Lernende Systeme

DIE PLATTFORM FÜR KÜNSTLICHE INTELLIGENZ

Datenschatz für KI nutzen, Datenschutz mit KI wahren

Technische und rechtliche Ansätze für eine datenschutzkonforme, gemeinwohlorientierte Datennutzung

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

acatech
DEUTSCHE AKADEMIE DER TECHNIKWISSENSCHAFTEN

WHITEPAPER

Müller-Quade, J., Houdeau, D. et al.
AG IT-Sicherheit, Privacy, Recht und Ethik

Inhalt

Zusammenfassung

Die stetig wachsende Menge und Verfügbarkeit an Daten – und die daraus resultierenden Werte für die (Weiter-)Verwertung konstituieren einen Datenschatz. Dieser bietet enorme Potenziale für die Entwicklung neuer Geschäftsmodelle, die Stärkung unseres Zusammenlebens und für digitale Nachhaltigkeit. Damit könnten wir auch datengetrieben unser Gemeinwohl stärken. Doch die Lage ist etwas kniffelig: Denn der bestehende Datenschutz-Rechtsrahmen in Europa dient zwar dazu, Datengebende umfassend zu schützen, schafft aber zugleich Interpretationsspielräume und Unsicherheiten in der Rechtsauslegung. Dies verkompliziert die Hebung des Datenschatzes für die Entwicklung datengetriebener Dienste und Services – insbesondere für KI-Systeme, die in hohem Maße auf Daten für das Training und die kontinuierliche Verbesserung angewiesen sind. Trotz ihrer großen Chancen für unser Gemeinwohl schrecken viele Unternehmen vor einem KI-Einsatz aus Datenschutzbedenken zurück. Dieses Spannungsfeld bedarf einer Auflösung, damit der Datenschatz gehoben werden kann.

Die Autorinnen und Autoren der Arbeitsgruppe IT-Sicherheit, Privacy, Recht und Ethik der Plattform Lernende Systeme zeigen in diesem Whitepaper (1.) das Interaktionsfeld zwischen Datenschutz und Datennutzung mitsamt Hemmnissen und Potenzialen auf, diskutieren (2.) technische Ansätze für eine datenschutzkonforme Datennutzung für KI-Systeme und artikulieren (3.) Gestaltungsoptionen, wie ein regulativer und technologischer Rahmen für eine zukunftssichere, gemeinwohlorientierte Datennutzung für KI ausgestaltet werden kann. Die Ergebnisse dieses analytischen Teils hat die Arbeitsgruppe zudem in drei Anwendungsszenarien aus dem Bereich Mobilität, Bildung und Gesundheit aufbereitet, um ihre zentrale Botschaft in kompakter Form darzustellen. Der Kern ihrer Botschaft lautet: Datenschutz und eine flexibilisierte Datennutzung müssen zusammengedacht werden! Denn ihre symbiotische Verknüpfung in technischen Ansätzen (Privacy Tech) basierend auf der europäischen Werteordnung kann Innovationsmotor für die europäische KI-Entwicklung und -Anwendung sein. Zum Wohle der Gesellschaft sollte daher die flexibilisierte Nutzung von Daten für KI-Systeme im Gemeinwohlinteresse handlungssichere Anerkennung finden. Dafür bedarf es eines Dialogs zwischen Regulatoren, Datenschützenden, Zivilgesellschaft und Unternehmen.

1 Einleitung

In unserer Gegenwart nehmen Daten eine nicht mehr wegzudenkende Rolle ein. Wir erzeugen als Privatpersonen Daten, wenn wir uns mit unseren Devices fortbewegen, sie in der Freizeit nutzen oder im Arbeitsumfeld mit smarten Produkten interagieren. Im industriellen Umfeld entstehen Daten durch Maschinen- oder Produktnutzung. Umgebungs- und Umweltdaten werden durch Satelliten oder Sensoren gesammelt. Kurzum: Wir sitzen auf einem riesigen Datenschatz, der stetig weiterwächst. Dieser Schatz ist eine wertvolle Ressource, gerade auch für das Gemeinwohl. Über Big Data können wir Muster erkennen, die es uns erlauben, Krankheitsausbrüche oder Extremwetterereignisse zu analysieren oder vorauszusagen. Wir können Daten so miteinander verknüpfen, dass wir neue Leistungen schaffen, die etwa Menschen mit Beeinträchtigungen neue Lebensqualität schenken, gefährliche Arbeitsprozesse optimieren oder die ökologische Nachhaltigkeit durch Ressourcenoptimierung stärken. Künstliche Intelligenz (KI) ist eine Schlüsseltechnologie, um solche Gemeinwohlpotenziale zu heben. Damit KI diese Potenziale entfalten kann, ist sie auf die Verwendung des Datenschatzes angewiesen.

Obwohl KI in der Wirtschaft als Schlüsseltechnologie wahrgenommen wird, setzen sie laut einer repräsentativen Umfrage nur 13 Prozent der deutschen Unternehmen ein (Bitkom Research 2021, S. 26), und dies, obwohl viele nicht-personenbezogene Daten, die im unternehmerischen Kontext anfallen, verhältnismäßig aufwandsarm genutzt werden könnten. Als wesentliche Hürde für den Einsatz von KI benennt jedes zweite Unternehmen die Anforderungen an den Datenschutz bei der Verwendung personenbezogener Daten. Die daraus resultierende Rechtsunsicherheit ist nach den hohen Investitionskosten das zweitgrößte Hindernis für den Einsatz von KI in der deutschen Wirtschaft (Bitkom Research 2021, S. 41). Der Datenschatz für KI wird also aktuell nicht umfassend gehoben.

Als Grundvoraussetzung für die gesellschaftliche Akzeptanz von KI als Schlüsseltechnologie gilt es zugleich, die informationelle Selbstbestimmung – also das Recht des Einzelnen, selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu entscheiden – zu sichern. Dies zeigt eine repräsentative Umfrage des TÜV-Verbandes, in der sich 78 Prozent der deutschen Bevölkerung für eine entsprechende gesetzliche Regelung aussprachen (TÜV-Verband 2020). Die Rechtsverordnung für personenbezogene Daten wird in der Europäischen Union (EU) aktuell vor allem über die Datenschutzgrundverordnung (DSGVO) und zukünftig auch über den derzeit in Erarbeitung befindlichen *Artificial Intelligence Act (AI Act)* der Europäischen Union vollzogen (Stiftung Datenschutz 2021, S.11f). Datenschutz ist im Sinne der Wahrung informationeller Selbstbestimmung ein Kernerfordernis und muss bei der Hebung des Datenschatzes gewährleistet sein.

Gleichzeitig ist es notwendig, den Datenschatz – bestehend aus noch häufig ungenutzten, aber wertvollen (un-)strukturierten Daten – insbesondere für eine gemeinwohlorientierte KI-Nutzung nutzbar zu machen: einerseits, um Wertschöpfungspotenziale durch KI (u. a. Enquete-Kommission KI 2020, S. 168f; PricewaterhouseCoopers 2018, S. 12f) zu realisieren und KI so als Schlüsseltechnologie zur Wahrung und Stärkung der Wettbewerbsfähigkeit der deutschen Wirtschaft zu etablieren; andererseits, um KI vor allem zur Stärkung und Förderung des Gemeinwohls einzusetzen. Gerade im letztgenannten Punkt liegt das besondere Potenzial des derzeit kaum gehobenen Datenschatzes: So könnten beispielsweise Gesundheitsdaten von Patientinnen und Patienten in großer Menge dafür verwendet werden, die Entstehung von Erkrankungen besser vorhersagen zu können (Plattform Lernende Systeme 2019; Dössel & Lenarz (Hrsg.) 2023) oder eine personalisierte Gesundheitsversorgung zu ermöglichen. Bewegungsdaten von Personen und Fahrzeugen könnten zur Routenoptimierung und für das Flottenmanagement eingesetzt werden, oder auch um Risiken im Straßenverkehr zu reduzieren, beispielsweise beim autonomen Fahren durch Objekterkennung (Plattform Lernende

Systeme 2021). Die Analyse von menschlichen Verhaltensdaten in großer Menge bietet zudem das Potenzial, maßgeschneiderte Unterstützungsdienstleistungen für individuelles Verhalten zu entwickeln – beispielsweise Lernangebote im Bildungsbereich. All die genannten Anwendungen bzw. Möglichkeiten machen deutlich, dass die Hebung dieses Datenschatzes durch KI weitreichende Potenziale für das Gemeinwohl bietet. Um diese Potenziale nicht zu verschenken, sollten Möglichkeiten der Datennutzung für KI in ein produktives Verhältnis zum Datenschutz gesetzt werden: im Sinne einer flexibilisierten Datennutzung im Gemeinwohlinteresse bei gleichzeitiger Wahrung des Datenschutzes.

KURZINFO

Gemeinwohlinteresse

… meint einen Nutzen, der möglichst vielen Mitgliedern eines Gemeinwesens zugutekommen soll, um ihr Wohl – etwa sozial, gesundheitlich oder ökonomisch – zu verbessern. Die Datennutzung im Gemeinwohlinteresse zielt folglich nicht allein auf individuelle – wirtschaftliche, eigennützige, freundschaftliche oder familiäre – Ziele, sondern ist durch eine Besserstellung des Gemeinwesens motiviert (Trésoret 2018), etwa die Nutzung von PatientInnendaten für KI-Systeme, um Krankheiten besser vorhersagen zu können.

Datenschatz

… meint die Menge an wertvollen (un-)strukturierten Daten, die Potenziale für die Entwicklung von digitalen Lösungen in diversen Anwendungsbereichen liefern. Sie sind häufig ungenutzt, weil etwa ihr Wert für die Entwicklung neuartiger Lösungen nicht erkannt wird, oder auch weil technische oder rechtliche Auflagen ihre Nutzung aktuell verunmöglichen. Für die Datennutzung für KI-Systeme trifft dies insbesondere auf Daten mit Personenbezug zu, etwa wenn PatientInnendaten genutzt werden könnten.

Es bestehen bereits technische Möglichkeiten einer flexibilisierten Datennutzung im Gemeinwohlinteresse bei gleichzeitiger technischer Wahrung des Datenschutzes (etwa durch Privacy-Preserving Machine Learning oder Datentreuhänder). Allerdings sind diese Möglichkeiten häufig wenig bekannt und juristisch kaum anerkannt, weshalb sie die Interpretationsspielräume und Unsicherheiten in der Rechtsauslegung bei der Nutzbarmachung des Datenschatzes noch nicht auflösen können. Die Folge davon ist: Die wirtschaftliche Anwendung von KI in der Breite sowie die Ausschöpfung möglicher Gemeinwohlpotenziale wird erschwert.

Mit dem Whitepaper stellt die Arbeitsgruppe IT-Sicherheit, Privacy, Recht und Ethik der Plattform Lernende Systeme mögliche technische und juristische Gestaltungsoptionen vor, wie eine gemeinwohlorientierte Datennutzung für Entwicklerinnen und Entwickler sowie Anwenderinnen und Anwender von KI-Systemen – seien es Unternehmen, Behörden oder Nichtregierungsorganisationen – unter Gewährleistung des Datenschutzes rechtssicher flexibilisiert werden kann. Die Zielsetzung liegt darin, über einen technikneutralen Ansatz Empfehlungen zu definieren, wie Interpretationsspielräume durch anwendungsspezifische Gesetzgebung geschlossen werden können, um die nachhaltige Realisierung ökonomischer Wertschöpfungspotenziale zum Wohle der Gesellschaft durch Hebung des Datenschatzes zu ermöglichen. Illustriert werden die Ausführungen durch Anwendungsszenarien aus besonders datenschutzrelevanten Anwendungsdomänen wie der Mobilität (siehe LEASYNG-Anwendungsszenario), der Bildung (siehe learn.digital-Anwendungsszenario) und der Gesundheit (siehe vAItality-Anwendungsszenario). So gelingt eine praktische Konkretisierung der aufgeworfenen Potenziale und Probleme im Interaktionsfeld Datenschutz und gemeinwohlorientierte Datennutzung.

Im Folgenden werden aktuelle Problemlagen für die datenschutzkonforme Datennutzung von KI-Systemen detailliert dargestellt (Kapitel 2). Darauf aufbauend werden pointiert inkrementelle und innovative Ansatzpunkte evaluiert, die über eine technische Sicherstellung des Datenschutzes eine Flexibilisierung der datenschutzwahrenden gemeinwohlorientierten Datennutzung für KI-Systeme ermöglichen könnten (Kapitel 3). Darunter fallen unter anderem eine Evaluierung von *Privacy-Preserving Machine Learning* (PPML)-Ansätzen und eine Auseinandersetzung mit Datentreuhänder-Ansätzen. Auch wird evaluiert, inwiefern *Erklärbare KI* (Explainable AI; kurz: XAI) zur Transparenzsteigerung und damit Datennutzungsliberalisierung eingesetzt werden könnte. Wie solche Ansätze in den jeweiligen Anwendungsdomänen juristisch anerkannt und zugleich praktisch umgesetzt werden könnten, wird in den Gestaltungsoptionen vorgezeichnet (Kapitel 4). Diese richten sich an Stakeholder aus der Wirtschaft (Unternehmen und Unternehmensverbände), Politik und Behörden sowie der Forschung. So wird ein Beitrag zur Diskussion geleistet, wie Daten als Ressource für einen vereinfachten und erleichterten Einsatz von KI-Systemen genutzt werden können.

2 Interaktionsfeld Datenschutz und gemeinwohlorientierte Datennutzung für KI-Systeme

Die Potenziale Künstlicher Intelligenz für das Gemeinwohl liegen auf der Hand. Im Gesundheitsbereich werden KI-Systeme basierend auf *Big Data Analytics* unter anderem erfolgreich für die Detektion und Therapie von Krankheiten eingesetzt (Plattform Lernende Systeme 2019), beispielsweise von Covid-19-Infektionen (Vaishya et al. 2020) oder Krebserkrankungen (Huang et al. 2020). Im Mobilitätsbereich helfen KI-Systeme beispielsweise bei der Routenoptimierung und reduzieren so den Energie- und Ressourcenverbrauch gemäß den UN-Nachhaltigkeitszielen; auch ihr Einsatz beim autonomen Fahren verspricht hier im Vergleich zu klassischen personengesteuerten Fahrzeugen eine Risikominimierung im Straßenverkehr (Plattform Lernende Systeme 2021). Im Bildungsbereich können KI-Systeme beispielsweise für die Optimierung einer altersgruppengerechten Nutzerführung in Lern-Apps eingesetzt werden. Im Bereich der öffentlichen Verwaltung besteht im Rahmen der Digitalisierung hohes Potenzial für KI-gestützte Automatisierung und Beschleunigung von Verwaltungsvorgängen (Antragsbearbeitung) und zu Planungszwecken (Infrastruktur-Bedarfsprognosen).

KI-Systeme sind in der Entwicklung für das Training ihrer *Machine Learning* (ML)-Modelle sowie für ihre Anwendung auf teilweise große Datenmengen angewiesen. Diese Datenmengen liegen bereits vor und sie wachsen stetig, weil unser Zusammen- und Arbeitsleben in großen Teilen digitalisiert ist. Die Hebung dieses Datenschatzes ist aber mitunter kompliziert. Während viele Unternehmensdaten für KI-Systeme wie Logistikdaten oder Qualitätsdaten „unbedenklicher" nutzbar sind (siehe Abbildung 1), werden besonders bei KI-basierten Dienstleistungen und Produkten, die für Endverbraucher konzipiert sind – sei es im Handel, im Mobilitäts- oder Gesundheitsbereich –, für klassische ML-Modelle in hohem Maße personenbezogene Daten erhoben und verarbeitet.

Abbildung 1: **Arten von Unternehmensdaten, die für KI-Systeme interessant sind**

Quelle: Eigene Darstellung.

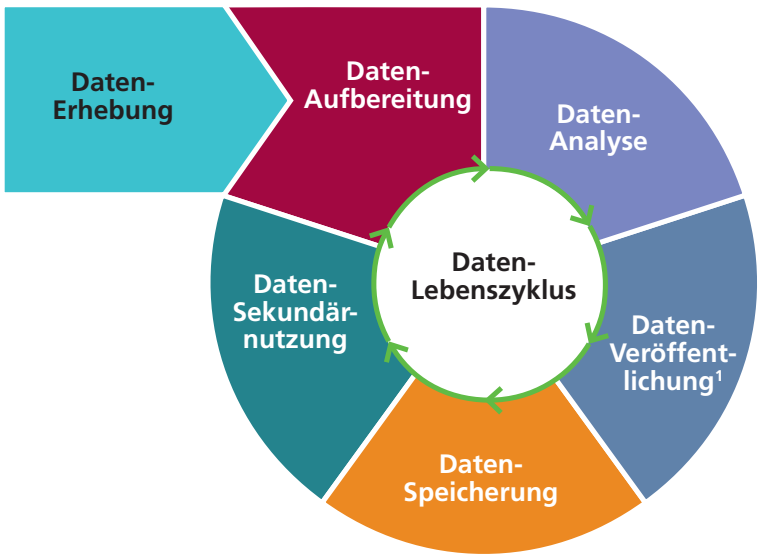
Mit der Datenschutzgrundverordnung (DSGVO) werden im bestehenden EU-Rechtsrahmen in Ermangelung eines KI-spezifischen Datenschutzrechts (Stiftung Datenschutz 2021, S. 11) hohe Anforderungen bei der Datennutzung für KI-Systeme gestellt. Mit der Zielsetzung der Wahrung informationeller Selbstbestimmung von Betroffenen gelten diese für Erhebende, Erwerbende, Nutzende und Weiterverwertende von personenbezogenen Daten – somit alle beteiligten Akteure in der Wertschöpfungskette der Datenökonomie. Es sind sowohl die Trainingsdaten als auch bereits trainierte Modelle zu schützen. Zudem ist es grundsätzlich unerheblich, ob eine etwaige Datennutzung im Gemeinwohlinteresse vollzogen wird. Dabei gelten für die einzelnen Stadien des Daten-Lebenszyklus (siehe Abbildung 2) verschiedene rechtliche Vorgaben, die beachtet werden müssen, wenn Unternehmen Daten für KI-Systeme nutzen möchten.

