KI: Wie lässt sich die DSGVO einhalten?

05. April 2022

Die künstliche Intelligenz wirft entscheidende und neue Fragen auf, insbesondere im Hinblick auf den Datenschutz. Die CNIL erinnert an die wichtigsten Grundsätze des Datenschutzgesetzes und der DSGVO, die es zu beachten gilt, sowie an ihre Positionen zu einigen spezifischeren Aspekten.

Einen Zweck festlegen

Das Prinzip

Um die DSGVO einzuhalten, muss ein System künstlicher Intelligenz (KI), das auf der Auswertung personenbezogener Daten beruht, immer **mit einem** klar definierten **Zweck** (Ziel) entwickelt, trainiert und eingesetzt werden.

Dieses Ziel muss bestimmt sein, d. h. es muss bereits im Vorfeld, bei der Konzeption des Projekts, festgelegt werden. Es muss auch legitim sein, d. h. mit den Aufgaben der Organisation vereinbar. Schließlich muss das Ziel explizit sein, d. h. es muss bekannt und verständlich sein.

> Mehr zum Thema: Einen Zweck festlegen

In der Praxis

Wie bei jeder Datenverarbeitung, aber noch mehr bei der Verarbeitung sehr großer Mengen personenbezogener Daten - wie es bei KI-Systemen sehr oft der Fall ist - muss sichergestellt werden, dass dieser Grundsatz eingehalten wird.

Insbesondere, weil es der Zweck ist, der sicherstellt, dass nur relevante Daten verwendet werden und dass die gewählte Aufbewahrungsdauer angemessen ist.

Lernen vs Produktion: Der besondere Fall von KI-Systemen

Die Einrichtung eines KI-Systems, das auf maschinellem Lernen beruht, erfordert die Abfolge von zwei Phasen:

1. **Die Lernphase**

In der Lernphase wird ein KI-System und insbesondere ein Modell entworfen, entwickelt und trainiert, d. h. eine Darstellung dessen, was das KI-System aus den Trainingsdaten gelernt hat.

2. **Die Produktionsphase**

In der Produktionsphase wird das in Schritt 1 erhaltene KI-System operativ eingesetzt.

Aus Sicht des Datenschutzes **erfüllen diese beiden Schritte nicht denselben Zweck und müssen daher getrennt werden.**

In beiden Fällen muss der Zweck der Verarbeitung personenbezogener Daten in jeder dieser Phasen bestimmt, legitim und explizit sein.

Eine Rechtsgrundlage schaffen

Das Prinzip

Wie jede Verarbeitung darf auch ein KI-System, das personenbezogene Daten nutzt, nur dann eingesetzt werden, wenn es einer gesetzlich vorgeschriebenen Rechtfertigung entspricht. In der DSGVO gibt es sechs solcher Rechtfertigungsgründe: Einwilligung, Erfüllung einer rechtlichen Verpflichtung, Erfüllung eines Vertrags, Erfüllung einer Aufgabe im öffentlichen Interesse, Wahrung lebenswichtiger Interessen, Verfolgung eines berechtigten Interesses. Konkret bedeutet dies, dass die Rechtsgrundlage das ist, was einer Organisation das Recht gibt, personenbezogene Daten zu verarbeiten. Die Wahl dieser Rechtsgrundlage ist daher ein erster unverzichtbarer Schritt, um sicherzustellen, dass die Verarbeitung den Vorschriften entspricht. Je nachdem, welche Grundlage gewählt wird, können die Pflichten der Organisation und die Rechte der Personen variieren.

Diese Wahl der Rechtsgrundlage muss vor der Durchführung der Datenverarbeitung erfolgen.

> Mehr zum Thema: Die rechtlichen Grundlagen

Auch wenn sich die Einrichtung eines KI-Systems nicht grundlegend von der Verarbeitung anderer personenbezogener Daten unterscheidet, gibt es doch einige Besonderheiten, auf die Sie achten sollten. KI-Systeme, insbesondere solche, die auf *maschinellem* Lernen basieren, erfordern die Auswertung von Daten in der Lernphase, bevor sie in der Betriebsphase auf andere angewendet werden können.

In jedem Fall darf ein KI-System sowohl in der Lern- als auch in der Betriebsphase nicht auf illegal erhobenen personenbezogenen Daten basieren. Der nächste Abschnitt,
"Eine Datenbank aufbauen" bietet weitere Informationen.

Wenn die Daten nach einem anderen System (wie z. B. dem der Polizei-Justiz-Richtlinie) erhoben wurden, fällt eine Verarbeitung personenbezogener Daten zu Lernzwecken außer in Sonderfällen außerdem unter die DSGVO, sofern :

- diese Lernphase unterscheidet sich deutlich von der Phase der operativen Umsetzung des KI-Systems (siehe Kasten "Lernen VS Produktion (z. B. "Der besondere Fall von KI-Systemen" aus dem vorherigen Abschnitt "Einen Zweck definieren") ;
- sein ausschließlicher Zweck in der Entwicklung oder Verbesserung der Leistung eines KI-Systems besteht.

