Universität Zürich[UZH]

Universität Basel

Juristische Fakultät

WHITE PAPER

Datenschutz

Florent Thouvenin
Stephanie Volz

Juni 2024

CENTER FOR INFORMATION TECHNOLOGY SOCIETY AND LAW — ITSL

e-PIAF
electronic Public Institutions and Administrations Research Forum

Dieses White Paper wurde im Projekt **«Nachvollziehbare Algorithmen: ein Rechtsrahmen für den Einsatz von Künstlicher Intelligenz»** entwickelt, das vom Center for Information Technology, Society, and Law (ITSL) der Universität Zürich und von der Forschungsstelle electronic Public Institutions and Administrations Research Forum (e-PIAF) der Universität Basel durchgeführt und von der Stiftung Mercator finanziell unterstützt wird. Dieses White Paper ist Teil einer Reihe von White Papers, die sich mit den zentralen Herausforderungen befassen, die mit dem Einsatz von Künstlicher Intelligenz (KI) in Unternehmen und in der Verwaltung verbunden sind.

Die White Papers und weitere Informationen zum Projekt finden sich auf **www.itsl.uzh.ch** und **www.ius.unibas.ch/e-piaf**.

Folgende White Papers sind bislang erschienen:

- **Manipulation**
- **Diskriminierung**
- **Datenschutz**
- **Transparenz**
- **Transparenz durch Begründung von Verfügungen**
- **Transparenz durch öffentliches Verzeichnis**

Das Kernprojektteam besteht aus folgenden Personen:

Prof. Dr. Florent Thouvenin, Professor für Informations- und Kommunikationsrecht an der Universität Zürich, Vorsitzender des Lenkungsausschusses des ITSL

Prof. Dr. Nadja Braun Binder, MBA, Professorin für öffentliches Recht an der Universität Basel

Dr. Stephanie Volz, wissenschaftliche Geschäftsführerin des ITSL und Lehrbeauftragte an der Universität Zürich

Liliane Obrecht, MLaw, Wissenschaftliche Mitarbeiterin und Doktorandin an der Juristischen Fakultät der Universität Basel

Bei der Entwicklung und Verwendung von Systemen der Künstlichen Intelligenz (KI) werden sehr grosse Datenmengen verwendet. In vielen Fällen handelt es sich dabei um Personendaten. Die Bearbeitung dieser Daten wird durch das Datenschutzrecht umfassend normiert. Damit stellt sich die Frage, wie die Bestimmungen des Datenschutzgesetzes (DSG) angewendet werden können, um die Herausforderungen bei der Bearbeitung von Personendaten durch KI zu erfassen, und ob sie zu diesem Zweck ergänzt oder präzisiert werden müssen.

Ziel dieses White Papers ist nicht eine umfassende Darstellung aller datenschutzrechtlicher Fragen, die sich bei der Entwicklung und Verwendung von KI-Systemen stellen können. Die Analyse fokussiert vielmehr auf diejenigen Aspekte, bei denen die Anwendung von KI besondere Fragen aufwirft. Sie zeigt, dass die Herausforderungen durch eine sinnvolle Auslegung und Anwendung des geltenden Rechts gelöst werden können. Namentlich wird klar, dass Personendaten nach dem DSG für das Trainieren, Validieren und Testen von KI-Systemen bearbeitet werden dürfen, weil es sich um eine Bearbeitung für nicht personenbezogene Zwecke handelt. Um Rechtssicherheit zu schaffen, könnte der Gesetzgeber diesen Fall aber in den gesetzlichen Katalog der Regelbeispiele für eine Rechtfertigung aus überwiegendem Interesse aufnehmen.