Datenschutzrecht und seine Auslegung bei der Datennutzung für KI entlang dem Daten-Lebenszyklus

Die Vorgaben bei der Verwendung personenbezogener Daten aus der DSGVO scheinen – auf den ersten Blick – klar formuliert. Jedoch behindert Unsicherheit in der Interpretation und Rechtsauslegung der DSGVO-Vorgaben aktuell den breiten Einsatz von KI (Stiftung Datenschutz 2021; Hoeren & Niehoff 2018). So können die Potenziale des Datenschatzes für KI-Anwendungen unter anderem für das Gemeinwohl zum aktuellen Zeitpunkt nicht vollständig ausgeschöpft werden.

Anwendende von KI-Systemen erheben Daten in der Regel, um in der Phase der Datenanalyse das KI-Modell auf den Daten anzuwenden und den daraus resultierenden Vorschlag durchzuführen. Ein Beispiel ist die Eingabe eines Navigationsziels durch eine Autofahrerin bzw. einen Autofahrer in eine Navigations-App, die daraus die schnellste Ankunftszeit durch einen KI-basierten Routenoptimierer ermittelt. Entwickelnde von KI-Systemen erheben Daten in der Regel, um sie in der Phase der Datenaufbereitung so aufzubereiten, dass sie als Trainingsdatensatz anschließend für das Training des KI-Systems verwendet werden können.

Abbildung 2: **Phasen des Daten-Lebenszyklus**



1 Datenveröffentlichung optional und in vielen Fällen nicht nötig, möglich oder erlaubt (z. B. aufgrund ärztlicher Schweigepflicht)
Quelle: Eigene Darstellung.

So können Daten aufbereitet werden, um ein System zu entwickeln, das Fahrzeiten unter Einfluss von Staus, Baustellen usw. voraussagt und optimiert. Das mit Daten trainierte KI-Modell und auch die Trainingsdaten können gespeichert werden, um sie im Rahmen einer Sekundärnutzung für weitere Trainingsiterationen des spezifischen KI-Systems oder für das Training verwandter KI-Systeme nutzen zu können. Je nach Nutzungsinteresse sind für KI-Anwendende und KI-Entwickelnde also verschiedene Phasen des Daten-Lebenszyklus relevant. Im Hinblick auf datenschutzrechtliche Vorgaben an die Datenverwendung erfordert dies eine umfangreiche und detaillierte Analyse einzelner Phasen im Daten-Lebenszyklus.

Schon bei der **Datenerhebung** bzw. **-generierung** muss – wenn es sich um personenbezogene Daten handelt (Art. 4 Nr. 1 DSGVO) – eine Einwilligung (Art. 4 Nr. 11 DSGVO) der betroffenen Personen (*Data Subjects*)[1] eingeholt werden, die über die Daten identifizierbar sind (sog. Einwilligungstatbestand). Im Sinne einer weiteren Auslegung ist es dabei unerheblich, ob der Bezug auf die Person direkt (Personenbezogenheit) oder indirekt (Personenbeziehbarkeit) herstellbar ist (siehe zur Unterscheidung Kurzinfo)[2].

Durch die Datenerhebenden ist sicherzustellen, dass Betroffene ausreichend darüber informiert sind, wie und für welchen Zweck ihre Daten verwendet werden (Grundsatz der Informiertheit) und dass die erhobenen Daten ausschließlich für den entsprechenden Zweck eingesetzt werden (Grundsatz der Zweckbindung: Art. 5 Abs. 1 lit. b) DSGVO). Die rechtlichen Vorgaben schreiben also fest, unter welchen Bedingungen die Erfassung von personenbezogenen Daten wie zu erfolgen hat – ihre praktische Umsetzung und damit Auslegung ist hingegen weniger eindeutig.

KURZINFO

Personenbezogene Daten

Unter personenbezogenen Daten werden „alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (betroffene Person) beziehen" (vgl. BMJV 2018a), verstanden. Es sind also Daten, die direkt mit einer Person verknüpft sind, wie z. B.:

- LEASYNG-Anwendungsszenario: LEASYNG (Unternehmen) speichert ab, dass Carla (Name, Telefonnummer, Geburtsdatum, Staatsangehörigkeit) mit ihrer Kreditkarte für 24 Stunden ein Auto gemietet hat.
- learn.digital-Anwendungsszenario: Für die Analyse der altersspezifischen App-Nutzung sind die Daten zum Alter der Nutzenden relevant.

Personenbeziehbare Daten

Diese Daten können einer natürlichen Person zugeordnet werden. Es sind also Daten ohne direkten Personenbezug, der sich aber aus ihnen indirekt herleiten lässt, wie z. B.:

- LEASYNG-Anwendungsszenario: Carla gibt als Startadresse für den Reiseassistenten im LEASYNG-Auto Ort, Straße und Hausnummer ein. Dort befindet sich nur das Einfamilienhaus von Carlas Eltern.
- vAltality-Anwendungsszenario: Ein Krebsatlas von vAltality zeigt in einer Karte die Häufigkeit von Kehlkopferkrankungen auf. Ein Ort mit weniger als 20 Einwohnern, in dem nur Anton eine Kehlkopferkrankung hat, wird gelistet.

1 Wenn im Folgenden „Betroffene" erwähnt werden, sind damit *Data Subjects* gemeint.
2 Wenn im Folgenden „personenbezogene Daten" erwähnt werden, werden diese immer im breiten Rechtsverständnis der DSGVO verstanden, beinhalten also sowohl personenbeziehbare als auch -bezogene Daten.

So ist unklar, wie granular die Einwilligung sein muss, welcher Grad an Informiertheit im Einzelfall ausreicht und wie die Zweckbindung interpretiert werden kann (siehe vAltality-Anwendungsszenario, Szenario 1). Selbst wenn personenbezogene Daten explizit für Forschungszwecke im Gemeinwohlinteresse erhoben werden sollen, ist „*broad consent*", also die breite Einwilligungserklärung der Betroffenen (*Data Subjects*) für die pseudonymisierte Verwendung ihrer Daten für Forschungszwecke, rechtlich nur eingeschränkt möglich (siehe vAltality-Anwendungsszenario, Szenario 1).

Zusammengefasst heißt das: Bei der Umsetzung der Datenschutzvorgaben für die Datenerhebung besteht für Unternehmen erhebliche Rechtsunsicherheit bei der Einholung von Einwilligungserklärungen der Betroffenen hinsichtlich erhobener Datenmengen. Da häufig zum Zeitpunkt der Datensammlung noch nicht klar ist, welche Features (konkrete Datenpunkte) am Ende genutzt werden sollen, gehen in der Praxis viele KI-anwendende Unternehmen eher dazu über, zuerst alle messbaren Features zu sammeln und erst später zu entscheiden, welche benötigt werden. So muss für eine Vielzahl von Datenpunkten hoher Aufwand bei der Datenerhebung betrieben werden. Dies erzeugt (Zusatz-)Kosten für die Unternehmen und kann die Datennutzung für KI-Systeme unattraktiv machen oder zumindest erheblich verteuern.

Wenn personenbezogene Daten bereits erhoben wurden, gilt für die **Datenaufbereitung**, also Datenprozessierung, der Grundsatz der Datenminimierung (Art. 5 Abs.1 lit. c) DSGVO). Dieser besagt, bezogen auf KI-Systeme, dass nur diejenigen personenbezogenen Daten verarbeitet werden dürfen, die für das Training bzw. die Anwendung des KI-Systems benötigt werden (Grundsatz der Erforderlichkeit). In der Praxis könnte dies eine künstliche Verknappung und Beschneidung von Datensätzen bedeuten, die eine Sekundärnutzung häufig unmöglich machen und auch die Potenziale von KI-Systemen aufgrund eingeschränkter Datenverfügbarkeit einschränken können (Gausling 2020, S. 13). Andererseits wird aus Datenschutzperspektive mit diesem „scharfen Schwert" die Wahrung informationeller Selbstbestimmung sichergestellt, da die Datenprozessierung gezielt auf den Zweck der Datenerhebung abgestimmt wird (siehe learn.digital-Anwendungsszenario, Szenario 1). In diesem Zusammenhang ist grundsätzlich je nach Anwendungskontext für das Training von KI-Systemen auch von Interesse, personenbezogene Daten mit nicht-personenbezogenen, statistischen Daten zu ersetzen: So könnte für die KI-unterstützte Kapazitätsallokation in Krankenhäusern eine statistische Auswertung der lokalen Verteilung von Herz-Kreislauf-Erkrankungen angewandt werden. Ähnliches ist auch für die Kapazitätsplanung von Sitzplätzen auf einer bestimmten Bahnstrecke denkbar. Auch könnten statistische monatliche Krankenstände in Großunternehmen für die Kalkulation von Kantinenessen angewandt werden. Es empfiehlt sich also eine sorgfältige Abwägung, ob personenbezogene Daten durch statistische Daten für das Training von KI-Systemen ersetzt werden könnten. Ist dies der Fall, sollten personenbezogene Daten grundsätzlich durch statistische Daten ersetzt werden, wenn sie vorhanden sind und es zumutbar ist, sie zu beschaffen. So werden mögliche Datenschutzrisiken by Design mitigiert.

Bei der **Datenanalyse**, also der konkreten Nutzung von Daten für das Training bzw. die Anwendung von KI-Systemen, sieht die DSGVO als Schutzinstrument für Betroffene mit den Artikeln 13 bis 15 besondere Auskunftspflichten für Datenprozessierende vor, sofern Entscheidungen automatisiert getroffen werden (Hoeren & Niehoff 2018). Denn grundsätzlich ist es verboten, automatisierte Entscheidungen auf Grundlage der personenbezogenen Daten der datengebenden Betroffenen zu treffen (Art. 22 Nr. 1 DSGVO). Die DSGVO stellt strenge Auflagen für Ausnahmen von diesem Verbot auf: So stehen datennutzende Unternehmen in der Pflicht, Betroffenen im Fall von automatisierten Entscheidungen „aussagekräftige Informationen über die involvierte Logik" zum Zeitpunkt der Datenerhebung mitzuteilen (Art. 13 Abs. 2 lit. f) und Art. 14 Abs. 2 lit. g) DSGVO). Denn die Betroffenen haben ein Auskunftsrecht (Art. 15 Abs. 1 lit. h) DSGVO) vor oder auch nach der Einwilligung zum Datenteilen und ein Widerspruchsrecht gegen ihre Datenverwendung (Art. 21 DSGVO), um gegen automatisierte Entscheidungen vorgehen zu können. Diese Informationspflichten und Auskunfts-

rechte erzeugen Interpretationsunsicherheit (Stiftung Datenschutz 2021) – vor allem im Hinblick darauf, ob KI-Systeme immer unter die Rechtsdefinition „automatisierte Entscheidungen" fallen (Hoeren & Niehoff 2018, S. 53). Unklar bleibt auch, was „aussagekräftige Informationen über die involvierte Logik" eines KI-Systems sind (siehe learn.digital-Anwendungsszenario, Szenario 2): Ist ihre Funktionsweise offenzulegen? Müssen nur ihre Ergebnisse oder gar das komplette Modell offengelegt werden und wie ist das angesichts des Black-Box-Charakters von Deep-Learning-Modellen überhaupt darstellbar? In der Praxis hat sich derweil etabliert, dass Datennutzende die vom KI-System getroffene „automatisierte Entscheidung so erklärbar präsentieren, dass sie für die Betroffenen überprüfbar und nachvollziehbar ist" (Hoeren & Niehoff 2018, S. 60). Aber auch das bleibt unkonkret in Bezug darauf, welcher Grad und welche Form von Erklärbarkeit ausreicht.

Angesichts hoher Sanktionsandrohungen – bei Verstößen gegen die Informationspflicht drohen Geldbußen bis zu 20 Millionen Euro bzw. bis zu vier Prozent des weltweit erzielten Jahresumsatzes (Art. 83 DSGVO) – sind die Unsicherheiten in der Rechtsauslegung in der Phase der Datenanalyse für KI-Systeme also gewichtige Hemmfaktoren für ihren Einsatz. Außerdem schafft die Datennutzung für KI-Anwendungen eine neue Qualität im Hinblick auf den Datenschutz, da über die Dauer der KI-Anwendung die Analyseergebnisse eines Algorithmus immer genauer werden (sollten), da er mit immer mehr Trainingsdaten zunehmend verbessert wird. Mit steigendem Training könnte ein KI-Modell also Personenbeziehbarkeit über Identifizierbarkeit herstellen und so, obwohl ex ante keine personenbezogenen Daten erhoben wurden, plötzlich über diese Daten eine Person identifizierbar machen. Auch dieses Szenario ist derzeit regulatorisch nicht aufgefangen. Darüber hinaus sind weitere Auflagen zu beachten: So sind Datenverarbeitende verpflichtet, eine Datenschutzfolgeabschätzung zu vollziehen (Art. 35 Abs. 4 lit. a) DSGVO) und ihrerseits eine Datenschutzbeauftragte bzw. einen Datenschutzbeauftragten zu benennen (Art. 37 DSGVO).

Auch die **Datenveröffentlichung** unterliegt datenschutzrechtlichen Vorgaben und erfordert neben der Einhaltung des Zweckbindungsgrundsatzes (Art. 5 Abs. 1 lit. b) DSGVO) auch einen Rechtfertigungsgrund (Art. 6 Abs. 1 DSGVO) bzw. eine ausdrückliche und jederzeit widerrufbare Einwilligungserklärung der Betroffenen zur Datenveröffentlichung (Art. 6 Abs. 1 lit. a) DSGVO). Für KI-Systeme ist die Datenveröffentlichung weniger relevant, weil Daten bei KI-Systemen in der Regel als Input für das Training von ML-Modellen oder für die Anwendung von KI-Systemen verwendet und nur selten veröffentlicht werden.

Hingegen sind Auflagen an die **Datenspeicherung** für die Entwicklung oder Anwendung von KI-Systemen von Bedeutung. Jede Datenspeicherung stellt bereits eine Datenverarbeitung im Sinne der DSGVO dar und unterliegt somit strengen Rechtsvorschriften (Art. 9 DSGVO). Ausgehend vom Datenminimierungsprinzip erfordert jede Speicherung personenbezogener Daten eine Rechtsgrundlage, also einen triftigen Grund, der unter anderem auch in einem hohen öffentlichen Interesse liegen kann (siehe learn.digital-Anwendungsszenario, Szenario 3). Die Datenspeicherung unterliegt zudem gesetzlichen Aufbewahrungs- und Löschfristen (Art. 17 DSGVO), die sich jedoch je nach Anwendungsfall unterscheiden: So ist im Gastgewerbe eine Speicherung der Übernachtungs- und Registrierungsdaten zwingend erforderlich, im Luftfahrtsektor besteht die Auflage, Passagierdaten für sechs Monate zu speichern, und auch bei der Anmietung eines Mietfahrzeugs ist die Speicherung personenbezogener Daten aufgrund etwaiger Verstöße von Mietenden gegen die Straßenverkehrsordnung für einen längeren Zeitraum erforderlich. Auch in weiteren Bereichen gelten unterschiedliche Aufbewahrungsfristen. Diese fehlende Einheitlichkeit der Aufbewahrungsfristen erzeugt Unsicherheit. Des Weiteren gilt, dass die Identifizierung betroffener Personen aus den gespeicherten Daten nur solange möglich sein darf, wie es der Zweck erfordert, für den die Daten verarbeitet wurden (Art. 5 Abs. 1 lit. e) DSGVO). Wiederum entstehen in der Interpretation dieser Vorgaben Unsicherheiten und Aufwände für KI-nutzende und -entwickelnde Unternehmen: Denn es bleibt unklar, worin der triftige Grund einer Datenspeicherung konkret liegt. Die Vorgabe der zweckgebundenen Identifizierbarkeit erfordert entweder eine Ano-

nymisierung des Datensatzes oder seine Löschung. Dies erzeugt Kosten für das Datenmanagement und macht eine Weiterverwendung entsprechender Daten wenig attraktiv, auch wenn sie gerade für KI-Systeme sinnvoll wären.