Achtung: Der Zweck der "wissenschaftlichen Forschung" kann für sich genommen keine Rechtsgrundlage für die Verarbeitung darstellen. Nur die in der DSGVO aufgelisteten Rechtsgrundlagen können eine Verarbeitung personenbezogener Daten ermöglichen.

Eine Datenbank aufbauen

KI-Systeme, insbesondere solche, die auf maschinellem Lernen basieren, erfordern die Verwendung großer Datenmengen. Diese Daten sind sowohl für das Training der Systeme als auch für die Bewertung, den Vergleich (*Benchmarking) und die* Validierung unerlässlich. Der Aufbau von Datenbanken war schon immer eine Herausforderung für die Informatikforschung und stellt einen großen Aufwand dar, da es sich

dabei um Folgendes handelt

die Daten mit Annotationen zu versehen, die die Daten beschreiben, und sie zu kategorisieren, zu bereinigen, zu normalisieren und so weiter. Daher ist dies eine wesentliche Herausforderung für die Verarbeitung von künstlicher Intelligenz.

In der Praxis

Es gibt zwei Hauptmöglichkeiten für den Aufbau von Datenbanken: die spezifische Erhebung personenbezogener Daten zu diesem Zweck und die Wiederverwendung von Daten, die bereits für einen anderen Zweck erhoben wurden. Im letzteren Fall stellt sich die Frage, ob die Zwecke, für die die Daten ursprünglich erhoben wurden, und die Bedingungen, unter denen die ursprüngliche Datenbank aufgebaut wurde, miteinander vereinbar sind.

In jedem Fall darf die Einrichtung von Datenbanken mit personenbezogenen Daten, die sehr oft auf langen Aufbewahrungsfristen für die Daten beruhen, nicht auf Kosten der Rechte der betroffenen Personen gehen. Insbesondere muss sie mit Informationsmaßnahmen einhergehen :

- entweder vor dem Einsammeln ;
- oder innerhalb eines Monats nach Erhalt der Grundlagen durch Dritte.

Diese Information ist wesentlich, um die Ausübung anderer Rechte (Zugang, Berichtigung, Löschung, Widerspruch) zu ermöglichen.

Beispiele

- Im Gesundheitsbereich hatte die CNIL die Gelegenheit, sich zur Einrichtung von Gesundheitsdatenlagern zu äußern. In einem kürzlich veröffentlichten Referenzsystem wird erläutert, in welchem Rahmen Daten gesammelt und in einer einzigen Datenbank über einen längeren Zeitraum gespeichert werden können, wenn dies im Rahmen von Aufgaben von öffentlichem Interesse und für spätere Forschungszwecke geschieht.
- Im Rahmen eines Beratungsersuchens konnten die Dienststellen der CNIL die Weiterverwendung von Bildern aus dem Videoschutz in einem bestimmten Kontext für die Durchführung wissenschaftlicher Forschungen zum Verständnis von Massenbewegungen, einer Aufgabe aus dem Bereich der Computervision, zulassen. Es wurde jedoch klargestellt, dass zu diesem Zweck die Erhebung :
  - innerhalb der gesetzlichen Aufbewahrungsfrist für Bilder von Videoschutzmaßnahmen (1 Monat) stattfinden; und
  - mit Informationen für die betroffenen Personen versehen sein.

Daten minimieren

Das Prinzip

Die erhobenen und verwendeten personenbezogenen Daten müssen **angemessen**, **relevant** und **auf das für** den festgelegten Zweck **erforderliche Maß beschränkt** sein: Dies ist der Grundsatz der Datenminimierung. Besondere Aufmerksamkeit muss der Art der Daten gewidmet werden und dieser Grundsatz muss besonders streng angewandt werden, wenn es sich bei den verarbeiteten Daten um sensible Daten handelt (Artikel 9 DSGVO).

Die derzeit prominentesten und am meisten diskutierten KI-Systeme basieren auf besonders leistungsstarken Methoden des *maschinellen Lernens (Machine Learning*). Die Verbesserung dieser Methoden wurde durch die kombinierten Effekte :

- der Forschung und der Entwicklung neuer Ansätze ;
- der Erhöhung der verfügbaren Rechenleistung, die komplexere Operationen ermöglicht; und
- der Zunahme der verfügbaren Datenmengen.

Die Verwendung großer Datenmengen steht zwar im Mittelpunkt der Entwicklung und des Einsatzes von KI-Systemen**, das Prinzip der Minimierung ist jedoch nicht per se ein Hindernis für die Durchführung**

solcher Verarbeitungen.

In der Praxis

Es ist notwendig, die Arten von Daten zu bestimmen, die für das Training und den Betrieb eines KI-Systems erforderlich sind, z. B. durch Experimente und Tests mit fiktiven Daten, d. h. Daten, die die gleiche Struktur wie echte Daten aufweisen, aber nicht personenbezogen sind. Diese Daten sind dann keine personenbezogenen Daten.