Problemstellung

Die Entwicklung von Systemen der Künstlichen Intelligenz (KI) beruht auf der Bearbeitung sehr grosser Datenmengen. In vielen Fällen handelt es sich dabei um **Personendaten**. Auch bei der Verwendung von KI-Systemen werden regelmässig Personendaten bearbeitet, bspw. bei Systemen zur Analyse von MRI-Scans oder für die Selektion von Bewerbungen. Soweit bei der Entwicklung und Verwendung von KI-Systemen Personendaten bearbeitet werden, müssen die Vorgaben des Datenschutzrechts eingehalten werden.

Die Normen des DSG sind **technologieneutral**. Sie erfassen damit jedes Bearbeiten von Personendaten, unabhängig von der verwendeten Technologie, und normieren das Bearbeiten von Personendaten durch KI-Systeme umfassend. Damit ist bereits gesagt, dass es keinen spezifischen Erlass zur Regelung des Bearbeitens von Personendaten durch KI-Systeme braucht. Erforderlich ist nur eine **angemessene Anwendung der bestehenden Bestimmungen des DSG** (und der sektorspezifischen datenschutzrechtlichen Bestimmungen) auf KI-Systeme. Das entspricht auch dem Ansatz der KI-Konvention des Europarates, nach der die Mitgliedstaaten Massnahmen ergreifen oder aufrechterhalten müssen, um das Recht auf Privatsphäre (privacy rights) der natürlichen Personen zu schützen und die bestehenden nationalen und internationalen Vorgaben des Datenschutzrechts einzuhalten (Art. 11 Bst. a KI-Konvention). Neben der Anwendung der Bestimmungen des DSG ist in der Folge auch zu prüfen, ob eine punktuelle Anpassung oder Ergänzung der bestehenden Regelung angezeigt ist, um KI-spezifische Fragen angemessen zu erfassen.

Da sich bei der Entwicklung, beim Anbieten und bei der Verwendung von KI-Systemen unterschiedliche Fragen stellen, werden diese drei Situationen je für sich betrachtet.

Entwicklung

Grundsätze der Datenbearbeitung

Das Beschaffen und Bearbeiten von Personendaten für das Trainieren, Validieren und Testen von KI-Systemen steht zu den **Grundsätzen der Datenbearbeitung** (Art. 6 DSG) in einem Spannungsverhältnis. Das Beschaffen und Bearbeiten ist für die betroffenen Personen in der Regel nicht erkennbar, weil für das Trainieren, Validieren und Testen von KI-Systemen regelmässig Daten aus verschiedenen Quellen verwendet werden, ohne dass die betroffenen Personen darüber informiert würden (**Grundsatz der Erkennbarkeit**). Dabei werden die Daten zu einem anderen Zweck bearbeitet als zu dem, für den sie ursprünglich beschafft worden sind (**Grundsatz der Zweckbindung**). Hier stellt sich die Frage, ob das Bearbeiten von Personendaten für das Trainieren, Validieren und Testen eines KI-Systems als Zweck zu verstehen ist, der mit dem ursprünglichen Zweck der Bearbeitung nicht vereinbar ist. Das dürfte regelmässig der Fall sein, wenn das KI-System einem ganz anderen Zweck dient als der Zweck, zu dem die Daten beschafft worden sind, bspw. wenn Fotos auf Social Media verwendet werden, um eine Gesichtserkennungssoftware zu trainieren. Dasselbe gilt, wenn die Daten für das Trainieren eines KI-Systems verwendet werden, das zu beliebigen Zwecken verwendet werden kann (sog. «foundation model» oder «general purpose AI system»). Anderes gilt bspw., wenn Röntgenbilder von Knochenbrüchen verwendet werden, um ein KI-System zu trainieren, das Knochenbrüche auf Röntgenbildern erkennen kann.