Die **Datensekundärnutzung** – also, wenn Daten ursprünglich für einen bestimmten Zweck erhoben wurden und nun abermals, aber für einen anderen Zweck genutzt werden – tritt grundsätzlich häufig auf. Bei KI-Systemen ist sie besonders verbreitet, weil die Verknüpfung unterschiedlicher Daten für das Training und die (Weiter-)Entwicklung von KI-Systemen besonders wichtig ist: Im Rahmen von ML-Lernprozessen kommt es häufig vor, dass neue Muster im ML-Modell erkannt werden, die das KI-System für andere weitere Zwecke interessant machen als für den Zweck, für den ursprünglich die Einwilligung der Betroffenen eingeholt wurde. Nur wenn diese Zweckänderung als „logischer nächster Schritt" gilt, sofern also der ursprüngliche Zweck mit dem neuen Zweck kompatibel ist, muss keine abermalige Einholung der Einwilligung erfolgen. Diesen Nachweis muss die datenprozessierende Entität führen (Art. 6 Abs. 4 DSGVO), was wiederum Interpretationsspielraum eröffnet und somit Unsicherheiten in der Rechtsauslegung und Aufwand erzeugt. In der Praxis muss bei der Datensekundärnutzung und der nachträglichen Zweckänderung also die nachträgliche Einwilligung der datengebenden Betroffenen eingeholt werden, weil die Datensekundärnutzung in der Regel nicht mehr dem Zweck der Datenersterhebung dient, dem die betroffene Person zugestimmt hatte (siehe LEASYNG-Anwendungsszenario, Szenario 1 und 2). Trotz vermeintlich klarer Regelung ist dies daher in der Praxis mit unwägbaren Problemen verbunden. So ist die Re-Identifizierung von Betroffenen für KI-Entwickelnde aufgrund des sehr hohen Aufwands kaum lohnenswert: in ML-Modellen werden personenbezogene Daten häufig zu neuen Datenpunkten in Trainingsdatensätzen verknüpft und müssten für die Re-Identifizierung erst wieder entwirrt werden (Gausling 2020, S. 42f). Eine Alternative zur nachträglichen Einholung der Einwilligung ist die individuelle Interessenabwägung (siehe LEASYNG-Anwendungsszenario, Szenario 1): Wenn Datensekundärnutzende in sorgfältiger Einzelfallabwägung begründet darstellen können, dass das Gemeinwohlinteresse einer Datenverwertung dem Schutzbedürfnis der persönlichen Daten der Betroffenen höhergestellt ist, können die Daten auch für andere Zwecke als ursprünglich eingewilligt verwendet werden (Art. 6 Abs. 1 S. 1 lit. f) DSGVO). Auch hier besteht in der Umsetzung aber Unklarheit, wann das Gemeinwohlinteresse überwiegt, wie dieses überhaupt definierbar und operationalisierbar ist und anhand welcher Indikatoren die Abwägung vollzogen werden kann. Festzuhalten ist: Auch bei der Datensekundärnutzung besteht Unsicherheit in der Rechtsauslegung, die oftmals eine aufwändige Neuerhebung von Daten sinnvoller macht. Aus Sicht des Datenmanagements ist dies aber nicht nur teuer, sondern auch ineffektiv.

Die Datensekundärnutzung kann nicht nur durch das Unternehmen selbst, das die Daten ursprünglich erhoben hat, vollzogen werden, sondern auch über Dritte. Diese Akteure kommen ins Spiel, wenn entweder das ersterhebende Unternehmen eine **Datenweitergabe** an einen anderen Akteur vollzieht oder diese per **Datenerwerb** in den Datenbesitz kommen (siehe vAltality-Anwendungsszenario, Szenario 2). Grundsätzlich ist ein Datenerwerb durch Dritte auf schuldrechtlicher Basis möglich. Beim Vollzug des Datenerwerbs, wenn Dritte also in den Besitz personenbezogener Daten von den Datenerhebenden kommen, ist erneut das Datenschutzrecht zu beachten. Dabei müssen diese entweder wiederum die nachträgliche Einwilligung für die Datennutzung von den betroffenen datengebenden Personen einholen oder eine individuelle Interessenabwägung vollziehen. Es gelten hier ähnliche Problemstellungen wie bei der Datensekundärnutzung: Interpretationsunsicherheit entsteht in der Umsetzung des Datenschutzrechts, was die Datennutzung für KI-Systeme hemmt.

Status quo: Interpretationsoffenheit hemmt – auch gemeinwohlorientierte – Datennutzung für KI

Insgesamt bleibt festzuhalten: Nicht das Recht an sich, sondern die Unsicherheit durch uneinheitliche Rechtsauslegung hemmt die Hebung des Datenschatzes für den KI-Einsatz. Obwohl die rechtlichen Vorgaben klar und eindeutig im Wortlaut erscheinen, erzeugen sie in der unternehmerischen Praxis erheblichen Interpretationsspielraum verbunden mit rechtlichen und/oder finanziellen Konsequenzen (siehe Abbildung 3). Dies beginnt bereits aufgrund des unklaren Anwendungsbereichs der DSGVO mit der Rechtsdefinition der Personenbezogenheit, die in der Praxis dazu führt, dass Daten im Zweifel als personenbezogen geführt werden (Stiftung Datenschutz 2021, S. 53), und erstreckt sich bis hin zu Interpretationsspielräumen bei der Einzelfallabwägung und Zweckbindung bei der Nutzung personenbezogener Daten. Die Nutzbarmachung von Daten für KI-Systeme ist für Unternehmen also nicht nur finanziell, sondern auch zeitlich, personell und rechtlich

Abbildung 3: Datenschutzauflagen an KI-Systeme und Interpretation in der Anwendung über den Daten-Lebenszyklus

PHASE IM DATEN-LEBENSZYKLUS	DATENSCHUTZ – RECHTLICHE VORGABEN	UNSICHERHEIT IN DER ANWENDUNG
DATEN-ERHEBUNG	· Einwilligung der betroffenen Personen (Art. 4 Nr. 11 DSGVO) · Informiertheit der betroffenen Personen (Art. 6 Abs. 1 lit. a) DSGVO) · Zweckbindung der Datenerhebung (Art. 5 Abs. 1 lit. b) DSGVO)	· Wie granular/detailliert muss die Einwilligung sein? · Welcher Grad an Informiertheit ist erforderlich? · Wie soll Zweckbindung interpretiert werden?
DATEN-AUFBEREITUNG	· Grundsatz der Datenminimierung (Art. 5 Abs.1 lit. c) DSGVO)	· Datenminimierung erschwert Sekundärnutzung der Datensätze
DATEN-ANALYSE	Bei automatisierten Entscheidungen gilt · Mitteilungspflicht nach Art. 23 Abs. 2 lit. f) und Art. 14 Abs. 2 lit. g) DSGVO, sobald Personen von automatisierten Entscheidungen nach Artikel 22 DSGVO betroffen sind · Auskunftsrecht nach Art. 15 Abs. 1 lit. h) DSGVO · Widerspruchsrecht nach Art. 21 Abs. 1 DSGVO gegen Datenverwendung sicherstellen	· Ab wann fällt ein KI-System unter die Rechtsdefinition „automatisierte Entscheidungen"?
DATEN-VERÖFFENTLICHUNG	· Zweckbindungsgrundsatz (Art. 5 Abs. 1 lit. b) DSGVO) · Rechtmäßigkeit der Verarbeitung (Art. 6 Abs. 1 DSGVO) · Jederzeit widerrufbare Einwilligungserklärung der Betroffenen (Art. 6 Abs. 1 lit.a) DSGVO)	
DATEN-SPEICHERUNG	· Aufbewahrungs-/Löschfristen (Art. 17 DSGVO) · Identifizierung betroffener Personen nur so lange erlaubt, wie es ursprünglicher Datenverarbeitungszweck erfordert	· Interpretationsspielraum in der Rechtsauslegung · Uneinheitlichkeit bei Aufbewahrungs-/Löschfristen · Schwierige Anonymisierbarkeit von Randgruppen · Semantische Segmentation schafft Personenbeziehbarkeit
DATEN-SEKUNDÄRNUTZUNG	· Datensekundärnutzung ohne Einwilligung der Betroffenen erlaubt, wenn sie als „logischer nächster Schritt" gilt oder im Gemeinwohlinteresse liegt (Art. 6 Abs. 1 S. 1 lit. f) DSGVO)	· „logischer nächster Schritt" bietet Interpretationsspielraum · Gemeinwohlinteresse ist nicht eindeutig definiert

teuer sowie risikobehaftet. Viele von ihnen nehmen deshalb teilweise oder ganz Abstand vom Einsatz bzw. der (Weiter-)Entwicklung von KI-Systemen – auch dann, wenn sich Potenziale für das Gemeinwohl bieten.

Diese Problemstellungen verdeutlichen vor dem Hintergrund ebenjener Gemeinwohlpotenziale das Erfordernis, eine flexibilisierte Datennutzung für KI-Systeme im Gemeinwohlinteresse mit dem Datenschutz in Symbiose zu bringen, anstatt sie als gegenseitige Hemmfaktoren im „Spannungsfeld" (Stiftung Datenschutz 2021; Gausling 2020) zu belassen. Die Auflösung des Spannungsfelds erfordert sowohl die Ausnutzung und Weiterentwicklung technischer Möglichkeiten zur indirekten Minimierung der Unsicherheiten in der Rechtsauslegung (siehe Kapitel 3) als auch ihre direkte Eliminierung durch gezieltere, anwendungsspezifische und technikneutrale Gesetzgebung (siehe Kapitel 4).

3 Technische Ansätze zur datenschutzwahrenden gemeinwohlorientierten Datennutzung

Unterschiedliche technische Ansätze zur datenschutzwahrenden gemeinwohlorientierten Datennutzung für verschiedene Kommunikationstypologien sind denkbar, um den geschilderten Problemlagen entgegenzutreten. Insbesondere mit einem weit verbreiteten Einsatz von Ansätzen datenschutzwahrenden maschinellen Lernens (*Privacy-Preserving Machine Learning*; kurz: PPML) könnte die datenbasierte Entwicklung von KI-Systemen und damit final auch die KI-Anwendung erleichtert werden. Diese PPML-Ansätze versprechen, Datenschutz qua Design sicherzustellen. Zugleich stellen sie damit auch einen möglichen Alternativweg für eine flexibilisierte Datennutzung im Allgemeinwohl im Vergleich zu den hohen Ex-ante-Anforderungen an die Datenverarbeitung aus der DSGVO dar. Auch andere, nicht direkt KI-Modell-bezogene technische Maßnahmen, wie der Einsatz von Personal Information Management-Systemen (PIMS) oder Datentreuhändern, könnten eine gemeinwohlorientierte flexibilisierte Datennutzung für KI-Systeme erleichtern und gleichzeitig die Souveränität von datengebenden Personen in der Datenökonomie stärken.

Anonymisierungs- und Pseudonymisierungsmaßnahmen

Schon bei der **Datenerhebung/-generierung** kann über verschiedene technische Maßnahmen die Personenbezogenheit von Daten eliminiert werden, wodurch die DSGVO nicht mehr greifen würde. Auf diese Weise könnten die Daten auch flexibel für weitere Zwecke eingesetzt werden. Die Krux ist jedoch dabei, solche Verfahren zu finden, die Datenschutzkonformität erlauben, ohne dass die Datenqualität für die Anwendung oder Entwicklung von KI-Systemen letztlich darunter leidet. Zudem müssen diese Ansätze hinreichend zukunftssicher und robust im Hinblick auf künftige mögliche Überwindungen von Datenschutzkonformität bei technischem Fortschritt sein (Leopoldina et al. 2018, S. 50). Folgende technische Ansätze und Verfahren in der Phase der Datenerhebung bestehen bereits:

- **Anonymisierung und Pseudonymisierung** (siehe learn.digital-Anwendungsszenario, Szenario 1): Während nach DSGVO bei einer Pseudonymisierung weiterhin Personenbezogenheit von Daten vorliegt (Schwartmann & Weiß 2017), verspricht die Anonymisierung in der Theorie die komplette Eliminierung derselbigen (Schwartmann & Weiß 2017, S. 12) und würde so Datennutzungsmöglichkeiten flexibilisieren. Bei einer Anonymisierung können Techniken wie Verrauschung (Hinzufügen von Noise) oder Vergröberung eingesetzt werden, um die Identifizierbarkeit von Personen zu verhindern. Beide Verfahren reduzieren jedoch die Datenqualität signifikant (Gausling 2020, S. 19). Nach geltender Rechtsauffassung stellt die Anonymisierung ihrerseits eine der DSGVO unterliegende Datenverarbeitung dar, weshalb das Datenschutzrecht weiter greift. Nur in dem Fall der Anonymisierung unmittelbar bei der Erhebung der Daten, also vor ihrer Speicherung, liegen zu keinem Zeitpunkt personenbezogene Daten vor (Stiftung Datenschutz, S. 23). Da Personenbezogenheit als Rechtsbegriff zudem nicht definiert ist, ist es für die Unternehmen unklar, inwieweit sie die Daten bis zur Eliminierung der Personenbezogenheit anonymisieren müssen. Es existiert kein technischer Standard, bei dessen Einhaltung vermutet wird, dass Daten anonymisiert sind. Eine Anonymisierung gerät deshalb häufig zur „*mission impossible*" (Stiftung Datenschutz 2021, S. 24).

KURZINFO

Pseudonymisierung

… verändert personenbezogene Daten so, dass sie „ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können" (Art. 4 Nr. 5 DSGVO). Da es aber aufgrund ebenjener zusätzlichen Informationen weiterhin möglich ist, Personen zu identifizieren, eliminiert die Pseudonymisierung laut DSGVO die Personenbezogenheit von Daten nicht.

- **Datensynthese** (siehe LEASYNG-Anwendungsszenario, Szenario 1): Mit der Datensynthese wird nicht versucht, die Personenbezogenheit im Originaldatensatz zu eliminieren. Stattdessen werden synthetische Datensätze ohne Personenbezug nach Vorlage der für den jeweiligen Anwendungsfall wichtigsten statistischen Eigenschaften des personenbezogenen Originaldatensatzes künstlich erzeugt. Die Datensynthese ist somit nicht nur für KI-Anwendende, sondern auch für KI-Entwickelnde für die Synthetisierung von Trainingsdatensätzen interessant. Die populärste Methode ist der Einsatz von *Generative Adversarial Networks* (GANs). Ein GAN besteht aus zwei konkurrierenden neuronalen Netzwerken, wobei das eine, der Generator, künstliche, aber möglichst realistische Daten erzeugt.

  Das zweite Netzwerk, der Diskriminator, bewertet, ob Daten künstlich erzeugt wurden oder real sind. Mit fortschreitendem Training werden beide Netzwerke immer besser. So kann der Generator beispielsweise lernen, realistische Gesichter zu erzeugen. Grundsätzlich erlaubt die Datensynthese eine komplette Eliminierung des Personenbezugs und erfährt deshalb unter anderem von der Datenethikkommission der Deutschen Bundesregierung Unterstützung (Datenethikkommission der Bundesregierung, Gutachten, Oktober 2019, Ziff. 4.2.3, S. 132). Dies ist bedeutend, da der Einsatz von GANs als wichtige Methode angesehen wird, um die Personenbezogenheit aus unstrukturierten Daten (v. a. Videos, Bildern) herauszufiltern, von denen die Betroffenen oft gar nicht wissen, dass diese über sie – häufig im Internet – verfügbar sind und potenziell für das Training von KI-Systemen verwendet werden könnten[3] (Liu et al. 2021, S. 26). Durch Datensynthese erzeugte künstliche Datensätze fallen nach geltender Rechtsauffassung nicht unter die DSGVO und können sogar öffentlich geteilt werden, was sie für eine flexibilisierte Datennutzung für KI-Systeme attraktiv macht. Die Praktikabilität der Datensynthese ist allerdings begrenzt. Denn gerade im B2C-Bereich sind viele KI-Systeme für das Training ihrer ML-Modelle auf statistische Merkmale angewiesen, die eine persönliche Identifizierbarkeit ermöglichen – und somit wiederum unter die DSGVO fielen. Allein der Erzeugungsaufwand dieser künstlichen Datensätze ist hoch. Zudem bleibt immer noch das Risiko, dass der synthetische Datensatz den originalen Datensatz nicht hinreichend repräsentiert. Trotzdem vermag die Datensynthese in bestimmten Anwendungsbereichen hilfreich zu sein, wenn auch nicht als Allheilmittel für die gesamte KI-Wirtschaft (Stiftung Datenschutz 2021, S. 25).

  Als Subform der Datensynthese kann *Deep Natural Anonymization* angesehen werden. Hier kann die Personenbezogenheit der Daten entfernt werden, wobei gleichzeitig andere für die Datenqualität relevante Attribute erhalten bleiben. So kann beispielsweise in einem Bild ein

3 Laut Vorhersagen der International Data Corporation werden 2025 unstrukturierte Daten 80 Prozent weltweiter Daten ausmachen.

Gesicht durch ein künstlich generiertes Gesicht ersetzt werden, indes die Blickrichtung als relevantes Attribut bestehen bleibt. Bei diesem Ansatz wird nicht ein kompletter Datensatz künstlich erzeugt, sondern nur die Teile eines Datensatzes, die Personenbezogenheit aufweisen. Diese Methode kommt beispielsweise in der kamerabasierten Fahrgastallokation im öffentlichen Nahverkehr zur Unkenntlichmachung von Gesichtern zum Einsatz. Der Synthetisierungsaufwand wird bei dieser Methode zwar erheblich reduziert, dennoch bestehen auch für *Deep Natural Anonymization* Unsicherheiten, ob eben jene Anonymisierung den Personenbezug rechtssicher eliminiert.