Die Menge der für den Antrieb des Systems erforderlichen Daten muss ebenfalls genau geschätzt und in Verbindung mit dem Grundsatz der Verhältnismäßigkeit dem Zweck der Verarbeitung gegenübergestellt werden.

Denn die Lernphase (oder Trainingsphase) dient der Entwicklung eines KI-Systems und damit der Erkundung der Möglichkeiten des maschinellen Lernens und kann eine große Anzahl von Daten erfordern, von denen sich einige in der Einsatzphase letztlich als nutzlos erweisen.

Die Daten müssen also vernünftig genutzt werden. In der Praxis wird empfohlen, unter anderem folgende Maßnahmen zu ergreifen

- die Art und Menge der zu verwendenden Daten kritisch zu bewerten ;
- die Leistung des Systems zu überprüfen, wenn es mit neuen Daten gefüttert wird ;
- klar zwischen den Daten unterscheiden, die in der Lern- und Produktionsphase verwendet werden ;
- Mechanismen zur Pseudonymisierung oder Filterung/Obfuskation von Daten zu verwenden ;
- eine Dokumentation darüber zu erstellen und bereitzuhalten, wie der verwendete Datensatz und seine Eigenschaften zusammengestellt wurden (Datenquelle, Stichproben der Daten, Überprüfung der Datenintegrität, durchgeführte Bereinigungsmaßnahmen usw.) ;
- die Risiken für die betroffenen Personen regelmäßig neu zu bewerten (Privatsphäre, Diskriminierungs-/Bias-Risiko usw.) ;
- für die Datensicherheit zu sorgen und insbesondere die Zugriffsberechtigungen genau zu regeln, um die Risiken zu begrenzen.

Beispiel

Im Rahmen einer klinischen Forschung zur Ermittlung von Erklärungsvariablen für Prostatakrebs verweigerte die CNIL einem Pharmaunternehmen die Verarbeitung von Daten aus der gesamten aktiven Warteschlange der Krankenakten der verschiedenen an der Studie beteiligten Zentren.

Tatsächlich enthielt diese aktive Warteschlange **mehrere hundert Millionen Akten von Personen, die nicht an dem untersuchten Leiden litten** (und sogar Akten von Personen weiblichen Geschlechts!). Der Wunsch, diese Daten zu verarbeiten, der wissenschaftlich durch die Notwendigkeit erklärt wird, "echte Negative" zu haben, um einen Klassifikator effektiv zu trainieren, erschien in der Tat **unverhältnismäßig im Hinblick auf den Zweck der Verarbeitung** und nicht notwendig für die Entwicklung eines KI-Systems mit guter Leistung.

Lernen vs Produktion - der Sonderfall von KI-Systemen

Während der Lernphase ist ein relativ flexibler Rahmen für den Zugang zu Daten in ausreichender Menge und Vielfalt möglich, sofern die Gegenleistungen in einem angemessenen Verhältnis zu den Risiken der Verarbeitung stehen (dabei sind insbesondere die Art der Daten, ihre Menge und der Zweck des KI-Systems zu berücksichtigen). Die Maßnahmen können bestehen aus :

- ein Zugang, der auf eine begrenzte Anzahl von berechtigten Personen beschränkt ist ;
- eine zeitlich begrenzte Verarbeitung, die Pseudonymisierung von Daten ;
- Umsetzung geeigneter technischer und organisatorischer Maßnahmen ;
- usw.

Erst nach Abschluss der Lernphase kann der Einsatz des KI-Systems in der Produktionsphase in Betracht gezogen werden. In dieser zweiten Phase, in der die "Laborumgebung" verlassen wird, müssen stärkere Beschränkungen für die Verarbeitung eingeführt werden.

So wird es beispielsweise notwendig sein, **die Typologie der personenbezogenen Daten auf diejenigen zu beschränken, die sich nach der** Lernphase **als unverzichtbar erwiesen haben,** und **entsprechende Maßnahmen zu ergreifen,** da sich die Produktionszwänge von den Design- und Entwicklungszwängen unterscheiden, vorausgesetzt, dass diese erste Phase keine besonderen Risiken für den Einzelnen birgt.

Beispiele

- Im Rahmen eines von einer Behörde eingereichten Projekts hatte die CNIL die Gelegenheit, sich zum Unterschied zwischen der Lern- (oder Entwicklungs-) und der Betriebsphase (oder Produktionsphase) eines KI-Systems zu äußern. Dabei war vorgesehen, dass die erste (Lern-)Phase per Dekret genehmigt werden sollte. Wenn sich diese Phase als zufriedenstellend erwiesen hätte, hätte ein zweites Dekret die praktische Umsetzung dieses Bezugsrahmens für Fachleute und die breite Öffentlichkeit regeln sollen.
- Im Gesundheitsbereich wird klar zwischen den Forschungsphasen, die eine Formalität bei der CNIL erfordern (Genehmigung, Einhaltung einer Referenzmethode usw.), und den Nutzungsphasen in einem Behandlungspfad, die ihrerseits keine Formalität bei der CNIL erfordern, **unterschieden.**

Eine Aufbewahrungsfrist festlegen

Das Prinzip

Personenbezogene Daten können nicht unbegrenzt aufbewahrt werden. Die DSGVO schreibt vor, dass ein Zeitraum festgelegt werden muss, nach dessen Ablauf die Daten gelöscht oder in bestimmten Fällen archiviert werden müssen. Diese Aufbewahrungsdauer muss von dem für die Verarbeitung Verantwortlichen anhand des Zwecks, der zur Erhebung dieser Daten geführt hat, festgelegt werden.