Beim Trainieren, Validieren und Testen von KI-Systemen führt die Verwendung von mehr Daten in der Regel zu besseren Ergebnissen. Dem Bedürfnis nach der Bearbeitung möglichst vieler Daten steht der **Grundsatz der Datenminimierung** an sich diametral entgegen. Dieser kann aber durchaus so ausgelegt und angewendet werden, dass ein Verstoss entfällt, weil die Verwendung möglichst vieler Daten dem Zweck des Trainierens, Validierens und Testens von KI-Systemen angemessen und für diesen Zweck notwendig ist. Dasselbe gilt für den **Grundsatz der Speicherbegrenzung**. Die für das Trainieren, Validieren und Testen von KI-Systemen verwendeten Personendaten sind zwar nicht als solche im trainierten KI-System gespeichert, sie werden aber für diesen Zweck und bisweilen auch nach Abschluss der Entwicklung in separaten Datenbanken gespeichert. Eine Speicherung nach Abschluss der Entwicklung kann aus technischen Gründen (bspw. spätere Weiterentwicklung oder Überprüfung des KI-Systems) sinnvoll oder aus rechtlichen Gründen (z. B. Vorgaben in der KI-Verordnung der EU, in Codes of Practices oder in technischen Standards) erforderlich sein. Die Speicherung der Daten ist damit für den Zweck der Bearbeitung erforderlich und nicht als Verstoss gegen den Grundsatz der Speicherbegrenzung zu qualifizieren.

Rechtfertigung

Die Analyse zeigt, dass die Bearbeitung von Personendaten für das Trainieren, Validieren und Testen von KI-Systemen regelmässig gegen einen oder mehrere Grundsätze der Datenbearbeitung verstösst. Damit stellt sich die Frage, ob die Bearbeitung durch Einwilligung der betroffenen Personen, durch ein überwiegendes privates oder öffentliches Interesse oder durch Gesetz gerechtfertigt werden kann (Art. 31 Abs. 1 DSG). Beim Trainieren, Validieren und Testen von KI-Systemen steht die Rechtfertigung durch ein überwiegendes Interesse des Verantwortlichen im Vordergrund. Das Speichern der Daten kann zudem durch Gesetz gerechtfertigt sein, wenn gesetzliche Dokumentationspflichten bestehen, die eine Pflicht zur Speicherung der Daten vorsehen. Das Einholen einer Einwilligung wird hingegen wegen der grossen Zahl der betroffenen Personen und der sehr grossen Menge an Daten sowie der verwendeten Quellen (bspw. Webseiten, Social Media Posts) meist nicht möglich sein.

Das DSG nennt eine Reihe von Konstellationen, in denen ein überwiegendes Interesse des Verantwortlichen regelmässig in Betracht kommt (Art. 31 Abs. 2 DSG), unter anderem die Rechtfertigung von **Bearbeitungen für nicht personenbezogene Zwecke**, insb. für Forschung, Planung oder Statistik (Art. 31 Abs. 2 lit. e DSG). Beim Trainieren, Validieren und Testen von KI-Systemen werden Personendaten nicht bearbeitet, um Erkenntnisse über die betroffenen Personen zu gewinnen oder Folgen für diese Personen abzuleiten, sondern um ein System zu entwickeln, das eine bestimmte Aufgabe (oder eine Vielzahl unbestimmter Aufgaben) möglichst gut erfüllen kann. Die Bearbeitung ist damit nicht personenbezogen und lässt sich grundsätzlich durch das überwiegende Interesse des Verantwortlichen rechtfertigten.

Die Rechtfertigung einer Bearbeitung für nicht personenbezogene Zwecke setzt allerdings voraus, dass **drei weitere Voraussetzungen** erfüllt sind: (i) Der Verantwortliche muss die **Daten anonymisieren**, sobald es der Bearbeitungszweck erlaubt; ist eine Anonymisierung unmöglich oder erfordert