- **Differential Privacy (DP)** (siehe vAltality-Anwendungsszenario, Szenario 3): Bei diesem Verfahren werden Daten so „verrauscht", dass in der Grundgesamtheit zwar immer noch korrekte statistische Aussagen getroffen werden können, ein einzelner Datenpunkt – also beispielweise der Nachname einer einzelnen Person – aber nicht mehr identifizierbar ist. Besonders geeignet ist Differential Privacy deshalb für die Datenschutzwahrung von Trainingsdaten (*Training Data Privacy*). Die Verhinderung der Aussonderung (singling out) über Verrauschung ist für die Anonymisierung medizinischer Forschungsdaten interessant und wird bereits im Big-Data-Kontext eingesetzt (Torkzadehmahani et al. 2020). In der praktischen Umsetzung für die Entwicklung und Anwendung von KI-Systemen ist dieses Verfahren aber nicht unproblematisch (Leopoldina et al. 2018, S. 51). Bei der Entwicklung von KI-Systemen ist die dynamische Erhebung von Daten sowie ihre Veränderung (z. B. Verknüpfung) bei der Datenprozessierung häufig erforderlich. Eine schnelle Aufhebung des Personenbezugs ist aufgrund des hohen Verrauschungsaufwands bei Differential Privacy schwer möglich, weshalb die Technik nicht für KI-Modelle geeignet ist, die in Echtzeit agieren (z. B. Echtzeit-Vorhersagen). Auch bei sehr kleinen Datensätzen ist dieses Verfahren nicht hilfreich, da die Modellgenauigkeit aufgrund des mit Datensatzverkleinerung steigenden Verrauschungsaufwands zur Verhinderung der Identifizierbarkeit sinkt. Wenn zudem die Aussagen über individuelle Datenpunkte weiter stimmen müssen, da diese sonst einer falschen Entscheidung des ML-Modells unterliegen, kann Differential Privacy kontraproduktiv sein (siehe vAltality-Anwendungsszenario, Szenario 3). Zudem ist es rechenintensiv und erfordert eine potenziell fehlerbehaftete menschliche Einordnung, welche Datenpunkte im Datensatz die Identifizierbarkeit von Personen ermöglichen würden. Differential Privacy bietet elegante mathematische Garantien zur Wahrung der Privatsphäre; allerdings erfordert dies ein „*Privacy budget*", welches in der operativen Produktumgebung leicht überschritten werden kann. Nur in begrenzten Fällen steigert Differential Privacy Rechtssicherheit und Skalierbarkeit zuverlässig.

Verschlüsselungsmaßnahmen

Wie beschrieben geht mit den oben genannten Anonymisierungstechniken der Personenbezogenheit in Datensätzen auch eine Reduzierung der Datenqualität einher, die solche technischen Maßnahmen unattraktiv für die Anwendung oder Entwicklung bestimmter ML-Modelle machen kann. Kryptografische datenschutzwahrende Maßnahmen setzen deshalb bei der **Datenprozessierung**, also der Datenaufbereitung, an, um die Personenbezogenheit in Datensätzen zu eliminieren bzw. den Datenschutz zu gewährleisten. Zu diesem Zeitpunkt im Datenlebenszyklus ist die Erhebung der personenbezogenen Daten im klassischen Prozess schon vollzogen, wodurch die DSGVO bereits greift. Danach könnten die im Folgenden vorgestellten Verfahren als mögliche Unterstützungsmaßnahmen zur Wahrung des Datenschutzes nach Art. 25 DSGVO gelten. Stellt sich nach eingehender Prüfung im praktischen Einsatz heraus, dass diese technischen Maßnah-

men den Datenschutz wirklich wahren und gleichzeitig die Datennutzung flexibilisieren, wäre durchaus über ihre sichere rechtliche Anerkennung nachzudenken (siehe Gestaltungsoptionen in Kapitel 4). Folgende kryptografische Techniken sind für die Phase der Datenprozessierung von Bedeutung:

- **Homomorphic Encryption** (siehe learn.digital-Anwendungsszenario, Szenario 4): Diese Verschlüsselungsmethode kann entweder initial auf einem Rohdatensatz durchgeführt werden, um die Rohdaten der Nutzenden zu schützen (*Input Privacy*), oder auch auf dem Output eines KI-Modells ausgeführt werden (*Output Privacy*): Dabei werden noch bei den datengebenden Betroffenen Datenpunkte mit Personenbezug verschlüsselt. Anschließend wird der Datensatz samt der verschlüsselten Datenpunkte an Prozessierende verschickt, die diese auf KI-Systeme anwenden oder mit diesen KI-Modelle trainieren. Diese versenden die Ausgaben der Modelle an die betroffenen, datengebenden Personen zurück, die diese Datensätze wiederum entschlüsseln können (Torkzadehmahani et al. 2020, S. 3). Über den ganzen Prozess hinweg sind die Daten verschlüsselt, der Prozessierende hat so keinen direkten Zugriff auf personenbezogene oder -beziehbare Informationen. In der Praxis ist die Komplexität dieser Verschlüsselung im Hinblick auf die Anforderungen an die Rechenlast ein gewichtiges Hemmnis für die schnelle Datenprozessierung (Liu et al. 2021, S. 12; Torkzadehmahani et al. 2020). Insbesondere das Training von Deep Neural Networks ist daher mit Homomorphic Encryption wenig praktikabel.

- **Secure Multiparty Computation (SMPC)** (siehe learn.digital-Anwendungsszenario, Szenario 4): Als Erweiterung der Verschlüsselung in Mehrparteien-Settings werden bei SMPC im ML-Kontext Dateninteraktionen durch Verschlüsselungsmechanismen und verdeckte Übertragungen so vollzogen, dass Datenprozessierende weder direkten Zugriff auf die Daten noch auf das Modell selbst haben (Liu et al. 2021, S. 12). Damit sollen *Input* und *Output Privacy* gewährleistet werden. SMPC ist dabei datenschutzwahrend, kann aber hohe Kommunikationskosten erzeugen und ist damit ähnlich wie Homomorphic Encryption in vielen Anwendungsbereichen unpraktikabel (Torkzadehmahani et al. 2020); auch wenn einige Ansätze Fortschritte hinsichtlich der Effizienz versprechen und verschiedene Kommunikationstopologien unterschiedlichen Einfluss auf die SMPC-Effizienz haben.

- **Confidential Computing** (siehe LEASYNG-Anwendungsszenario, Szenario 3): Dieser Ansatz schließt eine Verschlüsselungslücke im *Cloud Computing*, wo traditionell zwar die Übertragungswege der Daten in die Cloud (*Data in Transit*) sowie deren Speicherung (*Data at Rest*), nicht aber die Verarbeitung der Daten in der Cloud (*Data in Use*) geschützt sind. Klassischerweise werden die Daten hierfür entschlüsselt. Aus Datenschutzperspektive ist das kritisch, da Datenprozessierende durch eine nicht sichergestellte Ende-zu-Ende-Verschlüsselung ungeschützten Zugriff auf personenbezogene Daten haben. Beim Confidential Computing findet die Verarbeitung der Daten in einer hardwarebasierten, vertrauenswürdigen Ausführungsumgebung (*Trusted Execution Environment*; kurz: TEE) statt. Diese Ausführungsumgebung stellt sicher, dass kein nicht-autorisierter Zugriff auf die Daten während der Verarbeitung möglich ist. Hierdurch wird das Vertrauen in die Wahrung von *Privacy* deutlich erhöht, sofern der TEE-Hersteller selbst vertrauenswürdig ist. Für die Prüfungen der Zugriffsautorisierung und der Echtheit der TEE entstehen auch bei Confidential Computing Kommunikationskosten, selbst wenn diese im Vergleich zu anderen Ansätzen (z. B. SMPC) geringer sind.

Zusammengefasst erlauben diese kryptografischen bzw. isolierenden Ansätze zwar Datenschutzwahrung durch einen Schutz der Vertraulichkeit von Daten, sind aber über den *Ansatz bring data to computation*, also

zentralisiertes Lernen, aufgrund der hohen Rechenlast für die Bereitstellung für das Modelltraining schwächer hinsichtlich Skalierung (Torkzadehmahani et al. 2020), auch wenn Cloud-Anbieter dies durch ihre gute Skalierung der Rechenleistung abmildern können.

Verteiltes maschinelles Lernen und hybride Ansätze

Weitere Maßnahmen in späteren Phasen des Datenlebenszyklus rücken somit in den Fokus: Bei der **Datenanalyse**, also der konkreten Nutzung von Daten für das Training von KI-Modellen, kommt der Einsatz dezentraler ML-Methoden infrage, um Datenschutz und gemeinwohlorientierte Datennutzung produktiv zu kombinieren. Diese dem *Ansatz bring computation to data* folgenden technischen Maßnahmen erfordern grundsätzlich keine Ex-ante-Veränderung personenbezogener Daten, da die Daten lokal bei den Betroffenen verbleiben. Sie versprechen hohe Datenqualität und hohe individuelle Datensouveränität und sind vor allem für KI-Entwickelnde interessant. Zu diesen Ansätzen zählen:

- **Verteiltes maschinelles Lernen** (siehe vAItality-Anwendungsszenario, Szenario 3): Der Ansatz des verteilten maschinellen Lernens liegt darin, das Training von ML-Modellen mit (personenbezogenen) Daten anders als beim zentralisierten Lernen (*bring data to computation*) nicht auf einem zentralen Server der Datenprozessierenden, sondern auf Endgeräten (sog. Clients) der datengebenden Betroffenen zu vollziehen (*bring computation to data*). Somit entfällt das Erfordernis des Teilens personenbezogener Daten mit den Prozessierenden, da Betroffene selbst ihre Daten auf dem eigenen Endgerät für das Training des ML-Modells prozessieren. Betroffene behalten so die Souveränität über ihre (personenbezogenen) Daten. Dieser Ansatz weist daher besonderes Potenzial für datenschutzfreundliche KI auf (Schallbruch et al. 2021). Ansätze verteilten maschinellen Lernens sind aufgrund des fehlenden Erfordernisses des Teilens von Rohdaten durch das dezentrale Training von ML-Modellen nicht nur aus Datenschutz-Gesichtspunkten attraktiv, sondern auch für (1.) das Training von ML-Modellen über ein heterogenes Set an Datenpunkten, (2.) bei einer für einen Server nicht kontrollierbaren Datenmenge oder nicht auflösbaren Verteilung der Daten sowie (3.) für das Echtzeitlernen (Liu et al. 2021, S. 4f.). Dabei bestehen diverse technische Implementierungen verteilten maschinellen Lernens (z. B. *Split Learning, Collaborative Learning, Swarm Learning*), wobei Federated Learning (föderiertes Lernen) die bisher weiteste Verbreitung und stärkste Erforschung erfahren hat (Plattform Lernende Systeme 2022): Ein zentraler Server wählt Clients (Endgeräte) für das Training eines ML-Modells aus, die das ML-Modell von diesem Server herunterladen, das Modell lokal mit dem eigenen Datensatz trainieren und nur die Trainingsergebnisse (*Weights*) zurück an den Server senden. Dieser aggregiert die Trainingsergebnisse aller Clients und aktualisiert damit das ML-Modell, dessen Update Clients für weitere Trainingsrunden zur Verfügung gestellt werden. Unter den vorgestellten Maßnahmen sind Ansätze des verteilten maschinellen Lernens aus Skalierbarkeitsperspektive am besten geeignet, um Datenschutz by Design zu wahren, solange Rückschlüsse auf personenbezogene Daten aus den *Weights* und Modellparametern unmöglich sind (Torkzadehmahani et al. 2020, S. 11). Zu bedenken ist jedoch, dass bei einigen Ansätzen des verteilten maschinellen Lernens wie *Federated Learning* oder *Split Learning* hohe Kommunikationskosten zwischen *Clients* und *Servern* entstehen können (Torkzadehmahani et al. 2020, S. 8), die von der Größe des Modells abhängen. Gerade bei kleinen Modellen und großen Datensätzen scheint verteiltes Lernen performanter als zentralisiertes Lernen zu sein. Zu beachten ist, dass Ansätze des verteilten maschinellen Lernens generell gerade im Hinblick auf den Austausch der Trainingsergebnisse neue Angriffsvektoren öffnen (Liu et al. 2021), die den

Datenschutz und/oder die Modellintegrität bedrohen könnten (z. B. *Backdoor Attacks* auf *Clients* mit *Model Poisoning* oder *Data Poisoning*). Erste rechtliche Bewertungen für Lösungsansätze des verteilten maschinellen Lernens existieren bereits (XayNet & CMS 2020); trotzdem ist weiterhin unklar, ob ihr Einsatz datenschutzkonform ist.

- **Hybride Ansätze verteilten maschinellen Lernens** (siehe learn.digital-Anwendungsszenario, Szenario 4 und vAItality-Anwendungsszenario, Szenario 3): Kombinationen unterschiedlicher PPML-Ansätze versprechen, solche neuen Angriffsvektoren beim verteilten maschinellen Lernen schließen zu können, Performanz zu wahren und gleichzeitig Datenschutz und Sicherheit zu garantieren und so holistischen Schutz der Interaktion mit KI-Modellen herzustellen (*Model Privacy*). Eine Kombination verteilter Lernansätze mit datenschutzwahrenden Maßnahmen des zentralisierten Lernens (z. B. DP, SMPC, HE) ist möglich, wobei sich die Verknüpfung von *Federated Learning* (Skalierbarkeit über verteiltes Lernen) mit *Differential Privacy* in Anwendung und Forschung bisher am besten bewährt hat (Torkzadehmahani et al. 2020; Liu et al. 2021). Zwar reduziert sich die Modellgenauigkeit aufgrund der „Verrauschung" der Trainingsdaten bei diesem hybriden Ansatz stärker als in einer Kombination von *Federated Learning* mit *Secure Multiparty Computation*, die nur den Datenzugang, nicht aber die Daten selbst verschlüsseln. Allerdings verspricht die Kombination aus *Federated Learning* (FL) und *Differential Privacy* im Gegensatz zu den beiden anderen hybriden Modellen, Rechenlast zu minimieren und vor allem Datenschutz verlässlich sicherzustellen, auch wenn der Austausch der Trainingsergebnisse (*Weights*) zwischen Server und Clients weiterhin Ziel von datenschutzrelevanten Angriffen bleiben kann (Torkzadehmahani et al. 2020, S. 10). Um diesen Angriffsvektor zu schließen, wäre eine Ausführung von DP-basiertem *Federated Learning* in *Trusted Execution Environments* (TEEs) zielführend. Denn *Federated Learning* „glättet" bereits Model-Parameter und erschwert hierdurch Angriffe, die versuchen, Informationen über die genutzten Trainingsdaten aus dem Model zu extrahieren. Ferner anonymisiert die Aggregation die Beiträge der einzelnen Teilnehmenden. Sobald die einzelnen Modelle zu einem einzigen Model aggregiert wurden, ist es in der Regel nicht mehr möglich herauszufinden, von wem ein Trainingsbeispiel verwendet wurde, selbst wenn eine Angreiferin oder ein Angreifer es schaffen sollte, aus dem aggregierten Model noch Informationen zu extrahieren. Das Risiko einer Deanonymisierung durch „neugierige" Server (sog. *Honest-but-curious*) oder sogar Server, die vom vereinbarten Protokoll abweichen (sog. *fully malicious*), kann hierbei beispielsweise durch die Nutzung von *Secure Multiparty Computation* (Fereidooni et al. 2021) oder vertrauenswürdige Ausführungsumgebungen *Trusted Execution Environments* (Mo et al. 2021) minimiert werden. Je nach Einsatzgebiet und -zielen müssen beim Einsatz hybrider PPML-Ansätze also Abwägungen hinsichtlich der Wichtigkeit von Modellgenauigkeit, Datenschutz der Rohdaten, Datenschutz der ausgetauschten Trainingsergebnisse und der Skalierbarkeit vollzogen werden.

Vertrauen in KI-Systeme und Akzeptanz von KI-Systemen

Eine gesellschaftlich geforderte Randbedingung zur Akzeptanzsteigerung während der Datenanalyse ist der Ansatz, über eine bessere Erklärung der Funktionsweise und der Outputs von KI-Systemen einerseits Transparenz und andererseits so auch Vertrauen und Robustheit bei Betroffenen zu schaffen – sei es in zentralen oder dezentralen ML-Modellen. Durch diese Stärkung der Erklärbarkeit von KI-Systemen könnte eine Flexibilisierung der Datennutzung besonders dann angedacht werden, wenn eindeutiges Gemeinwohlinteresse vorliegt.

- **XAI (Explainable AI; kurz: XAI; deutsch: Erklärbare KI)** (siehe learn.digital-Anwendungsszenario, Szenario 2): Umfasst eine Menge von Methoden, um Transparenz und Erklärbarkeit von KI-Modellen sicherzustellen („*painting the black box white*") und so die Akzeptanz für die Nutzung personenbezogener Daten für KI-Systeme zu steigern. Aufgrund der Komplexität von KI-Systemen ist dies eine herausfordernde Aufgabe, weil selbst für KI-Entwickelnde das Verhalten eines KI-Systems häufig nicht vorhersehbar und kaum erklärbar ist. Eine vielversprechende Annäherung unter anderem ist der *LIME-Ansatz* (*Local Interpretable Model-agnostic Explanations*), bei dem ein einfach verständliches Modell, ohne in die technologische Tiefe zu gehen, eine Interpretation der Ergebnisse eines KI-Modells ermöglichen soll. Isoliert betrachtet verspricht XAI im Gegensatz zu den anderen vorgestellten Ansätzen nicht, Datenschutz qua Design sicherzustellen. Vielmehr könnte XAI eine Auflage an Datenprozessierende für die freie Verwendung personenbezogener Daten sein, wenn ihr KI-System auf einen von einer unabhängigen Stelle eindeutig attestierten gemeinwohlorientierten Zweck ausgerichtet ist. Aber auch Erklärbare KI hat „natürliche" Grenzen, die in anwendungsspezifischen Kontexten ersichtlich werden: So könnte im Mobilitätskontext ein plötzlicher, partieller Stromausfall dazu führen, dass ein Reiseassistent einige Transportmittel in dieser Region nicht empfiehlt. Reiseassistent-Nutzende erkennen den Zusammenhang nicht und sind über die Reiseempfehlung erstaunt, die sich nicht nachvollziehbar erklären lässt, weil der Sonderfall eines außerordentlichen Stromausfalls im implementierten LIME-System des Reiseassistenten nicht modelliert wurde. XAI ist somit kein Allheilmittel und stellt Datenschutz nicht qua Design her. XAI ist, da es die Unsicherheit in der Rechtsauslegung nicht verlässlich auflösen kann, nur eine inkrementelle, begleitende Option für eine flexibilisierte Datennutzung.