> Mehr zum Thema: Die Aufbewahrungsfristen für Daten

Die Implementierung eines KI-Systems kann in vielen Fällen die Aufbewahrung personenbezogener Daten für einen längeren Zeitraum als bei anderen Verarbeitungen erfordern. Dies kann der Fall sein, um Datensätze für das Training und die Entwicklung neuer Systeme zu erstellen, aber auch, um die Anforderungen an die Rückverfolgbarkeit und die Leistungsmessung im Laufe der Zeit zu erfüllen, wenn das System in Produktion geht.

Die Notwendigkeit, eine Aufbewahrungsdauer für die von einer Verarbeitung verwendeten Daten festzulegen, stellt kein Hindernis für die Durchführung von KI-Verarbeitungen dar. Diese Dauer muss immer in einem angemessenen Verhältnis zum verfolgten Zweck stehen: Beispielsweise muss der Zweck der Leistungsmessung ausdrücklich für die Verwendung vorgesehen sein und die Daten, die zu diesem Zweck länger aufbewahrt werden, müssen angemessen ausgewählt werden. Der bloße Zweck der Leistungsmessung im Zeitverlauf reicht a priori nicht aus, um eine lange Aufbewahrung aller Daten zu rechtfertigen.

Bei KI-Verarbeitungen, die zu wissenschaftlichen Forschungszwecken durchgeführt werden, ist es außerdem möglich, die Daten für längere Zeiträume aufzubewahren.

Kontinuierliche Verbesserung betreuen

Die Unterscheidung zwischen Lern- und Produktionsphase ist nicht immer für alle KI-Systeme klar ersichtlich. Dies gilt insbesondere für sogenannte "kontinuierliche Lernsysteme", bei denen die Daten aus der Produktionsphase auch zur Verbesserung des Systems verwendet werden, so dass eine vollständige Rückkopplungsschleife entsteht. Der Prozess des Neulernens kann in folgenden Bereichen stattfinden

unterschiedliche Frequenzen, z. B. nach einigen Stunden, Tagen oder Monaten, je nach Ziel.

Die Fragen, die Sie sich stellen sollten

Abgesehen von den Risiken von Fehlentwicklungen, die dem kontinuierlichen Lernen innewohnen (Einführung diskriminierender Verzerrungen, Verschlechterung der Leistung usw.), wirft eine solche Verwendung von Daten für zwei unterschiedliche Zwecke (den Zweck, für den das KI-System in Produktion geht, und die intrinsische Verbesserung des Systems) aus datenschutzrechtlicher Sicht Fragen auf:

- Inwiefern sind diese beiden Zwecke untrennbar miteinander verbunden?
- Ist es in jedem Fall möglich, eine Trennung zwischen Lern- und Produktionsphase vorzunehmen?
- Wenn der Algorithmus von einem Herausgeber bereitgestellt und von einem dritten für die Verarbeitung Verantwortlichen verwendet wird, wie sind dann die Verantwortlichkeiten für die beiden Phasen der Verarbeitung aufzuteilen?

Beispiele

- In den Fällen, zu denen die CNIL Stellung nehmen musste, hat sie stets die Auffassung vertreten, dass es möglich ist, die Lern- und Produktionsphasen zu trennen, auch wenn diese miteinander verflochten sind. In ihrem Weißbuch über Sprachassistenten analysiert die CNIL beispielsweise den Anwendungsfall der Wiederverwendung von Daten, die von einem Sprachassistenten gesammelt wurden, um den Service zu verbessern. Das Beispiel der Annotation neuer Lernbeispiele zur Verbesserung der Leistung von Systemen der künstlichen Intelligenz wird speziell angesprochen, und es wird klar zwischen dieser Verarbeitung und der Verarbeitung unterschieden, die zur Ausführung der vom Nutzer des Sprachassistenten erwarteten Dienstleistung eingesetzt wird.
- In Bezug auf die Verteilung der Verantwortlichkeiten zwischen den Akteuren hat sich die CNIL kürzlich zur Frage der Weiterverwendung von Daten, die von einem für die Verarbeitung Verantwortlichen anvertraut wurden, durch einen Auftragsverarbeiter geäußert. Angewandt auf den Fall von KI-Systemen ist eine Weiterverwendung durch einen Systemanbieter rechtlich möglich, wenn mehrere Bedingungen erfüllt sind: Genehmigung des für die Verarbeitung Verantwortlichen, Kompatibilitätstest, Information und Wahrung der Rechte von Personen sowie Konformität der neu implementierten Verarbeitung.