sie einen unverhältnismässigen Aufwand, reicht es, wenn der Verantwortliche angemessene Massnahmen trifft, um die Bestimmbarkeit der betroffenen Personen zu verhindern (Art. 31 Abs. 2 lit. e Ziff. 1 DSG). Diese Voraussetzung lässt sich bei der Bearbeitung von Personendaten für das Trainieren, Validieren und Testen von KI-Systemen in aller Regel durch geeignete technische Massnahmen erfüllen. (ii) Werden **besonders schützenswerte Personendaten** bearbeitet, dürfen diese Dritten nur so bekanntgegeben werden, dass die betroffenen Personen nicht bestimmbar sind; ist dies nicht möglich, muss gewährleistet sein, dass die Dritten die Daten nur zu nicht personenbezogenen Zwecken bearbeiten (Art. 31 Abs. 2 lit. e Ziff. 2 DSG). Auch diese Voraussetzung wird sich erfüllen lassen. (iii) Schliesslich dürfen die Ergebnisse **nur so veröffentlicht werden, dass die betroffenen Personen nicht bestimmbar** sind (Art. 31 Abs. 2 lit. e Ziff. 3 DSG). Diese Voraussetzung zielt auf die Regelbeispiele der Bearbeitung für Forschung, Planung oder Statistik; für diese Konstellationen ist sie wichtig und sinnvoll und in aller Regel auch ohne weiteres zu erfüllen. Für das Trainieren, Validieren und Testen von KI-Systemen passt die Voraussetzung aber kaum. Ob sie erfüllt ist, hängt wesentlich davon ab, ob man davon ausgeht, dass mit der Inbetriebnahme eines KI-Systems auch die Trainingsdaten «veröffentlicht» werden, wie das teilweise vertreten wird. Diese Auffassung vermag allerdings nicht zu überzeugen, weil sich (gewisse) Trainingsdaten zwar mit hinreichenden technischen Kenntnissen und mit einigem Aufwand aus dem trainierten KI-System ermitteln («reverse engineering») lassen, die Personendaten mit dem KI-System aber nicht im Sinn dieser Bestimmung veröffentlicht werden. Das gilt allerdings nur, wenn KI-Systeme so ausgestaltet werden, dass sich die beim Trainieren verwendeten Personendaten daraus möglichst gar nicht oder jedenfalls nur mit einem derart grossen Aufwand ermitteln lassen, dass ein solches Ermitteln zwar theoretisch möglich ist, nach der allgemeinen Lebenserfahrung aber nicht damit zu rechnen ist, dass ein Dritter diesen Aufwand auf sich nehmen wird.

Selbst wenn gewisse Zweifel bestehen sollten, ob die drei Voraussetzungen des spezifischen Rechtfertigungsgrundes der Bearbeitung für nicht personenbezogene Zwecke stets erfüllt sind, ist eine **Rechtfertigung** der Bearbeitung von Personendaten beim Trainieren, Validieren und Testen von KI-Systemen **ohne weiteres möglich**. Denn der spezifische Rechtfertigungsgrund ist nur ein Regelbeispiel für das Vorliegen eines **überwiegenden Interesses** des Verantwortlichen. Ein solches Interesse kann auch dann bejaht werden, wenn die Voraussetzungen des Regelbeispiels nicht erfüllt sind. Da die Bearbeitung von Personendaten beim Trainieren, Validieren und Testen von KI-Systemen für die betroffenen Personen keinerlei Konsequenzen hat, dürfte das Interesse des Verantwortlichen an der Bearbeitung das Interesse der betroffenen Personen an der Nicht-Bearbeitung in aller Regel überwiegen.

Angesichts der grossen Bedeutung von KI-Systemen könnte es allerdings sinnvoll sein, dass der **Gesetzgeber die Liste der Regelbeispiele** für das Vorliegen eines überwiegenden Interesses des Verantwortlichen (Art. 31 Abs. 2 DSG) um ein weiteres Beispiel **ergänzt**, das die Bearbeitung von Personendaten für das Trainieren, Validieren und Testen von KI-Systemen erfasst. Alternativ könnte es auch genügen, wenn der **Eidgenössische Datenschutz- und Öffentlichkeitsbeauftragte (EDÖB)** in einem **Leitfaden oder Merkblatt** klarstellen würde, dass die Inbetriebnahme eines KI-Systems nicht als Veröffentlichung von Ergebnissen im Sinn der dritten Voraussetzung des Rechtfertigungsgrundes (Art. 31 Abs. 2 lit. e Ziff. 3 DSG) zu verstehen ist.

