichkeiten für Erstellung, Besitz und Verwendung"). Installiert ein Verwender eines KI-Modell eine Kopie davon auf seiner eigenen Infrastruktur, wird er allerdings bezüglich des Modells auskunftspflichtig (und "Besitzer" in der vorliegenden Terminologie).

Notwendigkeit einer Datenschutz-Folgenabschätzung

Soll eine Anwendung basierend auf einem Modell für generative KI zum Einsatz kommen, so beurteilt sich die Notwendigkeit einer Datenschutz-Folgenabschätzung (DSFA) nach dem Risiko, das damit voraussichtlich verbunden ist. Der Umstand, dass im KI-Modell Personendaten vorkommen, für sein Training Personendaten zum Einsatz gekommen sind oder der Output Personendaten enthalten können, stellt für sich noch kein hohes Risiko dar. Mit anderen Worten zieht der Einsatz eines KI-Modells nicht per se eine DSFA nach sich.

Art. 22 Abs. 2 DSG erwähnt zwar die Verwendung neuer Technologien in Bezug auf die Beurteilung der Möglichkeit eines hohen Risikos einer Datenbearbeitung für die betroffenen Personen, erwähnt diese allerdings nur als zu berücksichtigender Faktor. Entscheidend sind nach wie vor die Art, der Umfang, die Umstände und der Zweck der Bearbeitung. Setzt ein Unternehmen ein KI-Modell zur Vorformulierung von Werbeanschreiben für seine Kunden ein, ist beispielsweise nicht von einem hohen Risiko für die betroffenen Personen auszugehen, weshalb keine Datenschutz-Folgenabschätzung erforderlich sein wird. Dies kann z.B. im Rahmen einer Schwellenwertanalyse ermittelt werden, wenn der Fall nicht klar ist.

Verletzung der Datensicherheit

Führt die Verwendung eines KI-Modells zu einer unrechtmässigen Preisgabe von Personendaten, so stellt dies nur dann eine Verletzung der Datensicherheit im Sinne von Art. 5 Bst. h DSG dar, wenn dies unplanmässig erfolgte, namentlich weil Massnahmen zur Kontrolle oder Beschränkung des Outputs unvorhergesehen versagten oder ausser Kraft gesetzt wurden, und so Personendaten von unbefugten Dritten wahrgenommen oder bearbeitet werden konnten. Es muss daher immer zuerst der operationelle Grund des Vorfalls ermittelt werden.

Nicht jede Verletzung des Datenschutzes ist eine "Verletzung der Datensicherheit" im Sinne des DSG. Letztere erfordert erstens eine Verletzung der Sicherheit, d.h. eine aus Sicht des Verantwortlichen planwidrige Verletzung der Vertraulichkeit, Integrität oder Verfügbarkeit von Personendaten, und diese Daten zweitens – vorliegend – darum gegenüber Unbefugten offengelegt oder zugänglich gemacht werden (wobei es genügt, dass eine mehr als theoretische Möglichkeit besteht, dass sie die Daten wahrnehmen oder bearbeiten können). Fehlt eine der beiden Voraussetzungen, muss die Frage einer Meldepflicht bzw. der Information der betroffenen Person nicht geprüft werden. Ist ein KI-Modell im Training mit Personendaten trainiert bzw. "gefüttert" worden, die es nicht enthalten dürfte, stellt deren Preisgabe für sich keine Verletzung der Sicherheit dar, sondern ist eine Verletzung des Datenschutzes, z.B. der Verhältnismässigkeit oder Zweckbindung. Ein solcher Input ist zudem eine von der Verwendung unabhängige Datenbearbeitung. Auch wo ein Filter für Personendaten schlicht an seine Grenzen stösst oder ein erzeugtes Bild zufällig dem einer realen Person entspricht, liegt keine Verletzung der Datensicherheit vor (aber möglicherweise des Datenschutzes). Sorgt ein Verwender des KI-Modells jedoch dafür, dass ein Kontrollmechanismus deaktiviert oder umgangen wird, oder geschieht dies aufgrund einer Fehlfunktion, ist die für die Datenbearbeitung geschaffene Sicherheit tangiert.

Vertrauen auf Vertragstreue

Sichert ein Anbieter einer KI-Anwendung eine DSG-konforme Bearbeitung des Inputs und Outputs vertraglich zu (z.B. im Rahmen eines Auftragsbearbeitungs-vertrags unter Ausschluss der Eigen-verwendung der Daten) und gibt es keinen Grund zur Annahme, dass er sich nicht an den Vertrag (und die darin vereinbarten technischen und organi-satorischen Massnahmen) hält, darf auf die Einhaltung des Vertrags vertraut werden.

Diese Regel gilt im Datenschutzrecht generell. Sie ist Grundstein jeder arbeitsteiligen Wirtschaft. Zwar ist die Einhaltung von in Verträgen vereinbarten technischen und organisatorischen Mass-nahmen in angemessener Weise (d.h. risikobasiert) zu überprüfen, aber der Gesetzgeber hat den Vertrag als Instrument zur Absicherung des Daten-schutzes grundsätzlich anerkannt (vgl. z.B. Art. 16 Abs. 2 Bst. b, d und e DSG).

Verein Unternehmens-Datenschutz VUD
www.vud.ch | info@vud.ch

Der Verein Unternehmens-Datenschutz VUD ist ein Zusammenschluss von Schweizer Unternehmen im betrieblichen Datenschutz. Er ist der selbständigen und unabhängigen Meinungsbildung im Bereich Datenschutz verpflichtet.

Autor: David Rosenthal