Sich gegen die Risiken von KI-Modellen absichern

Die größten Risiken

Maschinelles Lernen beruht auf der Erstellung von Modellen. Diese sind Darstellungen dessen, was die KI-Systeme aus den Trainingsdaten gelernt haben. Seit etwa 2010 hat sich ein Forschungsfeld in der Informatik herausgebildet, das sich mit der Sicherheit von KI-Modellen und insbesondere mit den Möglichkeiten der Informationsgewinnung befasst, die erhebliche Auswirkungen auf die Vertraulichkeit persönlicher Daten haben kann.

Man spricht daher häufig von "**Membership Inference"-Angriffen**, **"Template Exfiltration"-Angriffen** oder auch "Template **Inversion"-Angriffen** (siehe LINC-Artikel "Kleine Taxonomie der Angriffe auf KI-Systeme").

Beispielsweise haben zahlreiche Studien gezeigt, dass große Sprachmodelle (GPT-3, BERT, XLM-R usw.) dazu neigen, sich bestimmte Textelemente, auf die sie trainiert wurden, zu "merken" (Name, Vorname, Adresse, Telefon- und Kreditkartennummer usw.). Die Möglichkeit, solche Angriffe durchzuführen und Informationen zu extrahieren, stellt die Natur dieser neuen, durch künstliche Intelligenz eingeführten Objekte in Frage. Daher müssen sowohl technische als auch organisatorische Maßnahmen ergriffen werden, um die Risiken zu minimieren (siehe LINC-Veröffentlichungen zur Sicherheit von KI-Systemen).

Außerdem kann ein KI-Modell, das mit personenbezogenen Daten trainiert wurde, nicht standardmäßig selbst als personenbezogene Daten (oder genauer gesagt als ein Satz personenbezogener Daten) betrachtet werden.
Ihre Gründung muss jedoch auf einer rechtmäßigen Nutzung der Daten im Sinne der DSGVO beruhen. Einige

Regulierungsbehörden konnten so die Löschung von KI-Modellen verlangen, die auf der Grundlage von illegal gesammelten Daten erstellt wurden (z. B. die *Federal Trade Commission* in den USA).

Schließlich kann es eine Datenverletzung darstellen, wenn ein KI-Modell Gegenstand eines **erfolgreichen Datenschutzangriffs ist** (z. B. durch Inferenzen der Zugehörigkeit, Exfiltration oder Inversion). In diesem Fall muss das betreffende Modell so schnell wie möglich entfernt werden und eine Datenverletzungsmeldung an die zuständige Datenschutzbehörde erfolgen, wenn die Verletzung wahrscheinlich zu einem Risiko für die Rechte und Freiheiten der betroffenen Personen führt.

Beispiele

Die CNIL hatte die Gelegenheit, sich mit verschiedenen Organisationen über den Status von KI-Modellen im Hinblick auf die DSGVO auszutauschen. Bisher ist die CNIL nicht der Ansicht, dass ein KI-Modell, das mit personenbezogenen Daten trainiert wird, notwendigerweise personenbezogene Daten enthält.

Da es jedoch echte Risiken für die Privatsphäre gibt, empfiehlt die CNIL, dass geeignete Maßnahmen ergriffen werden, um diese Risiken zu minimieren. So stellte sich im Rahmen der Begleitung eines der Gewinnerprojekte der "Sandbox" für personenbezogene Daten die Frage nach der Natur der lokal gelernten KI-Modelle, die bei der Anwendung von föderierten Lernmethoden an ein Orchestrierungszentrum zurückgespielt werden.

Information und Erklärbarkeit sicherstellen

Das Prinzip :

Der Transparenzgrundsatz der DSGVO verlangt, dass alle Informationen oder Mitteilungen über die Verarbeitung personenbezogener Daten **kurz**, **transparent**, **verständlich und leicht zugänglich** sein müssen, und zwar **in einfachen und klaren Worten**.

> Mehr zum Thema: Wie kann man Menschen informieren und für Transparenz sorgen?

In der Praxis

Während die Hauptprinzipien der DSGVO und des Datenschutzgesetzes im Falle von KI-Systemen gelten, können die Informationen, die den Personen gegeben werden müssen, variieren:

- wenn die Daten nicht direkt von dem Verantwortlichen, der das KI-System implementiert, gesammelt wurden und es schwierig ist, zu den betroffenen Personen zurückzukehren. Diese Problematik ist nicht spezifisch für KI-Verarbeitungen, findet sich aber häufig in diesen, insbesondere bei der Verwendung von Lerndatenbanken ;
- für die Ausübung bestimmter Rechte (insbesondere aus Art. 22 DSGVO) ist es unerlässlich, der betroffenen Person genaue Erläuterungen zu den Gründen zu geben, die zu der betreffenden Entscheidungsfindung geführt haben. Die Komplexität und Undurchsichtigkeit einiger KI-Systeme kann die Bereitstellung dieser Elemente erschweren.