Rechte und Pflichten

Das DSG sieht eine Reihe von Pflichten der Verantwortlichen und mehrere Rechte der betroffenen Personen vor. Bei der Entwicklung von KI-Systemen stehen die Informationspflicht und das Auskunftsrecht im Vordergrund. Zudem fragt sich, ob die betroffenen Personen ein Recht auf Berichtigung und einen Anspruch auf Löschung der Personendaten oder ein Recht auf Widerspruch gegen das Bearbeiten haben.

Für das Bearbeiten von Personendaten bei der Entwicklung von KI-Systemen ist verantwortlich, wer allein oder zusammen mit anderen über den Zweck und die Mittel der Bearbeitung entscheidet (Art. 5 lit. j DSG). **Verantwortlicher** ist damit in aller Regel das Unternehmen, welches das KI-System trainiert, validiert und testet und über die dafür erforderlichen Daten verfügt, also der Anbieter («provider»). Dasselbe gilt für die Weiterentwicklung eines KI-Systems; Verantwortlicher ist auch hier das Unternehmen, welches die Weiterentwicklung vornimmt. Wird ein KI-System von einem Anbieter individuell für einen Betreiber («deployer») (weiter-)entwickelt oder (weiter-)trainiert, werden Anbieter und Betreiber für die Bearbeitung der Trainingsdaten als gemeinsam Verantwortliche zu qualifizieren sein. Wird ein vortrainiertes KI-System von einem

Betreiber für seine eigenen Zwecke mit weiteren Daten weiter trainiert, wird der Betreiber in der Regel für diese Bearbeitung allein verantwortlich sein.

Der Verantwortliche ist verpflichtet, die betroffenen Personen über die Beschaffung und Bearbeitung der sie betreffenden Personendaten zu informieren. Das gilt auch, wenn die Daten nicht bei der betroffenen Person, sondern bei einem Dritten beschafft werden (Art. 19 Abs. 1 DSG). Die **Informationspflicht** entfällt, wenn die Information nicht möglich ist oder einen unverhältnismässigen Aufwand erfordert (Art. 20 Abs. 2 DSG). Beschafft der Verantwortliche die Daten direkt bei den betroffenen Personen, wird er diese in aller Regel über die Bearbeitung informieren können und damit auch müssen. Der Verantwortlich wird die Daten, die er für das Trainieren, Validieren und Testen von KI-Systemen verwendet, allerdings meist bei Dritten beschaffen. Eine direkte Information der Betroffenen wird deshalb in aller Regel nicht möglich oder mit einem unverhältnismässigen Aufwand verbunden sein. In diesem Fall wird man aber verlangen, dass der Verantwortliche die Öffentlichkeit in allgemeiner Weise über die Bearbeitung der Personendaten informiert, bspw. in Form einer Mitteilung auf seiner Webseite.

Betroffene Personen haben das Recht, vom Verantwortlichen Auskunft darüber zu verlangen, ob Personendaten über sie bearbeitet werden (Art. 25 Abs. 1 DSG). Dieses **Auskunftsrecht** besteht auch beim Bearbeiten von Personendaten für die Entwicklung eines KI-Systems. Da die für das Trainieren, Validieren und Testen verwendeten Personendaten in der Regel in einer Datenbank gespeichert sind, wird es für den Anbieter durchaus möglich sein, einem Auskunftsbegehren zu genügen. Fraglich ist allerdings, ob die betroffenen Personen auch vom Betreiber eines KI-Systems Auskunft darüber verlangen können, ob das System mit sie betreffenden Personendaten trainiert, validiert oder getestet worden ist. Das ist schon deshalb zu verneinen, weil der Betreiber für das Bearbeiten von Personendaten bei der Entwicklung des KI-Systems nicht als Verantwortlicher zu qualifizieren ist. Ein Anspruch auf Auskunft gegenüber dem Betreiber könnte zudem nur bejaht (und erfüllt) werden, wenn die bei der Entwicklung verwendeten Personendaten als solche im KI-System gespeichert wären und damit vom Betreiber bearbeitet würden. Das ist aber nicht der Fall.