- **Safe AI:** Eine andere vertrauensstärkende Maßnahme in spezifischen Anwendungskontexten ist Safe AI (z. B. die Verifikation eines *Deep-Learning-Modells* zur Klassifizierung von Bildern). Dieser Ansatz beinhaltet Lösungen für die Bewertung der Sicherheit und Robustheit von KI mit Schwerpunkt auf Computer-Vision, um auf diese Weise die Transparenz von KI-Systemen und folglich das Vertrauen in sie zu stärken. So soll gewährleistet werden, dass kleine Fehler im Input-Datensatz (Inputpertubationen) nicht zu einer Fehlklassifizierung im Output führen. Der Ansatz kann zudem dazu beitragen, datenschutzverletzende Angriffe (bspw. *Membership Inference Attacks*) zu vermeiden. Aktuell befindet sich Safe AI noch in der Erprobungsphase und steht somit nicht direkt als Anwendungslösung zur Verfügung. Außerdem kann Safe AI „*stand alone*" kaum Transparenz-Anforderungen der DSGVO erfüllen und bedarf deshalb einer Kombination mit weiteren technischen Ansätzen.

Datenzugriff-Managementsysteme

Innovative Ansätze, die mehrere Phasen im Daten-Lebenszyklus – unter anderem die für KI äußerst relevante Datenspeicherung und Datensekundärnutzung, aber auch die Datenveröffentlichung – abdecken, setzen weniger an technischen Besonderheiten der Datenverarbeitung für KI-Systeme an, sondern betrachten vielmehr datenschutzkonforme Datennutzung ausgehend vom Ansatz eines integrierten Datenmanagements. Das Paradigma der von Datenprozessierenden zu beachtenden Datensparsamkeit bei der Datenerhebung aus der DSGVO wird durch die Maximen der Datensouveränität und Datensorgfalt der Betroffenen abgelöst, um einerseits Big-Data-basierte Entwicklung und -Anwendung von KI-Systemen zu ermöglichen, und andererseits Datenschutz und informationelle Selbstbestimmung sicherzustellen. Betroffenen kommt hierbei eine aktivere und gleichberechtigtere Rolle in der Datenökonomie zu als unter Maßgabe der DSGVO. Sie können so an ihrer Datenverwendung für die KI-Entwicklung bzw. -Anwendung aktiv partizipieren.

- **Datentreuhänder** (siehe vAltality-Anwendungsszenario, Szenario 1): Als Datenintermediär agieren Datentreuhänder als vertrauenswürdige dritte Instanz (*Trusted Third Entity*) vermittelnd zwischen Datengebenden und Datenerhebenden in unterschiedlichen Ausgestaltungen (Stevens & Boden 2022; Blankertz 2020; Blankertz, Specht-Riemenschneider 2021). Die Grundidee bei Datentreuhändern ist die Datenfreigabe für einen bestimmten Zweck, die Aufhebung der Personenbezogenheit und die anschließende freie Verwendung dieser Daten, wobei der Datentreuhänder das dafür benötigte Vertrauen zwischen Datengebenden und Datennutzenden schafft. Dabei müssen Datentreuhänder aber selbst nicht zwingend Daten speichern. Insbesondere mehrseitige Treuhänder sind im KI-Kontext interessant. Sie agieren als neutrale Instanz bei der Datenvermittlung und -auswertung und können sowohl zum Gemeinwohl (z. B. für medizinische Forschung) als auch als privatwirtschaftliche Treuhand datenmonetarisierend agieren. In einigen Fällen, wenn beispielsweise datenprozessierende Unternehmen zu hohe Marktmacht akkumulieren, könnte ihr Einsatz verpflichtend sein. Alternativ könnten den Betroffenen selbst in anderen Fällen – wie bei der beabsichtigten Verwendung von Daten für privatwirtschaftliche Zwecke – Opt-out- und Widerspruchsmöglichkeiten eingeräumt werden (Blankertz, Specht-Riemenschneider 2021). Das Konzept der **Datenenklave bei der Treuhand** ist aus Gemeinwohlperspektive von besonderem Interesse. Denn bei diesem werden die Daten vom Treuhänder verwaltet – nicht aber herausgegeben –, und stattdessen werden virtuelle beaufsichtigte Arbeitsumgebungen zur Verarbeitung der Daten durch berechtigte Stellen angeboten (Buchner et al. 2021, S. 810). Betroffene könnten ihre Daten für KI-Anwendende und KI-Entwickelnde by Default freigeben, wenn jene nachweisen, dass die eingesetzten KI-Systeme dem Gemeinwohl dienen. Dafür bedürfte es aber zuvor der Festlegung von Gemeinwohlkriterien. Die Anwendung der KI-Systeme mit den Daten der Betroffenen würde dann in einer geschützten Umgebung vollzogen werden, wo die Treuhand als vertrauenswürdige dritte Instanz die datenschutzwahrende Datennutzung überwachen würde. So könnte *broad consent* – nicht nur für Forschungs-, sondern sogar für Gemeinwohlzwecke – technisch implementiert werden. Rechtlich könnte die Datentreuhand als angemessene und spezifische Schutzmaßnahme (Art. 9 DSGVO) gelten und damit die Interessenabwägung zugunsten des Gemeinwohls erleichtern, was die (nachträgliche) Einholung der Einwilligung der datengebenden Betroffenen überflüssig machen würde. Allerdings ist unklar, inwieweit der Personenbezug von Daten reduziert werden müsste, um Anwendungssicherheit sicherzustellen. Zudem ist die zentrale Speicherung der Daten in der Enklave bei der Treuhand aufwändig und kostenintensiv. Auch der Datentreuhänder ist also kein Allheilmittel.

- **Personal Information Management System (PIMS)** (siehe vAltality-Anwendungsszenario, Szenario 2): Ein PIMS als Unterform eines Datentreuhänders ist ein technisches Hilfsmittel und stellt als virtuelle Einwilligungsinstanz für datenfreigebende Betroffene (*Data Subjects*) deren Informiertheit, Selbstbestimmtheit und Datensouveränität sicher. Dabei existieren verschiedene Ausgestaltungen sowohl mit dezentraler als auch zentraler Datenablage. PIMS, die nicht nur *User-Managed Access Controllers* (z. B. UMA und Web Tech), sondern Datenkontrolleure sind und somit die individuelle Datenhoheit stärken können, stehen besonders im Fokus. Das PIMS kann als Web-Portal und/oder App ausgestaltet sein, bei der Datenfreigebende ein persönliches Konto haben. Über dieses Tool können betroffene Personen steuern, für welche Initiativen (Zweckbindung) – seien es Forschungszwecke oder unternehmerische Ziele – sie ihre Daten freigeben möchten (Einwilligung). Dabei werden sie über diese Initiativen in ansprechender User Experience (z. B. Erklärfilme) informiert (Informiertheit). PIMS stellen damit als vertrauenswürdige dritte Instanz (*Trusted Third Entity*) zwischen datenverarbeitenden Entitäten und

Datenfreigebenden die Einhaltung der DSGVO sicher und vereinfachen so die Datenerhebung: So können die Datenfreigebenden ihre Präferenzen im PIMS hinterlegen und automatisch Daten für bestimmte Zwecke freigeben oder die Freigabe für bestimmte andere Zwecke ausschließen. Dies gilt auch für nachträgliche Einwilligungen für Datensekundärnutzung oder den Datenerwerb. PIMS implementieren damit ein Datenmanagementparadigma, bei dem die Daten nicht primär bei Prozessierenden liegen, sondern unter der Hoheit der Betroffenen verbleiben. Diese können ihre Daten bedarfsorientiert und informiert an Interessenten allokieren – unentgeltlich oder gegen ein Entgelt. Betroffene werden damit zu aktiven Teilnehmenden in der Datenökonomie, anstatt als passive Datenquelle von Datenprozessierenden abhängig zu sein. Obwohl erste Piloten bestehen (z. B. ViCon für den Gesundheitssektor), fehlt es für den flächendeckenden Einsatz von PIMS noch an ihrer marktreifen Ausarbeitung. Hinzu kommt eine sehr eng gefasste Regulierung von PIMS (über den *Data Governance Act* (DGA) und den § 26 des Telekommunikation-Telemedien-Datenschutz-Gesetzes (TTDSG)) sowie ein Mangel an Konkretisierung, wie die rechtlichen Vorgaben in der praktischen Umsetzung von PIMS zu berücksichtigen sind, insbesondere im Hinblick auf deren Akkreditierung und Zertifizierung, die Festlegung praktikabler Transparenz- und Berichtspflichten sowie Anforderungen an IT-Sicherheit und Zugriffsverbote (Blankertz 2021; Specht-Riemenschneider, Kerber 2021).

Datentreuhänder – sowie PIMS als ihre spezifische Ausgestaltung – versprechen, dass Betroffene über ihre Datenhoheit in der Datenökonomie nicht bloß reine Datengebende bleiben, sondern selbst an der Monetarisierung ihrer Daten partizipieren können, wodurch das Abhängigkeitsverhältnis zu Datenprozessierenden reduzierbar ist. PIMS und Datentreuhänder sind also nicht nur Hebel für eine datenschutzkonforme Datennutzungsflexibilisierung, sondern auch für eine Datennutzungsflexibilisierung mit der Aufwertung der Rolle Datengebender als souveräne Partizipierende in der ökonomischen datenbasierten Wertschöpfung – der „Persönlichen Digitalen Datenwirtschaft" (Fraunhofer MOEZ, 2013). In B2B- oder B2C-Konstellationen können nachgeschaltete Akteure in der Wertschöpfungskette bzw. Unternehmenskunden gleichberechtigter mit KI-nutzenden Anbietern interagieren. Ähnliches gilt für individuelle Personen im B2C-Verhältnis mit KI-nutzenden Anbietern. Ein solcher Paradigmenwechsel könnte eine höhere gesellschaftliche Akzeptanz des Datenteilens hervorrufen, die für KI-Systeme angesichts ihres Bedarfs an personenbezogenen Daten elementar ist. Frühzeitig sollte allerdings beachtet werden, dass Betreiber von PIMS und von Datentreuhändern aus ihrer neuralgischen Stellung heraus in der Datenwertschöpfung signifikante Macht akkumulieren könnten. Daher sollten sie bereits by Design auf das Gemeinwohl und auf das Agieren im Sinne der Datengebenden verpflichtet werden.

Kaum Rechtssicherheit für existente technische Ansätze zu datenschutzkonformer Datennutzung für KI

Zusammenfassend bestehen über die einzelnen Phasen im Datenlebenszyklus also vielschichtige technische Ansatzpunkte, um eine praktikable datenschutzwahrende Datennutzung für KI-Systeme im Gemeinwohlinteresse zu flexibilisieren. Vor dem Hintergrund unterschiedlicher Leistungsfähigkeiten der Maßnahmen (siehe Tabelle 1) hinsichtlich Genauigkeit, Kommunikationsaufwand, Rechenlast, Datenmanagement sowie Invasivität bezogen auf die Datenintegrität ist eine sorgfältige Abwägung und Auswahl der jeweilig für den Anwendungskontext geeigneten Maßnahme vorzunehmen.

Tabelle 1: Einordnung der technischen Maßnahmen nach jeweiliger Leistungsfähigkeit

Maßnahme	Gewährleistung Datenschutz	Wahrung Datenqualität	Aktuelle Rechtssicherheit	Praktikabilität (hinsichtlich Aufwand)	Potenzial für KI	Illustriert in Anwendungsszenario
Anonymisierung	by Design	Verrauschung	DSGVO-Anwendungsbereich	Ex-ante-Manipulation des Datensatzes	schwache Datenqualität, Rechtsunsicherheit	learn.digital, Szenario 1
Pseudonymisierung	by Design	Verrauschung	DSGVO-Anwendungsbereich	Ex-ante-Manipulation des Datensatzes	schwache Datenqualität, Rechtsunsicherheit	learn.digital, Szenario 1
Datensynthese	by Design	Synthetisierung	fällt nicht unter DSVGO	Synthetisierung dauert	Standardverfahren für ML außer für Echtzeitsysteme	LEASYNG, Szenario 1
Differential Privacy	by Design	begrenzte Verrauschung	Human in the Loop als Fehlerquelle	ressourcen-/personenintensiv	nicht für Echtzeitsysteme, ressourcenintensiv	vAltality, Szenario 3
Homomorphic Encryption	by Design	keine Datenmanipulation	unklar	Rechenlast	nicht für Echtzeitsysteme, ressourcenintensiv	learn.digital, Szenario 4
Secure Multiparty Computation	by Design	keine Datenmanipulation	unklar	Kommunikationskosten	nicht für Echtzeitsysteme, ressourcenintensiv	learn.digital, Szenario 4
Confidential Computing	by Design	keine Datenmanipulation	unklar	Rechenlast und Kommunikationskosten	nicht für Echtzeitsysteme, ressourcenintensiv	LEASYNG, Szenario 3
Verteiltes Lernen	by Design	keine Datenmanipulation	nicht reguliert	Enablement von Clients für Edge KI	Datensouveränität, gute Performance, aber Angriffsvektoren	vAltality, Szenario 3
Federated Learning + Differential Privacy	by Design	Verrauschung der Weights	nicht reguliert	Verrauschungsaufwand der Weights ressourcenintensiv	Datensouveränität, Datenschutz, mittlere Performanz	vAltality, Szenario 3
XAI	eher Gewährleistung von Transparenz, Informiertheit	keine Interferenz	nicht reguliert	Erklärbarkeit nicht definierbar	Transparenz kann Vertrauen stärken	learn.digital, Szenario 2
Personal Information Management Systems	Datensouveränität/-partizipation als Paradigma	keine Interferenz	Anknüpfung an Informiertheit und Zweckbindung	Set-up-Kosten, dann gute Skalierbarkeit	Datensouveränität/-partizipation; Flexibilisierung	vAltality, Szenario 2
Datenenklave in der Datentreuhand	Datensouveränität und Trusted Third Entity	keine Interferenz	Anknüpfung an Informiertheit und Zweckbindung, broad consent	Set-up-Kosten, dann gute Skalierbarkeit	Datensouveränität/-partizipation; Flexibilisierung	vAltality, Szenario 1

Grundvoraussetzung für die praktische Anwendung all dieser Maßnahmen ist, dass zuvor ihre juristische Anerkennung erfolgt, um bestehende Unsicherheit(en) in der Rechtsauslegung der DSGVO nachhaltig aufzulösen, anstatt sie bloß auf andere Bereiche – wie die rechtliche Einstufung datenschutzwahrender technischer Ansätze – zu verschieben. Dafür bedarf es rechtlicher Anpassungen.

4. Gestaltungsoptionen und Ausblick

Für einen modernen und funktionsfähigen Rechtsrahmen gilt sowohl im Allgemeinen als auch im Speziellen für eine flexibilisierte datenschutzwahrende Datennutzung im Gemeinwohlinteresse, dass er diejenigen Instrumente zulässt und fördert, die zur Eindämmung dieser Unsicherheit in der Rechtsauslegung beitragen und gleichzeitig kollidierende Rechte und Rechtsgüter umfassend schützt. Dies bedeutet:

1. Ein holistischer Ansatz zur rechtlichen Anerkennung einer flexibilisierten Datennutzung im Gemeinwohlinteresse ausgehend von den vorgestellten technischen Maßnahmen sowie dem integrierten Datenmanagement im Gemeinwohlinteresse ist zielführend.

2. Anpassungen bezogen auf die einzelnen vorgestellten technischen Maßnahmen (*Privacy by Design*) sind zusätzlich notwendig, wenn die Verwendung personenbezogener Daten für Prozessierende alternativlos ist.

Beide Ansätze sollten in einen technikneutralen Datenschutzrechtsrahmen eingegossen werden, der anwendungsorientierte und klar interpretierbare Vorgaben spezifisch für die Datennutzung für KI-Systeme definiert, um so Rechts- und Handlungssicherheit für die KI-Entwicklung zu schaffen. Darüber hinaus sind weitere Anpassungen für die flexibilisierte Datennutzung im Gemeinwohlinteresse unter Datenschutzwahrung empfehlenswert.

4.1 Datennutzung für KI bei nachgewiesenem Gemeinwohlinteresse flexibilisieren

> *Wir empfehlen einen technikneutralen und entwicklungsoffenen regulatorischen Ansatz, der die Maßgabe realisiert, Handlungs- statt Verbotsräume für eine gemeinwohlorientierte Datennutzung zu schaffen.*

Einheitliche Definition des Gemeinwohlinteresses

Voraussetzung für eine flexibilisierte Datennutzung im Gemeinwohlinteresse ist eine **einheitliche Definition, was unter Gemeinwohlinteresse zu verstehen ist**. Gemeinwohlinteresse sollte nicht durch die Verengung der Definition auf „das, was durch Forschung und zivilgesellschaftliche Aktivitäten entsteht", genutzt, sondern breit im eigentlichen Wortsinn als Verbesserung vom aktuellen gesellschaftlichen Status quo verstanden werden und somit auch privatwirtschaftlich erbrachte gemeinwohlfördernde Leistungen umfassen. Einem weiten Ansatz folgend schlagen wir vor, Gemeinwohlinteresse zu definieren als „Tätigkeit, deren Erbringung nicht allein durch individuelle – wirtschaftliche, eigennützige, freundschaftliche oder familiäre – Ziele motiviert ist, sondern sich zumindest auch als Ausdruck gesellschaftlicher Verantwortung erweist" (Trésoret 2018). Der Entwurf des *AI Act* der Europäischen Union versteht sich als Regulierung im Sinne von Produktsicherheit; wir empfehlen, solch einen präventiven Ansatz in dieser Regulierung mit einem schöpferischen zur Datennutzungsflexibilisierung im Gemeinwohlinteresse zu ergänzen, um Potenziale des Datenschatzes für das Gemeinwohlinteresse zielführend heben zu können.