In einigen Fällen ist es möglich, vom Recht auf Information abzuweichen, wenn die Daten nicht direkt bei den betroffenen Personen erhoben wurden, insbesondere wenn sich die Information dieser Personen nachweislich als unmöglich erweist oder einen unverhältnismäßigen Aufwand erfordert, z. B. bei KI-Verarbeitungen, die für Zwecke der wissenschaftlichen Forschung durchgeführt werden. In den jüngsten Veröffentlichungen der CNIL zum Thema wissenschaftliche Forschung (außerhalb des Gesundheitswesens) wird in einem der Praxisblätter speziell auf die Modalitäten für Ausnahmen vom Recht auf Information der Personen eingegangen.

Beispiel

Nach der Kontrolle einer Plattform, die die Voranmeldung für das erste Jahr einer Ausbildung nach dem Abitur ermöglichte, stellte die CNIL fest, dass keine Informationen über die Verwendung eines Algorithmus und dessen Funktionsweise für die Klassifizierung und Zuweisung von Personen innerhalb von Hochschuleinrichtungen vorlagen.

Diese Tatsachen stellten einen Verstoß gegen Artikel 39.I.5 des Gesetzes über Informatik und Freiheiten dar: "*Jede natürliche Person, die ihre Identität nachweist, hat das Recht, den Verantwortlichen einer Verarbeitung personenbezogener Daten zu befragen, um Folgendes zu erhalten: Informationen, die es ermöglichen, die Logik, die der automatisierten Verarbeitung zugrunde liegt, zu kennen und anzufechten, falls eine Entscheidung auf ihrer Grundlage getroffen wird und Rechtsfolgen für den Betroffenen hat.*".

Die CNIL forderte daher, dass das Treffen von Entscheidungen, die Rechtswirkungen für Personen haben, allein auf der Grundlage einer automatisierten Datenverarbeitung eingestellt wird. Insbesondere wurde die Einführung einer menschlichen Intervention gefordert, die es ermöglicht, die Bemerkungen der Personen zu berücksichtigen.

Die Ausübung von Rechten umsetzen

Das Prinzip :

Die von einer Verarbeitung betroffenen Personen haben Rechte, um die Kontrolle über ihre Daten zu behalten. Der für die Datei Verantwortliche muss ihnen erklären, wie sie diese ausüben können (bei wem? in welcher Form?, etc.).
Wenn sie ihre Rechte ausüben, müssen die Personen in der Regel innerhalb eines Monats eine Antwort erhalten.

> Mehr zum Thema: Die Rechte der Menschen respektieren

Wenn das KI-System die Verarbeitung personenbezogener Daten beinhaltet, muss sichergestellt werden, dass die in der DSGVO festgelegten Grundsätze für die Ausübung von Rechten durch Einzelpersonen eingehalten werden: Zugang (Artikel 15), Berichtigung (Artikel 16), Löschung (Artikel 17), Einschränkung (Artikel 18), Übertragbarkeit (Artikel 20) und Widerspruch (Artikel 21). Diese Rechte stellen einen wesentlichen Schutz für Einzelpersonen dar, indem sie es ihnen ermöglichen, nicht die Folgen eines automatisierten Systems zu erleiden, ohne die Möglichkeit zu haben, die Verarbeitung von Daten, die sie betreffen, zu verstehen und sich gegebenenfalls dagegen zu wehren. In der Praxis finden diese Rechte während des gesamten Lebenszyklus des KI-Systems Anwendung und umfassen daher auch personenbezogene Daten :

- die in den für das Lernen verwendeten Datenbanken enthalten sind ;
- in der Produktionsphase verarbeitet werden (dies kann auch die vom System erzeugten Outputs umfassen).

Die für die Verarbeitung Verantwortlichen müssen sich daher bereits bei der Konzeption des Systems darüber im Klaren sein, dass sie geeignete Mechanismen und Verfahren einbauen müssen, um auf Anfragen, die möglicherweise eingehen, reagieren zu können. Ausnahmen von der Ausübung bestimmter Rechte können im Fall von KI-Verarbeitungen, die zu wissenschaftlichen Forschungszwecken durchgeführt werden, mobilisiert werden.

Darüber hinaus können auch gelernte KI-Modelle personenbezogene Daten enthalten:

- **durch Konstruktion,** wie es bei einigen speziellen Algorithmen der Fall ist, die Bruchteile von Lerndaten enthalten können (z. B. SVM oder einige Clustering-Algorithmen) ;
- **durch Unfall,** wie im Abschnitt "Sich gegen die Risiken von KI-Modellen absichern" beschrieben.

Im ersten Fall, je nach den angebotenen technischen Möglichkeiten und der Fähigkeit des Verantwortlichen für die

Verarbeitung die betroffene Person (wieder) zu identifizieren, kann daher die Ausübung der Rechte von Personen erfolgen.