Beim Recht auf **Berichtigung**, beim Anspruch auf **Löschung** und beim Erheben von **Widerspruch** gegen die Bearbeitung von Personendaten ist zwischen der Entwicklung und dem entwickelten KI-System zu unterscheiden. Die betroffenen Personen können ihre Rechte und Ansprüche gegenüber dem Bearbeiten der sie betreffenden Personendaten beim Trainieren, Validieren und Testen von KI-Systemen geltend machen, wenn die Voraussetzungen des jeweiligen Rechts bzw. Anspruchs erfüllt sind. Ist das der Fall, dürfen die Daten nicht für das Trainieren, Validieren und Testen verwendet werden und sie müssen gelöscht oder berichtigt werden. Mit Bezug auf das fertig entwickelte Modell können die Betroffenen diese Rechte bzw. Ansprüche aber nicht geltend machen, weil dieses die Personendaten nicht als solche enthält. Anderes gilt nur, wenn ein KI-System nicht hinreichend gegen Angriffe geschützt ist (siehe dazu sogleich).

Anbieten

Das Anbieten von KI-Systemen auf dem Markt ist datenschutzrechtlich relevant, wenn dabei Personendaten übermittelt oder zugänglich gemacht und damit bekannt gegeben werden (Art. 5 lit. e DSG). Ein Bekanntgeben liegt aber nur vor, wenn das KI-System die bei der Entwicklung verwendeten Personendaten nach Abschluss der Entwicklung enthält, wenn diese also im KI-System gespeichert sind. Bei den heute im Vordergrund stehenden Systemen, die auf maschinellem Lernen beruhen, ist das aber nicht der Fall.

Auch wenn Daten nicht als solche in KI-Systemen gespeichert sind, ist es doch möglich, den Systemen durch **geeignete Abfragen («prompts»)** oder **gezielte Angriffe** Daten zu entnehmen, die für das Training der Systeme verwendet wurden, bspw. durch sog. «inversion attacks». Anders als bei klassischen Datenbanken werden die Daten den KI-Systemen dabei aber nicht entnommen. Bei generativen KI-Systemen werden die Daten vielmehr aufgrund eines «prompts» erneut erzeugt, während es bei gewissen Angriffen, bspw. «inversion attacks», möglich ist, durch den Vergleich eines selbst trainierten Modells mit dem angegriffenen Modell Rückschlüsse auf die beim Training des angegriffenen Modells verwendeten Daten zu ziehen.

Diese Vorgänge können als (unfreiwilliges) **Bekanntgeben von Personendaten** (Art. 5 lit. e DSG) verstanden werden, weil der Anbieter mit dem KI-System zugleich die darin enthaltenen Personendaten an den Betreiber des KI-Systems übermittelt oder diesem zugänglich macht. Wie bei der Anonymisierung von Personendaten drängt sich aber auch hier eine **differenzierte Betrachtung**

auf: Dass es technisch möglich ist, einem trainierten KI-System Personendaten zu entnehmen, bedeutet nicht, dass die Daten im datenschutzrechtlichen Sinn in diesem System gespeichert sind, bei dessen Verwendung bearbeitet und Dritten bekannt gegeben werden. Ist ein KI-System durch geeignete technische und organisatorische Massnahmen hinreichend gegen Angriffe geschützt und das Risiko der Entnahme von Personendaten gering, ist das Anbieten eines trainierten Modells deshalb nicht als Bekanntgeben von Personendaten zu qualifizieren. Ist der Schutz aber ungenügend, liegt ein Bekanntgeben vor.