Recht und Zertifizierung

Darauf aufbauend sollte **Gemeinwohlinteresse entlang konkreter Anwendungskontexte rechtssicher definiert** werden und im Falle des Vorliegens Datennutzungsflexibilisierungen als Rechtsfolge einräumen. Die Beurteilung, ob Gemeinwohlinteresse vorliegt, sollte dabei nicht nur im isolierten Kontext vollzogen werden, sondern in holistischer Sicht Implikationen und mögliche Externalitäten anhand eines klaren Kriterienrahmens messbar machen. So kann die Datenfreigabe für gemeinwohlorientierte KI-Systeme von großen Digitalkonzernen auf den ersten Blick einen direkten gemeinwohlorientierten Nutzen im Anwendungskontext aufweisen (z. B. Google Books: Demokratisierung des Wissenszugangs); wenn aber durch anschließende Kapitalisierung daraus durch einen Digitalkonzern negative Externalitäten für das Gemeinwohl entstehen (Monopol auf Wissenszugang durch Marktmacht von Google Books), ist dies nicht im Sinne einer flexibilisierten Datennutzung im Gemeinwohlinteresse. Ein Kriterienrahmen kann hier Abhilfe leisten.

Grundsätzliche rechtliche Anerkennung einer flexibilisierten Datennutzung im Gemeinwohlinteresse

Ausgehend von der erforderlichen juristischen Übersetzung der Anwendungskontext-spezifischen Definition des Gemeinwohlinteresses muss eine **grundsätzliche rechtliche Anerkennung der Möglichkeit einer flexibilisierten Datennutzung im Gemeinwohlinteresse im bestehenden Datenschutzrechtsrahmen** erfolgen. Die Datennutzung im Gemeinwohlinteresse unter strengen Sicherheitsmaßnahmen und Zugriffskontrollen könnte als rechtliche Ausnahme in der DSGVO formuliert werden und der Regulierung in anderen, konkreteren Rechtsvorschriften einen Rechtsvorrang einräumen (vorrangige Spezialgesetzgebung). Dies würde bessere Möglichkeiten einer zielgenaueren und anwendungsspezifischeren Datenschutzgesetzgebung für die Datennutzung für KI-Systeme eröffnen, als sie in der dafür nicht geeigneten DSGVO zu regeln (Stiftung Datenschutz 2021).

Daraus folgt das Erfordernis, **Rechtsfolgen der Flexibilisierung zu definieren**, um Datenprozessierenden Handlungssicherheit für den Umgang mit bereits von ihnen gespeicherten Daten zu verschaffen sowie datengebenden Betroffenen klare Orientierung zu geben, was mit den sie betreffenden Daten im Sinne einer Verwendung für KI-Systeme im Gemeinwohlinteresse passieren wird.

4.2 Privacy by Design stärken und rechtlich anerkennen

> *Wir empfehlen, bestehende Privacy-Preserving-Machine-Learning-Ansätze weiterzuentwickeln und zu evaluieren und das Vehikel Privacy by Design rechtlich anzuerkennen.*

Wir bekräftigen die Grundüberzeugung im Sinne des Datenschutzes, dass für das Training von KI-Systemen nicht-personenbezogene Daten grundsätzlich personenbezogenen Daten vorgezogen werden sollten, sofern sie die gleiche Datenqualität aufweisen. Wir empfehlen deshalb, die **Verfügbarkeit nicht-personenbezogener Daten zu stärken**, um die Ersetzung von personenbezogenen Daten für das Training von KI-Systemen zu ermöglichen. Vermehrte Forschungs- und Entwicklungsinitiativen sollten beispielsweise in den Aufbau interoperabler Datenräume fließen.

Ist die Verwendung personenbezogener Daten für das Training eines KI-Systems alternativlos, bedarf es einer umfassenden Stärkung des Vehikels „*Privacy by Design*" durch vertiefte Forschung und Entwicklung, Standardisierungs- und Zertifizierungsinitiativen, technische Weiterentwicklung sowie wirtschafts-, forschungs- und bildungspolitische Initiativen und auch ihrer rechtlichen Anerkennung.

Forschung und Entwicklung

Erklärbare KI könnte ein Vehikel sein, die Entscheidungen eines KI-Systems vor dem Hintergrund ihres Einsatzes im Gemeinwohlinteresse transparent und verständlich zu erklären. Noch kann sie diesem Versprechen aber aufgrund mangelnder Klarheit, wie weit Erklärbarkeit greift, nicht gerecht werden. Neue Forschungsansätze verstehen XAI als einen Dialog, in dem dynamische Interaktion von Anwendenden und Nutzenden mit einem KI-System bessere Erklärbarkeit schaffen kann. Hier bedarf es einer **strukturierten Aufbereitung**, inwiefern **Erklärbare KI** betroffenen Nutzenden die Funktionsweise und das Gemeinwohlinteresse eines KI-Systems transparent und einfach verständlich erklären kann und so die selbstverantwortlichen Entscheidungsprozesse **der Betroffenen** stärkt.

Für die Ausgestaltung von PIMS ist zu klären, **inwieweit diese so zu gestalten sind, dass sie die Datensouveränität datengebender Betroffener stärken können**. Ziel sollte sein, dass PIMS Nutzende bewusst durch den Prozess der Freigabe ihrer personenbezogenen Daten führen und ihnen mögliche Implikationen einer Datenfreigabe in einer solchen Weise verdeutlichen können, sodass diese eine wirklich selbstbestimmte Entscheidung über die Freigabe ihrer persönlichen Daten treffen können. Hier sollte interdisziplinäre Expertise beispielsweise aus der Psychologie und Kognitionsforschung einbezogen werden.

Recht und Zertifizierung

Im Sinne eines technikneutralen Ansatzes empfehlen wir die Stärkung eines **Standardisierungs- und Zertifizierungsregimes technischer Maßnahmen für die Anonymisierung personenbezogener Daten**, bei deren Einhaltung und bei deren erfolgter Zertifizierung die erfolgreiche Anonymisierung personenbezogener Daten durch eine technische Maßnahme rechtlich vermutet wird. Dies entspräche einer signifikanten Datennutzungsflexibilisierung. Ähnlich wie in anderen technischen Bereichen (z. B. Automobilzulassung) sollten unabhängige Behörden oder Konsortien Zertifizierungsschemata einführen und die technischen Maßnahmen entsprechend dem in Standards festgelegten aktuellen Stand der Technik klassifizieren können. Eine Standardisierung nach DIN-Normen könnten hier ein Anknüpfungspunkt sein.

Wenn technisch sichergestellt ist, dass Erklärbare KI die Selbstbestimmung Betroffener durch Erklärung der Funktionsweise und des Gemeinwohlinteresses eines KI-Systems steigern kann, sind entsprechende **Erklärbare-KI-Ansätze zu standardisieren, um sie rechtssicher als Schutzmöglichkeit personenbezogener Daten anzuerkennen** und ihre flexibilisierte Datennutzung im Gemeinwohlinteresse als Rechtsfolge einzuräumen.

Hinsichtlich der Maßnahmen des integrierten Datenmanagements (PIMS und Datentreuhänder) bekräftigen wir die Empfehlung, die **Delegierbarkeit von Datenrechten von Betroffenen an PIMS (technisch) bzw. Datentreuhänder (institutionell) rechtssicher und DSGVO-konform anzuerkennen** (Blankertz 2021; Blankertz 2020).

Der *AI Act* der Europäischen Union schreibt für KI-Produkte einer gewissen Risikostufe ex ante oder ex post Bewertungen oder Zertifizierungen von Produktanforderungen vor, die teils von Drittparteien durchgeführt werden müssen. Um Engpässe für die KI-Entwicklung zu vermeiden, sind gezielte Maßnahmen erforderlich, eine belastbare Infrastruktur an zertifizierenden Stellen aufzubauen.

Wirtschafts- und Forschungspolitik

- Für die Steigerung der technischen Exzellenz der vorgestellten Maßnahmen sind weitere Forschungs- und Entwicklungsschritte vonnöten. Es bedarf eines **fundierten Förderrahmens zu Privatsphäre-fördernden Technologien zum Nutzen von Gesellschaft und Wirtschaft**. Dieser Förderrahmen sollte Forschungsaktivitäten zu den vielversprechendsten technologischen Ansätzen incentivieren.

- Voraussetzung für den gemeinwohlorientierten Einsatz von Datentreuhändern oder PIMS ist die Sicherstellung der rechtlichen und finanziellen Unparteilichkeit der Betreiber. Wir bekräftigen deshalb die Empfehlung einer **nötigen Anschubfinanzierung von Datentreuhändern** aus öffentlichen Quellen, wie sie beispielsweise mit dem *Mobility Data Space* erfolgt ist.

- Der *AI Act* der Europäischen Union sieht vor, auch die KI-Forschung zu regulieren. Dies sollte sehr sorgsam und im Einvernehmen mit der Forschungsgemeinschaft geschehen, um das Innovationspotenzial und die Wertschöpfung aus KI für das Gemeinwohl in der EU zu stärken.

Bildungs- und Arbeitsmarktpolitik

- **Interdisziplinäre Kompetenzvermittlung für KI stärken:** Im Ausland und sukzessive auch in Deutschland existieren bereits Bildungsabschlüsse, die sich auf Data Science, KI und deren Anwendungen fokussieren (siehe KI-Landkarte der Plattform Lernende Systeme). Dieser wachsende Bildungsmarkt bietet eine willkommene Möglichkeit, eine technische Kernausbildung in KI mit Kompetenzen im rechtlichen und ethischen Raum anzureichern und zu verbinden, sodass technische Fähigkeiten in der Produktentwicklung in der beruflichen Breite mit rechtlichem und ethischem Sachverstand gemeinwohlorientiert umgesetzt werden können. Eine zukunftsorientierte, gemeinwohlorientierte Nutzung von Daten für KI kann durch solch neue Ansätze in der Lehre und der praktischen Ausbildung erheblich gefördert werden. Wir empfehlen daher die interdisziplinäre Vermittlung von ethischen, rechtlichen und sozialwissenschaftlichen Kompetenzen gemeinsam mit technischen, naturwissenschaftlichen und produktorientierten Aspekten. Interdisziplinäre Ansätze, die im Ausland erprobt werden (z. B. im Gateway der University of California at Berkeley), sollten als Blaupausen eine intensivierte Diskussion um die Zukunft der KI-Ausbildung in Deutschland anregen.

- **Allgemeine Datenkompetenzen stärken:** Insbesondere für den breiteren Einsatz von PIMS und die Aufwertung der Stellung von datengebenden Betroffenen in der Datenökonomie empfehlen wir eine Stärkung von Konzepten, die die Ausbildung individueller Datenkompetenzen fördern (Blankertz 2021). Dies betrifft nicht nur die (Aus-)Bildung junger Generationen, sondern auch die Aufklärung älterer Mitbürgerinnen und Mitbürger. *Data Privacy Literacy* muss eine bildungs- und arbeitspolitische Priorität werden.

LEASYNG – Mobilität

Datenzweitverwertung

Das Berliner Unternehmen **LEASYNG** – ein nachhaltiger Mobilitätsleasingdienst – verleiht in und rund um Berlin per App und über seine Onlineplattform Mietfahrzeuge aller Art zu individuellen Konditionen. Zahlreiche Daten, darunter viele KundInnendaten, liegen LEASYNG aufgrund der hohen Nutzungsrate vor. Diese Daten möchte das Unternehmen gerne zweitverwerten, um die KundInnenzufriedenheit zu steigern, den ökologischen Fußabdruck des Unternehmens zu reduzieren und die Flottenauslastung zu optimieren.

Datenschatz
Personenbezogene/-beziehbare Daten

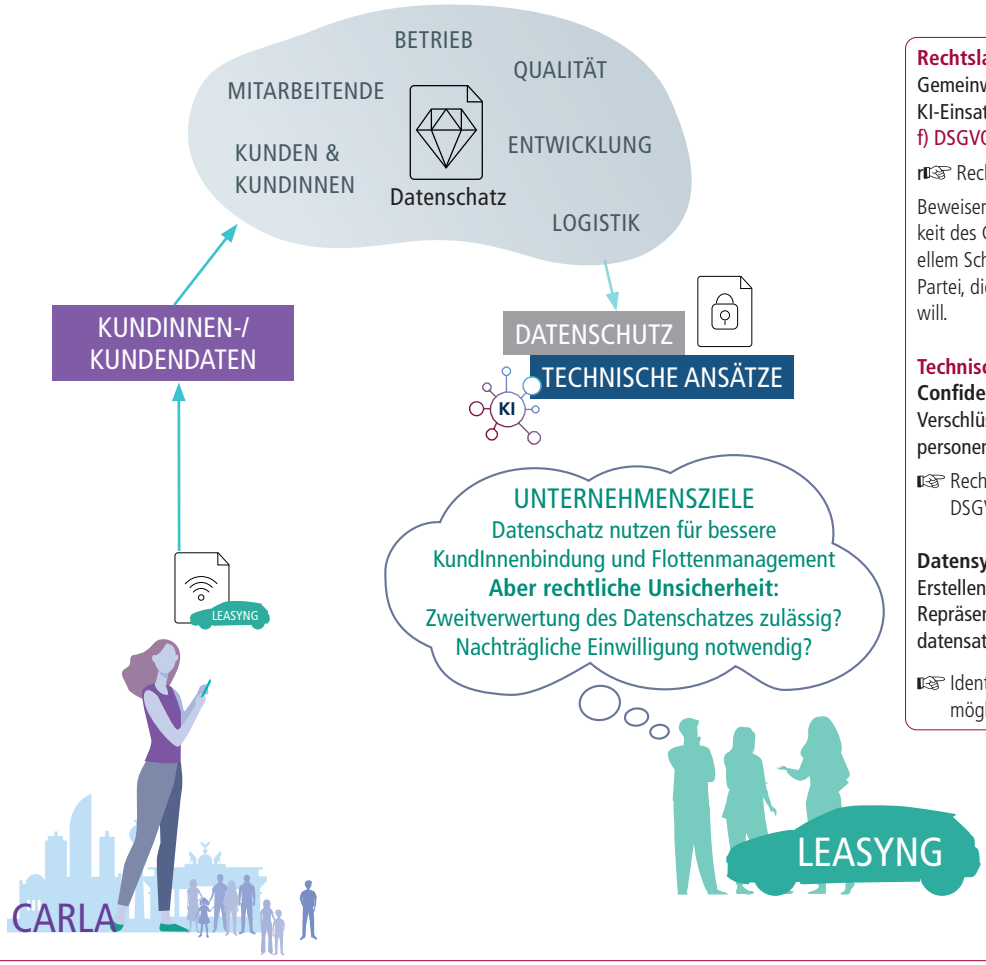
- **Biografische Daten**
  Alter, Geschlecht, Wohnort, Name etc.
- **Daten zum Mobilitätsverhalten**
  Häufige Standorte, genutzte Verkehrsmittel, zurückgelegte Strecken etc.
- **Daten zum App-Nutzungsverhalten**
  Häufigkeit/Regelmäßigkeit der Fahrzeuganmietung, Standort/Uhrzeit des App-Zugriffs etc.

Datenschutz
Datenzweitverwertung

Vorliegende KundInnendaten sollen für weitere Zwecke weiterverarbeitet werden.

Sekundärverwertung zulässig nach, ...
Art. 6 DSGVO: wenn die betroffene Person zur Datenverarbeitung für weiteren bestimmten Zweck einwilligt.

Art. 6 Abs. 1 S. 1 lit. f) DSGVO: wenn höherwertiges Gemeinwohlinteresse gemäß individueller Interessenabwägung nachweislich belegt ist.



INFO

Rechtslage
Gemeinwohlorientierter KI-Einsatz (Art. 6 Abs. 1 S. 1 lit. f) DSGVO)

☞ Rechtliche Auslegung unsicher.

Beweiserbringung der Höherwertigkeit des Gemeinwohls vor individuellem Schutzinteresse liegt bei der Partei, die die Daten verwenden will.

Technische Ansätze
Confidential Computing: Verschlüsselungstechnik für personenbezogene Daten

☞ Rechtlich anerkannt und DSGVO-konform? Unklar.

Datensynthese: Erstellen einer künstlichen Repräsentation eines Originaldatensatzes

☞ Identifizierbarkeit trotzdem möglich? Unklar.

Carla lebt und arbeitet in Berlin und hat aufgrund des guten ÖPNV-Angebots kein eigenes Auto. Für ihre Familienbesuche am Wochenende nach Brandenburg leiht sie sich ein Mietauto über den Mobilitätsanbieter LEASYNG. Um ihre Freundin in Tschechien besuchen zu können, nutzt sie ab und an einen LEASYNG-Mietwagen, den sie bei ihren Eltern abstellt, um mit dem familieneigenen Auto die Reise fortzusetzen.

LEASYNG möchte über den Einsatz von Künstlicher Intelligenz seine Services im Kundenbereich, in der Preisgestaltung sowie bei der Verfügbarkeit der Fahrzeuge verbessern. Hierfür könnte ein Teil des Datenschatzes – personenbezogene KundInnendaten – genutzt und zweitverwertet werden. Vorab ist aber zu klären, unter welchen (datenschutz-)rechtlichen Bedingungen dieser Datenschatz genutzt werden kann und welche technischen Ansätze sich für eine datenschutzkonforme Datennutzung für KI-Systeme anbieten. Die Unternehmensführung bittet deshalb die hauseigene Rechtsabteilung um eine Einschätzung (datenschutz-) rechtlicher Vorgaben und die IT-Abteilung um Ideen für die technische Umsetzung.