Im zweiten Fall kann es schwierig oder gar unmöglich sein, die Rechte der betroffenen Personen auszuüben und zu erfüllen.

Der für die Verarbeitung Verantwortliche darf keine zusätzlichen Informationen zur Identifizierung der betroffenen Person allein zum Zweck der Einhaltung der DSGVO erheben oder aufbewahren (Artikel 11). Daher kann sich die Identifizierung von Personen in einigen Fällen als kompliziert erweisen. Wenn der für die Verarbeitung Verantwortliche nachweist, dass er dazu nicht in der Lage ist, kann er dann die Rechte ausschließen, ohne dass dies die Personen daran hindert, zusätzliche Informationen bereitzustellen, mit denen sie in der Verarbeitung erneut identifiziert werden könnten. Dies wird insbesondere dann der Fall sein, wenn eine Person der Meinung ist, dass ein KI-System sie auf besondere Weise behandelt.

Die Befolgung einer Aufforderung zur Berichtigung oder Löschung von Lerndaten bedeutet daher nicht zwangsläufig die Berichtigung oder Löschung des/der KI-Modelle(s), das/die aus diesen Daten erzeugt wurde(n).

Rahmen für die automatisierte Entscheidungsfindung

Das Prinzip

Personen haben das Recht, nicht Gegenstand einer vollautomatisierten Entscheidung zu sein (Art. 22 DSGVO) - die häufig auf Profiling beruht -, die eine rechtliche Wirkung hat oder diese erheblich beeinträchtigt. Eine Organisation darf diese Art von Entscheidung dennoch automatisieren, wenn :

- die Person hat ihre ausdrückliche Zustimmung gegeben ;
- die Entscheidung ist für einen Vertrag mit der Organisation notwendig; oder
- die automatisierte Entscheidung durch besondere gesetzliche Bestimmungen erlaubt ist.

In diesen Fällen muss es für die Person möglich sein :

- darüber informiert zu werden, dass eine vollautomatische Entscheidung gegen sie getroffen wurde ;
- nach der Logik und den Kriterien fragen, die zur Entscheidungsfindung herangezogen wurden;
- die Entscheidung anzufechten und ihre Meinung zu äußern ;
- die Intervention eines Menschen zu verlangen, der die Entscheidung noch einmal überprüfen kann.

Mehr zum Thema: Profiling und vollautomatische Entscheidung

In der Praxis

KI-Systeme sind häufig Bestandteil von Verarbeitungsprozessen, die Mechanismen zur automatisierten Entscheidungsfindung implementieren können.

Der für die Verarbeitung Verantwortliche muss daher in seinem Fall die Möglichkeit eines menschlichen Eingreifens seinerseits vorsehen, um der betroffenen Person zu ermöglichen, eine erneute Prüfung ihrer Situation zu erhalten, ihren Standpunkt darzulegen, eine Erklärung für die getroffene Entscheidung zu erhalten und die Entscheidung anzufechten. Im Falle einer Unterstützung bei der Entscheidungsfindung sind ebenfalls Garantien erforderlich, insbesondere in Bezug auf die Information.

Beispiele

Es stellt sich die Frage nach dem Umriss der Definition dessen, was eine automatisierte individuelle Entscheidung ist, und nach dem Grad des wünschenswerten menschlichen Eingreifens im Falle von

KI-Systemen.

In ihrem Entwurf eines Leitfadens zur Personalbeschaffung analysiert die CNIL die Verwendung bestimmter Tools zur automatischen Klassifizierung oder sogar Bewertung von Bewerbungen. Solche Lösungen können dazu führen, dass eine "*ausschließlich auf einer automatischen Verarbeitung beruhende Entscheidung*" per Design getroffen wird, wenn Bewerbungen aussortiert werden oder wenn Bewerbungen beispielsweise aus Zeitmangel auf eine zweitrangige, nicht von Menschen kontrollierte Ebene verbannt werden. Aufgrund der Risiken, die mit dieser für die Bewerber oft undurchsichtigen Art der Entscheidungsfindung verbunden sind, sind solche Verfahren nach der DSGVO grundsätzlich verboten. Ihr Einsatz ist nur unter außergewöhnlichen Bedingungen zulässig und unterliegt der Umsetzung besonderer Garantien, die die Rechte und Interessen der Bewerber sichern sollen.

Die CNIL hatte die Gelegenheit, eine Stellungnahme zu einer Datenverarbeitung abzugeben, die von einer Behörde durchgeführt wurde und darauf abzielte, versuchsweise auf die Auswertung von online frei zugänglichen Inhalten auf Plattformen zurückzugreifen, auf denen mehrere Parteien zum Zwecke des Verkaufs einer Ware, der Erbringung einer Dienstleistung oder des Austauschs oder der gemeinsamen Nutzung eines Inhalts, einer Ware oder einer Dienstleistung miteinander in Verbindung gebracht werden. In dieser Stellungnahme stellte die CNIL klar, dass die durch die Verarbeitung modellierten Daten keinesfalls zu einer automatischen Planung von Steuerprüfungen oder gar zu Entscheidungen führen dürften, die den Steuerpflichtigen direkt entgegengehalten werden können.