Das Bekanntgeben ist ein Bearbeiten, das den Vorgaben des Datenschutzrechts genügen muss (Art. 5 lit. d DSG). Da dieser Vorgang für die betroffenen Personen nicht erkennbar ist, liegt in aller Regel ein Verstoss gegen den **Grundsatz der Transparenz** und häufig auch ein Verstoss gegen den **Grundsatz der Zweckbindung** vor. Ist ein KI-System ungenügend gegen die Entnahme von Personendaten geschützt, sind zudem die Vorgaben des **Grundsatzes der Datensicherheit** nicht eingehalten. Denn dieser verlangt, dass bei der Entwicklung, beim Anbieten und bei der Verwendung eines KI-Systems unter Berücksichtigung des Standes der Technik und mit Blick auf das Risiko alle geeigneten technischen und organisatorischen Massnahmen getroffen werden, die erforderlich sind, um ein «reverse engineering» der beim Trainieren verwendeten Personendaten, zu vermeiden (Art. 8 DSG). Während ein Verstoss gegen die Grundsätze der Transparenz und der Zweckbindung wohl durch ein überwiegendes Interesse gerechtfertigt werden kann, lässt sich ein Verstoss gegen den Grundsatz der Datensicherheit nach verbreiteter Auffassung nicht rechtfertigen.

Damit ist klar, dass beim Anbieten von KI-Systemen alle geeigneten technischen und organisatorischen Massnahmen getroffen werden müssen, die erforderlich sind, um ein «reverse engineering» von Personendaten zu verhindern. Ist das nicht der Fall, liegt ein Bekanntgeben von Personendaten und ein Verstoss gegen den Grundsatz der Datensicherheit vor, der nicht gerechtfertigt werden kann.

Verwendung

Auch bei der Verwendung von KI-Systemen sind die Vorgaben des Datenschutzrechts einzuhalten. Bei vielen KI-Systemen besteht der Input aus Personendaten, bspw. bei Systemen für die Bewertung von Bewerbungen oder bei Systemen zur Berechnung eines Kreditscores. Auch der Output dieser Systeme wird regelmässig aus Personendaten bestehen, bspw. der Kreditscore einer bestimmten Person oder die Empfehlung des Systems, eine Bewerberin für ein Gespräch einzuladen. Aus datenschutzrechtlicher Sicht bestehen allerdings **keine grundlegenden Unterschiede** zwischen der Bearbeitung von Personendaten durch ein KI-System und der Bearbeitung durch andere Systeme. Es ist deshalb nur auf einige besondere Fragestellungen hinzuweisen:

Für die Bearbeitung von Personendaten bei der Verwendung von KI-Systemen ist als **Verantwortlicher** zu qualifizieren, wer allein oder zusammen mit anderen über den Zweck und die Mittel der Bearbeitung entscheidet (Art. 4 lit. j DSG). Wer (allein oder gemeinsam) als Verantwortlicher und wer allenfalls als **Auftragsbearbeiter** zu qualifizieren ist, hängt von der konkreten Konstellation ab und kann nur im Einzelfall beurteilt werden. Sind mehrere Beteiligte involviert, kann die Frage durchaus komplex sein. Es lassen sich deshalb nur einige allgemeine Aussagen machen: Werden Personendaten als Input von standardisierten KI-Systemen (bspw. ChatGPT oder Gemini) verwendet, ist der Betreiber des Systems für die Bearbeitung dieser Daten als Verantwortlicher zu qualifizieren, der Anbieter des Systems und der von diesem eingesetzte Cloud-Provider als Auftragsbearbeiter. Betreibt der Betreiber das System auf eigenen Servern, ist er alleiniger Verantwortlicher. Zum Verantwortlichen wird der Anbieter nur, wenn er die Inputdaten auch für eigene Zwecke verwendet, bspw. für die Weiterentwicklung seines KI-Systems. Zum unbeteiligten Dritten würde der Anbieter hingegen, wenn er bei der Verwendung des KI-Systems keine Personendaten bearbeiten würde, weil die Daten «end-to-end» verschlüsselt wären. Das wird allerdings kaum vorkommen, weil die Daten vom System in aller Regel nicht in verschlüsselter Form bearbeitet werden können. Besteht der Output eines KI-Systems aus Personendaten, entscheidet der Betreiber in aller Regel allein über die Zwecke und Mittel der Bearbeitung dieser Daten und ist damit auch alleiniger Verantwortlicher.