Ziele des Unternehmens, auch mit KI-Einsatz

- **Selbstlernendes Flottenmanagement:** Reduzierung der Steh- und Transferzeiten
- **Individualisiertes Pricing:** Anreiz, Fahrzeuge im Kerngeschäftsbereich abzustellen
- **KI-Chatbots:** KundInnenorientierte Kommunikation
- **Optimale Zugänglichkeit der Fahrzeuge:** Steigerung der Attraktivität
- **Reduktion der CO_2-Emissionen:** Freiwilliger Beitrag für eine nachhaltige Entwicklung

Datenzweitverwertung zum KI-optimierten Flottenmanagement

SZENARIO 1

Ausgangssituation

LEASYNG möchte **personenbezogene KundInnendaten** (Daten zum Mobilitätsverhalten) zur Optimierung der Flottenauslastung zweitverwerten.

Rechtliche Lage

Aktuelle Rechtslage eröffnet zwei Interpretationsspielräume

1. Interpretationsspielraum
Datenzweitverwertung für einen Zweck, der nicht wesensähnlich zur ursprünglichen Einwilligung ist (Art. 6 Abs. 4 DSGVO), bedarf einer nachträglichen Einwilligung. Außer: Zweckänderung gilt als „**logischer nächster Schritt**": ursprünglicher Zweck ist mit neuem Zweck kompatibel (datenprozessierende Entität hat Nachweispflicht!).

2. Interpretationsspielraum
Ist das **Gemeinwohlinteresse** der Datenverwertung gemäß individueller Interessenabwägung **höherwertiger** als das individuelle Schutzbedürfnis einzustufen (Art. 6 Abs. 1 S. 1 lit. f) DSGVO), ist eine nachträgliche Einwilligung nicht notwendig.

☞ 1. Die Zweitverwertung von Daten zum Flottenmanagement kann nicht als „logischer nächster Schritt" der ursprünglichen Datenerhebung zum Zwecke des Abschlusses eines Leasingvertrags zwischen Carla und LEASYING angesehen werden.

2. Wenn LEASYNG die Höherwertigkeit des Gemeinwohlinteresses (Reduzierung der CO_2-Emissionen, bessere Fahrzeugverfügbarkeit) durch das Flottenmanagement rechtssicher nachweist, ist eine nachträgliche Einwilligung obsolet.

Datenschutz technisch sichergestellt?

Funktionsweise Datensynthese: Die personenbezogenen Daten zum Mobilitätsverhalten werden direkt bei der Erhebung mithilfe trainierter Generativer Adversarial Networks (GAN) in einen künstlichen synthetischen Datensatz umgewandelt, der nur die für die Routenoptimierung relevanten Informationen enthält und alle anderen personenbezogenen Daten eliminiert. LEASYNG speichert und nutzt diesen synthetisierten Datensatz.

Die Datensynthese ermöglicht es, den Datenschutz bei Erhebung und Verarbeitung der personenbezogenen/-beziehbaren KundInnendaten zu wahren: Für das KI-System sind nur die Daten zum Mobilitätsverhalten relevant und keine weiteren personenbezogenen/-beziehbaren Daten wie Namen, Kommunikationspräferenzen etc., die eine Identifizierbarkeit erlauben und besonders datenschutzkritisch wären.

☞ Nicht sichergestellt, ob synthetische Daten der Datensynthese eine Identifizierbarkeit erlauben (Rückschlüsse bspw. durch Geotracking auf Wohnort).

nein / ja

Technische Sicherstellung des Datenschutzes – rechtlich anerkannt?

☞ **Datensynthese** aktuell rechtlich nicht anerkannt zur DSGVO-konformen Nutzung

nein

Höherwertigkeit Gemeinwohlinteresse?

- Prüfschritt 1: **Plausibilität** CO_2-Reduzierung und bessere Fahrzeugverfügbarkeit? Eher hoch.
- Prüfschritt 2: **Rechtssicherheit** Gemeinwohlvorrang vor Einzelfallentscheidung?

nein/unklar

ja

Keine DSGVO-konforme Aktion notwendig

ja

Keine Datennutzung für KI-System

LEASYNG sieht wegen der teuren nachträglichen Einholung der Einwilligungen vom KI-Einsatz zum Flottenmanagement ab. Einige KundInnen – auch Carla – wechseln zu Wettbewerbern, die eine bessere Verfügbarkeit der Services bieten und damit den eigenen ökologischen Fußabdruck geringer ausfallen lassen.

(-) KundInnenvorteil
(-) Unternehmensvorteil
(+) Datenschutz
(/) Gemeinwohl

Entscheidung Kosten/Nutzen bei Datennutzung für KI-System

nein / ja

DSGVO-konforme Aktion notwendig: Nachträgliche Einwilligung

DSGVO-konforme Datennutzung für KI-System

LEASYNG setzt das KI-optimierte Flottenmanagement ein, muss aber nachträglich die Einwilligung aller KundInnen einholen. Trotz der besseren User Experience geben einige KundInnen – auch Carla – keine Einwilligung und wechseln den Anbieter.

(/) KundInnenvorteil
(/) Unternehmensvorteil
(+) Datenschutz
(+) Gemeinwohl

DSGVO-konforme Datennutzung für KI-System

LEASYNG setzt das KI-optimierte Flottenmanagement ein, muss aber die Höherwertigkeit des Gemeinwohlinteresses gegenüber dem individuellen Schutzinteresse nachweisen, was ein finanzielles Risiko für LEASYNG darstellt.

(+) KundInnenvorteil
(-) Unternehmensvorteil
> bessere Fahrzeugbereitstellung (+)
> gerichtlicher Nachweis „Gemeinwohlorientierung" (-)
(+) Datenschutz
(+) Gemeinwohl
> Nachhaltigkeit/ CO_2-Reduktion (+)

Technische Sicherstellung einer DSGVO-konformen Datennutzung für KI-System

LEASYNG kann seine Effizienz und Nachhaltigkeit sowie sein Nutzenversprechen durch bessere Bereitstellung der Fahrzeuge für KundInnen steigern, was den Umsatz erhöht. LEASYNG bildet Rücklagen aufgrund der rechtlichen Unsicherheit.

(+) KundInnenvorteil
(+) Unternehmensvorteil
> bessere Fahrzeugbereitstellung (+)
> Rücklagen/rechtliche Unsicherheit (-)
(+) Datenschutz
(+) Gemeinwohl
> Nachhaltigkeit/ CO_2-Reduktion (+)

Datenschützende Datennutzung über DSGVO-Ansatz

Datenschützende Datennutzung über technischen Ansatz

Datenzweitverwertung für individualisiertes Pricing als finanzielles Anreizsystem

SZENARIO 2

Ausgangssituation

LEASYNG möchte **personenbezogene KundInnendaten** (Daten zum Nutzungsverhalten) einsetzen für eine individualisierte Preisklassifikation.

Rechtliche Lage

Unregulierte Datennutzung kann zu unzulässiger (Preis-)Diskriminierung führen, falls Anpassung allein durch das Nutzungsverhalten determiniert ist.

☞ **Szenario einer unregulierten Datennutzung:** Aufgrund der Preisgestaltung durch das KI-System müsste Carla höhere Mietpreise für das Auto zahlen: Ihre personenbeziehbaren Daten zum Mobilitätsverhalten zeigen, dass sie das Auto häufig außerhalb des Haupteinzugsbereichs abstellt.

Anerkennung der unzulässigen Datennutzung?

ja

Keine Datennutzung für KI-System

LEASYNG nimmt statt einer individualisierten Preisklassifikation eine für alle KundInnen geltende dynamische, uhrzeitabhängige Preisklassifikation vor und kann so – wenn auch nicht im gleichen Maße – darauf einwirken, dass Mietfahrzeuge eher im Berliner Einzugsbereich abgestellt werden, ohne dass KundInnen diskriminiert werden.

(+) KundInnenvorteil
> Schutz vor Diskriminierung (+)
> bessere Verfügbarkeit (+)
(-) Unternehmensvorteil
> bessere Auslastung durch bessere Verfügbarkeit (+)
(+) Datenschutz
(+) Gemeinwohl
> bessere Verfügbarkeit (+)
> Schutz vor Diskriminierung (+)

Datenzweitverwertung für KI-Chatbots zur Steigerung der KundInnenzufriedenheit

SZENARIO 3

Ausgangssituation

LEASYNG möchte **personenbezogene KundInnendaten** (Daten zum Kommunikationsverhalten) für eine interaktive und individualisierte KundInnenkommunikation per KI-Chatbot zweitverwerten.

Rechtliche Lage

Zweitverwertung personenbezogener Daten (sofern sie unter die DSGVO fallen) bedarf der Einholung einer nachträglichen Einwilligung, sofern sie nicht im ursprünglichen Zweck liegt (Art. 5 Abs. 1 lit. b) DSGVO).

Zweitverwertung ohne nachträgliche Einwilligung möglich, wenn Personenbezug technisch eliminiert ist oder das Gemeinwohlinteresse höher wiegt.

☞ Die Zweitverwertung der Daten für den Chatbot entspricht nicht dem ursprünglichen Zweck der Datenerhebung: Abschluss des Auto-Mietvertrags.

Datenschutz technisch sichergestellt?

Funktionsweise Confidential Computing:
Carlas personenbezogene Daten werden bei der App-Interaktion verschlüsselt, gespeichert und in einer geprüften Enklave hochgeladen, wo sie mit attestierter ML-Software (u. a. Natural Language Processing) verarbeitet werden. LEASYNG hat keinen Einblick in die Daten, muss lokal keine ML-Infrastruktur aufsetzen und der KI-Chatbot interagiert basierend auf den Ende-zu-Ende-verschlüsselten Daten von Carla mit ebendiesen.

Confidential Computing ermöglicht die Einhaltung des Datenschutzes bei der Verarbeitung personenbezogener/-beziehbarer Daten: Datenverschlüsselung bei Erhebung ohne Reduktion der Datenqualität.

☞ Ende-zu-Ende-Verschlüsselung erlaubt nicht zwingend Elimination von Personenbezug.

Technische Sicherstellung des Datenschutzes – rechtlich anerkannt?

nein

☞ **Confidential Computing** aktuell rechtlich nicht anerkannt zur DSGVO-konformen Nutzung

ja

Keine DSGVO-konforme Aktion notwendig

Höherwertigkeit Gemeinwohlinteresse?

nein

- Prüfschritt 1: **Plausibilität** Bessere Interaktion mit allen Altersgruppen? Eher gering.
- Prüfschritt 2: **Rechtssicherheit** Gemeinwohlvorrang vor Einzelfallentscheidung?

nein/unklar

ja

Entscheidung Kosten/Nutzen bei Datennutzung für KI-System

nein

ja

DSGVO-konforme Aktion notwendig: Nachträgliche Einwilligung

Keine Datennutzung für KI-System

LEASYNG sieht wegen der teuren Einholung der Einwilligungen von einem Chatbot-Einsatz ab. Einige KundInnen – auch Carla – wechseln zu Wettbewerbern, die per Chatbot eine lebendigere KundInneninteraktion bieten.

(-) KundInnenvorteil
(-) Unternehmensvorteil
(+) Datenschutz
(/) Gemeinwohl

DSGVO-konforme Datennutzung für KI-System

LEASYNG setzt den Chatbot ein, muss aber nachträglich die Einwilligung aller KundInnen einholen. Trotz der besseren User Experience geben einige KundInnen – auch Carla – keine Einwilligung und wechseln den Anbieter.

(+) KundInnenvorteil
> Wechsel (-)
> Verbesserung (+)
(-) Unternehmensvorteil
> Einsatz (+)
> Aufwand (-)
(+) Datenschutz
(/) Gemeinwohl

DSGVO-konforme Datennutzung für KI-System

LEASYNG setzt den Chatbot ein, da der Datenschutz über die Höherwertigkeit des Gemeinwohls sichergestellt ist. Basierend auf Carlas biografischen Daten adaptiert der KI-Chatbot eine junge weibliche Stimme mit Berliner Akzent und interagiert intelligent mit Carla.

(+) KundInnenvorteil
(+) Unternehmensvorteil
(+) Datenschutz
(/) Gemeinwohl

Technische Sicherstellung einer DSGVO-konformen Datennutzung für KI-System

LEASYNG setzt den Chatbot ein, da der Datenschutz technisch sichergestellt und rechtlich anerkannt ist. Basierend auf Carlas biografischen Daten adaptiert der neu eingesetzte KI-Chatbot eine junge weibliche Stimme mit Berliner Akzent und interagiert intelligent mit Carla.

(+) KundInnenvorteil
(+) Unternehmensvorteil
(+) Datenschutz
(/) Gemeinwohl

Datenschützende Datennutzung über DSGVO-Ansatz

Datenschützende Datennutzung über technischen Ansatz

learn.digital – Bildung

Datenverknüpfung für individualisiertes digitales Lernen und Verbesserung der User Experience

learn.digital – ein digitaler Lernplattformanbieter – hat sich zum Ziel gesetzt, über alle Altersgruppen hinweg lebenslanges Lernen zu fördern und über ihre digitale Lern-App die Digital Literacy der Nutzenden zu stärken. Hierfür beabsichtigt learn.digital, das Eye-Tracking-System des Anbieters eye.ai zu nutzen, um aus den darüber gewonnenen Daten der Blickaufzeichnungen von den Nutzenden eine angepasste altersspezifische Interaktion für ihre Lerntools abzuleiten. Dazu sind die learn.digital-KundInnendaten mit den zu erfassenden Daten aus dem Eye-Tracking-System von eye.ai zu verknüpfen.

Datenschatz
Personenbezogene/-beziehbare Daten

- **Blickaufzeichnungsdaten**
  Fixationsort, -dauer etc. zur Ableitung kognitiver Prozesse
- **Verhaltensdaten**
  Verweildauer, Click-Through-Rate

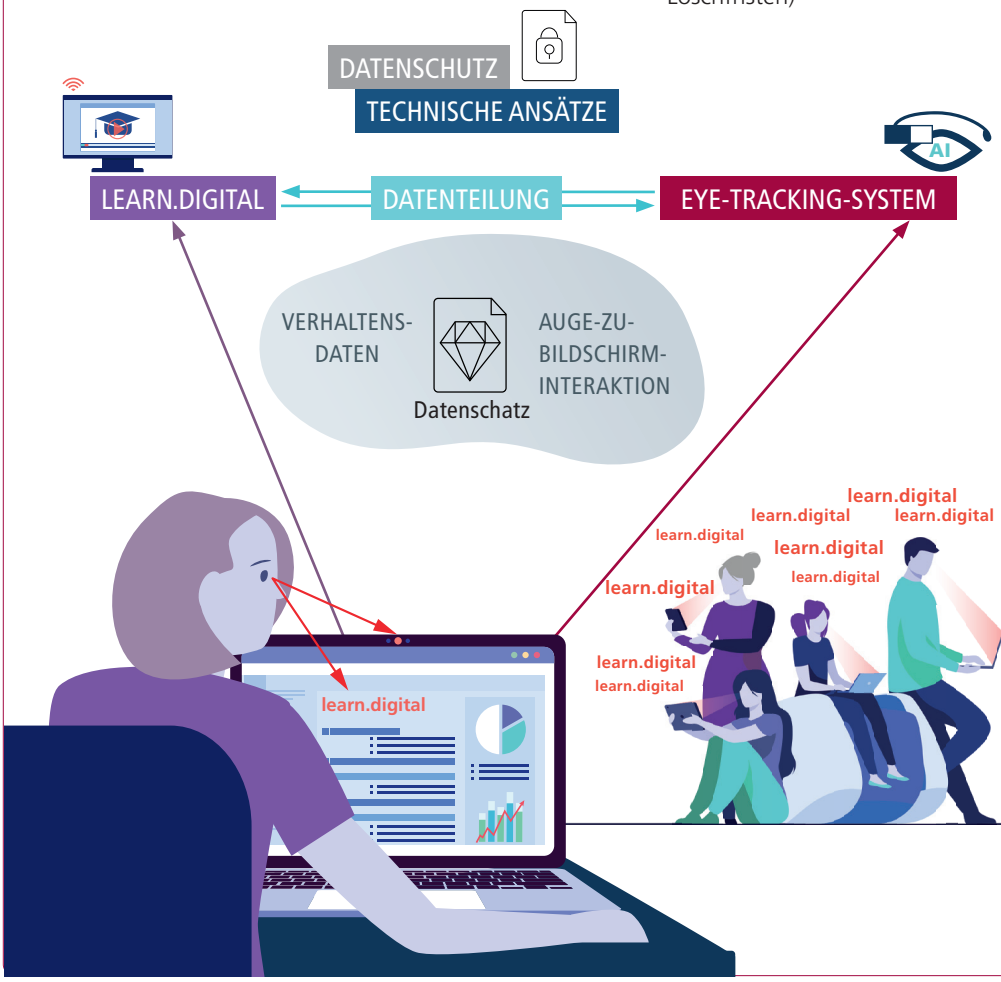
Datenschutz
Datenverknüpfung

Der datenschutzkonforme Datenumgang ist für folgende Phasen im Datenzyklus besonders wichtig:

Datenaufbereitung und Datenprozessierung: Art. 5 Abs.1 lit. c) DSGVO (Datenminimierung)

Datenanalyse: Art. 22 Nr. 1 DSGVO (Mitteilungspflicht, Auskunfts-/Widerspruchsrecht gegen Datenverwendung bei automatisierten Entscheidungen)

Datenspeicherung: Art. 17 DSGVO (Aufbewahrungs-/Löschfristen)



INFO

Eye-Tracking-System
Das System des externen Anbieters eye.ai analysiert über Zugriffe auf die integrierte Kamera des Endgeräts (Smartphone, Laptop etc.) Blickmuster und Pupillendaten in menschlichen Interaktionen mit der App und erlaubt so, visuelle Aufmerksamkeitsschwerpunkte, den Grad kognitiver Beanspruchung und Prozesse der Entscheidungsfindung abzuleiten.