Bewerten Sie das System

Die Bewertungswinkel

Die Bewertung von KI-Systemen ist ein zentrales Thema und steht im Mittelpunkt des Verordnungsentwurfs der Europäischen Kommission. Aus Sicht des Datenschutzes ist diese unerlässlich, um :

- **Validierung des Ansatzes, der in der Entwurfs- und Entwicklungsphase des Systems** (der sogenannten "Lernphase") **getestet wurde**. Dabei soll auf möglichst wissenschaftliche und ehrliche Weise überprüft werden, ob es gemäß den Erwartungen der Entwickler funktioniert und ggf. gut für den Einsatz in der Produktionsphase geeignet ist.
- **Minimierung des Risikos, dass das System im Laufe der Zeit abdriftet**. Zum Beispiel, weil es sich an Personen mit anderen Profilen richtet als die Personen, deren Daten die Lernbasis bilden, oder weil das System regelmäßig neu trainiert wird, was zu einer Verschlechterung der Leistung führen kann, die potenziell schädlich für die betroffenen Personen ist.
- **Sicherstellen, dass das System nach dem Einsatz in der Produktion die betrieblichen Anforderungen erfüllt, für die es konzipiert wurde**. Es ist wichtig, die Leistung des Systems in der Lernphase von der Leistung des Systems in der Produktionsphase zu trennen, da die Qualität der Lernphase nicht die Qualität der Produktionsphase beeinflusst.

Beispiel

Im Zusammenhang mit der Erprobung einer Gesichtserkennungstechnologie verlangte die CNIL, dass der ihr übermittelten Bilanz auch ein strenges Bewertungsprotokoll beigefügt wird, mit dem der Beitrag dieser Technologie genau gemessen werden kann. In der Praxis forderte die CNIL insbesondere folgende Informationen an:

- objektive Leistungsmetriken, die von der wissenschaftlichen Gemeinschaft üblicherweise verwendet werden ;
- eine systematische Analyse von Systemfehlern und deren operativen Auswirkungen ;
- Elemente, die sich auf die Versuchsbedingungen beziehen (z. B. bei einem Computer-Vision-System: Tag/Nacht, Wetterbedingungen, Qualität der verwendeten Bilder, Widerstand gegen mögliche Anstößigkeiten usw.) ;
- Denkanstöße zu den potenziellen Diskriminierungsrisiken, die mit dem Einsatz dieses KI-Systems verbunden sind, speziell ;
- Elemente, die sich auf die Auswirkungen dieses Systems beziehen, wenn es in einem operativen Rahmen unter Berücksichtigung der Realitäten vor Ort eingesetzt wird (z. B. hat eine False-Positive-Rate von 10 % bei 10 Warnungen nicht die gleichen operativen Auswirkungen wie 10 % bei 1000 Warnungen).

Vermeidung von algorithmischer Diskriminierung

Die Herausforderungen

Der Einsatz von KI-Systemen kann auch das Risiko der Diskriminierung mit sich bringen. Die Gründe dafür sind vielfältig und können folgende Ursachen haben

- der für das Lernen verwendeten Daten, z. B. weil sie nicht repräsentativ sind oder weil sie zwar die "reale Welt" gut repräsentieren, aber dennoch einen diskriminierenden Charakter widerspiegeln (z. B. die Reproduktion von Lohnunterschieden zwischen Frauen und Männern); oder
- des Algorithmus selbst, der Konstruktionsfehler aufweisen würde. Diese Dimension, die auch im Verordnungsentwurf der Europäischen Kommission stark vertreten ist, erfordert eine besondere Berücksichtigung durch die für die Verarbeitung Verantwortlichen.

Beispiele

Bei der Kontrolle einer Organisation, die ein System zur automatischen Bewertung von Video-Lebensläufen einsetzte, die von Bewerbern im Rahmen einer Einstellungskampagne aufgenommen worden waren, stellte die CNIL diskriminierende Verzerrungen fest. In diesem Fall war das System, das die *sozialen* Kompetenzen (*social skills*) der Personen bewerten sollte, nicht in der Lage, die Vielfalt der Akzente der Personen zu berücksichtigen.

Die CNIL hatte die Gelegenheit, den Défenseur des droits (DDD) bei der Veröffentlichung des Berichts Algorithms: prévenir l'automatisation des discriminations (Algorithmen: die Automatisierung von Diskriminierung verhindern) zu unterstützen. Dieser ruft insbesondere zu einer kollektiven Bewusstseinsbildung auf und verpflichtet die Behörden und betroffenen Akteure, greifbare und praktische Maßnahmen zu ergreifen, um zu verhindern, dass Diskriminierungen durch diese Technologien reproduziert und verstärkt werden.