Besteht der **Output eines KI-Systems** aus Personendaten, müssen die Grundsätze der Datenbearbeitung auch mit Bezug auf den Output eingehalten werden. Da dieser nicht notwendigerweise richtig ist, kann ein Verstoss gegen den **Grundsatz der Richtigkeit** vorliegen. Besondere Probleme können sich bei der Verwendung von generativen KI-Systemen ergeben, weil

diese bisweilen «halluzinieren», also frei erfundene Angaben über Personen erzeugen. Auch in diesem Fall liegen falsche Personendaten vor. Der Grundsatz der Richtigkeit gilt allerdings nicht absolut; die Richtigkeit ist vielmehr mit Bezug auf den Zweck der Bearbeitung zu bestimmen. Ist es für diesen irrelevant, ob eine bestimmte Angabe zutrifft (bspw. eine falsche Telefonnummer in einem System für den Versand von Newsletters), ist der Grundsatz der Richtigkeit nicht verletzt. Auch in diesen Fällen hat die betroffene Person aber Anspruch auf Berichtigung unrichtiger Personendaten (Art. 32 Abs. 1 DSG).

Die Verwendung von KI-Systemen wird in vielen Konstellationen als **automatisierte Einzelentscheidung** zu qualifizieren sein. Eine solche liegt vor, wenn eine Entscheidung ausschliesslich auf der automatisierten Bearbeitung von Personendaten beruht und die Entscheidung für die betroffene Person mit einer Rechtsfolge verbunden ist oder sie erheblich beeinträchtigt (Art. 21 Abs. 1 DSG), bspw. wenn ein KI-System darüber entscheidet, ob einer Konsumentin ein Kredit angeboten wird. Wird eine Einzelentscheidung automatisiert gefällt, muss der Verantwortliche die betroffene Person informieren (Art. 21 Abs. 1 DSG) und ihr auf Antrag die Möglichkeit geben, ihren Standpunkt darzulegen (Art. 21 Abs. 2 DSG). Zudem kann die betroffene Person verlangen, dass die automatisierte Einzelentscheidung von einer natürlichen Person überprüft wird (Art. 21 Abs. 2 DSG). Diese Pflichten entfallen, wenn die Entscheidung in unmittelbarem Zusammenhang mit dem Abschluss oder der Abwicklung eines Vertrags steht und dem Begehren der betroffenen Person stattgegeben worden ist oder wenn die betroffene Person eingewilligt hat, dass die Entscheidung automatisiert erfolgt (Art. 22 Abs. 3 DSG). Zudem haben die betroffenen Personen das Recht, vom Verantwortlichen Informationen über die Logik zu erhalten, auf der die automatisierte Einzelentscheidung beruht (Art. 25 Abs. 2 lit. f DSG). Diese Vorgaben lassen sich bei der Verwendung von KI-Systemen für automatisierte Einzelentscheidungen durchaus erfüllen, auch wenn es eine Herausforderung ist, die Erklärbarkeit und Nachvollziehbarkeit solcher Systeme zu gewährleisten.

Impressum

Herausgeberin:
Center for Information Technology,
Society, and Law (ITSL)
Universität Zürich
Rämistrasse 74|38
8001 Zürich