Technische Ansätze
Homomorphic Encryption: Verschlüsselungsmethode
☞ Reidentifikation? Unklar.

Anonymisierung/Pseudonymisierung:
☞ Führt zur Reduktion der Datenqualität!

Begleitende Optionen
LIME-Ansatz:
Erklärmodell, als vertrauensschaffende Maßnahme für KI-Systeme (Erklärbare KI).
☞ Juristisch nicht datenschutzkonform!

App-Nutzende profitieren von den Online-Tools und der Lern-App der Plattform learn.digital, die sie beim Lernen individuell unterstützen, die Lerninhalte verständlich zu vermitteln und jederzeit flexibel einzusetzen – individuell auf das Alter der Lernenden angepasst.

Die Lernplattform **learn.digital** möchte die Digital Literacy ihrer App-Nutzenden stärken. Dazu nutzt das Unternehmen App-basierte Lernmethoden, die auf ihrer Plattform gebündelt werden. Analysen zeigen, dass ältere App-Nutzende oft nur eine kurze Verweildauer haben und dadurch geringe Lernerfolge erzielen. Um zu prüfen, ob die Lern-Apps für diese Zielgruppe zu wenig intuitiv sind, plant das Unternehmen den Einsatz der KI-basierten Eye-Tracking-Software von eye.ai. Die gesammelten Daten der Lernenden sollen mit den Daten von eye.ai verknüpft werden, um das Nutzungsverhalten aller Altersgruppen zu verstehen und die Lerninhalte entsprechend anzupassen. Die datenschutzrechtliche Prüfung betrifft die Zusammenführung personenbezogener Daten aus diesen beiden Quellen.

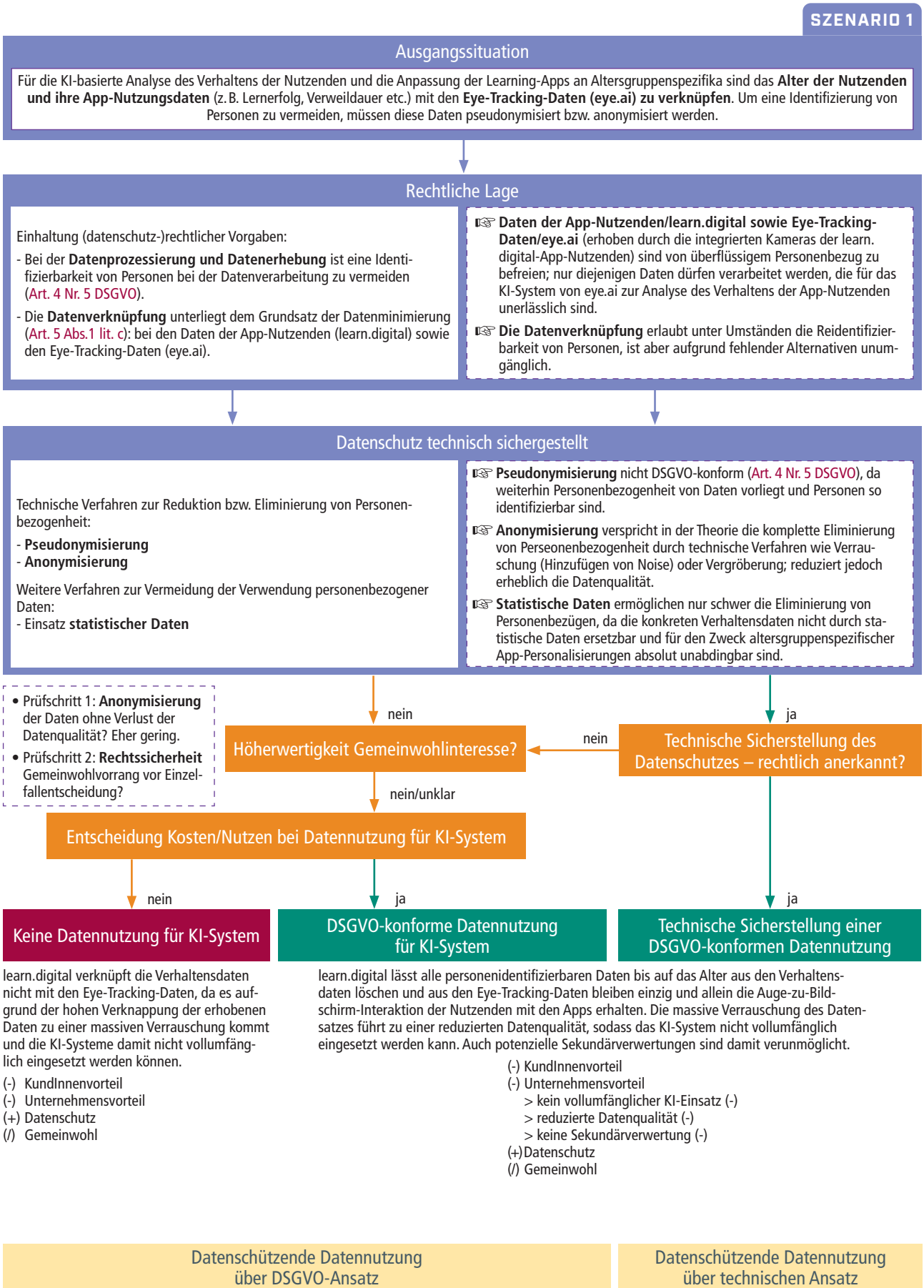
eye.ai – ein externer Anbieter für Eye-Tracking-Systeme, kann mithilfe seiner speziellen Software und mit dem Zugriff auf die integrierten Kameras von Endgeräten (wie Smartphones, Laptops) Blickmuster und Pupillendaten in menschlichen Interaktionen mit den learn.digital-Apps analysieren. Auf Basis dieser gewonnenen Daten möchte learn.digital das KI-System nutzen, um eine altersspezifische Interaktion für ihre eigenen Lerntools abzuleiten.

Ziele des Unternehmens, auch mit KI-Einsatz

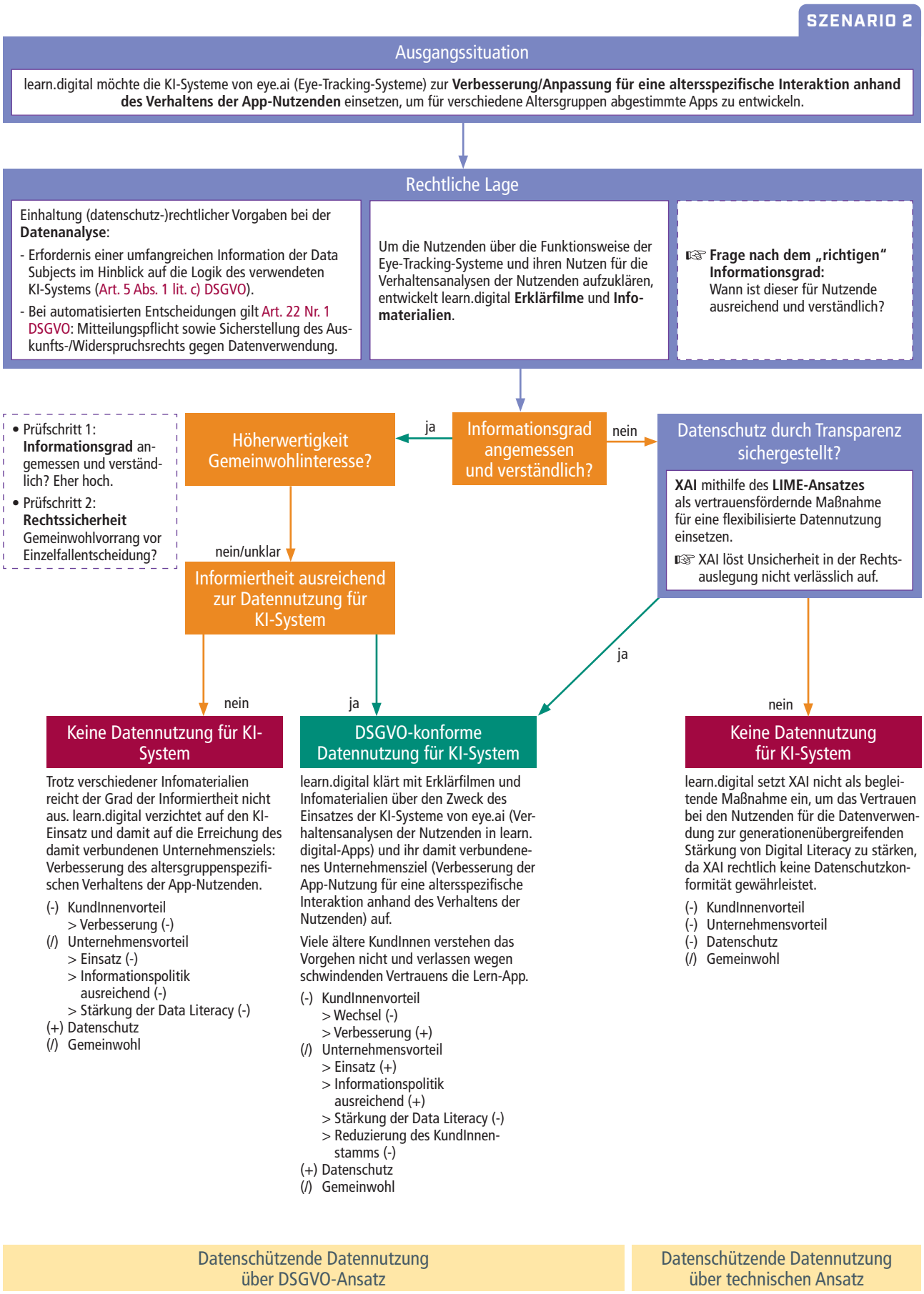
Einsatz eines Eye-Tracking-Systems des externen Anbieters eye.ai zur

- **Stärkung der Digital Literacy der App-Nutzenden**
- **Weiterentwicklung eines App-basierten lebenslangen Lernens aller Altersgruppen für individuellen Lernerfolg**
- **altersspezifische Interaktion für die Lerntools**

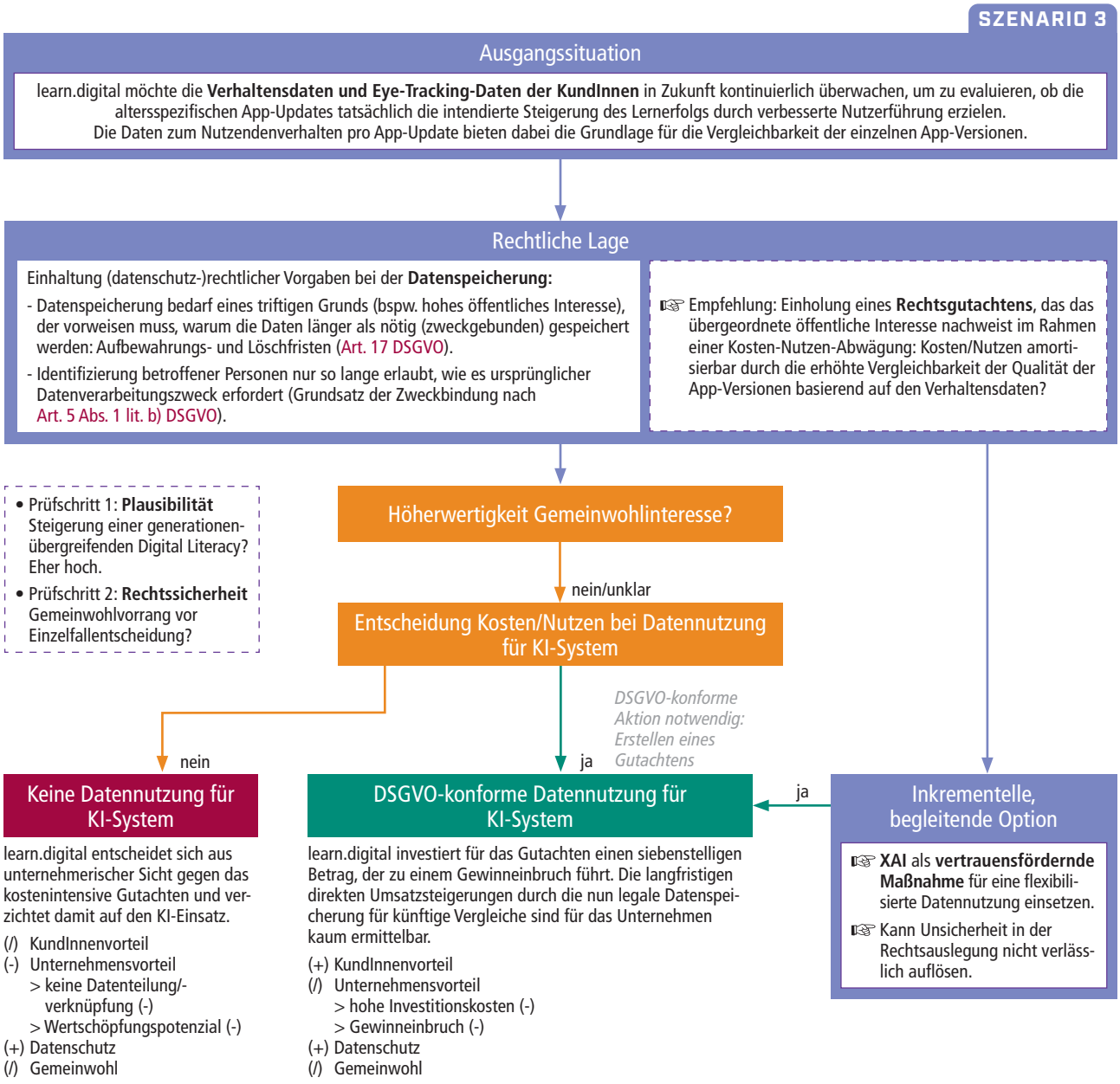
Datenverknüpfung – Kernphase Datenprozessierung



Datenanalyse



Datenspeicherung



Datennutzung mit technischen Verfahren flexibilisieren – Verschlüsselungsmethoden

SZENARIO 4

Ausgangssituation

learn.digital überlegt, **Verschlüsselungsmethoden zur datenschutzwahrenden flexibilisierten Datennutzung** für die Eye-Tracking-Analyse mit den KI-Systemen von eye.ai anzuwenden.

Datenschutz technisch sichergestellt?

Technische Verfahren zur Verschlüsselung von Daten:

- **Homomorphic Encryption (HC):** alle personenbezogenen Datenpunkte werden beim Data Subject verschlüsselt, insbesondere biografische Informationen.
- **Secure Multiparty Computation (SMPC):** kein Zugriff auf die Daten, die nur für die KI-Systeme von eye.ai genutzt werden, um das Verhalten der Nutzenden zu analysieren.

☞ **Homomorphic Encryption:** hohe Rechenlast; zu teuer im Verhältnis des zu erwartenden Ergebnisses.
☞ **Secure Multiparty Computation:** Verschlüsselung der Daten wie der Übertragung selbst; sehr hohe Rechenlast; keinen Zugriff auf die Datenpunkte; teurer als HC.

nein

ja

- Prüfschritt 1: **Plausibilität** Verschlüsselungsmethoden effektiv einsetzbar? Eher gering.
- Prüfschritt 2: **Rechtssicherheit** Gemeinwohlvorrang vor Einzelfallentscheidung?

Höherwertigkeit Gemeinwohlinteresse?

nein

Technische Sicherstellung des Datenschutzes – rechtlich anerkannt?

nein/unklar

Entscheidung Kosten/Nutzen bei Datennutzung für KI-System

nein

ja

DSGVO-konforme Aktion notwendig: Nachträgliche Einwilligung

ja

Technische Sicherstellung einer DSGVO-konformen Datennutzung für KI-System

learn.digital setzt die Verschlüsselungsmaßnahmen ein. Bildet aber Rücklagen, da nicht sicher ist, ob eine Reidentifikation möglich ist.

(-) KundInnenvorteil
(/) Unternehmensvorteil
> Einsatz (-)
> Keine Wirtschaftlichkeit (-)
> Stärkung der Data Literacy (-)
(+) Datenschutz
(/) Gemeinwohl

Keine Datennutzung für KI-System

Aufgrund der geringen Digital Literacy der Nutzenden und der hohen Rechenlast setzt learn.digital keine Verschlüsselungsmaßnahme ein. Zudem wäre auch die Verschlüsselung der Datenpunkte nach Einschätzung der IT-Abteilung teurer als die Vorteile, die learn.digital aus der Analyse der Verhaltensdaten der Nutzenden mittels der KI-Systeme von eye.ai ziehen könnte.

(-) KundInnenvorteil
> Stärkung der Data Literacy (-)
(/) Unternehmensvorteil
> Einsatz (-)
> Keine Wirtschaftlichkeit (-)
(+) Datenschutz
(/) Gemeinwohl

DSGVO-konforme Datennutzung für KI-System

Trotz der geringen Digital Literacy der Nutzenden und der hohen Rechenlast setzt learn.digital die Verschlüsselungsmaßnahmen ein.

(-) KundInnenvorteil
(/) Unternehmensvorteil
> Einsatz (-)
> Keine Wirtschaftlichkeit (-)
> Stärkung der Data Literacy (-)
(+) Datenschutz
(/) Gemeinwohl

Datenschützende Datennutzung über DSGVO-Ansatz | Datenschützende Datennutzung über technischen Ansatz

vAItality – Gesundheitswesen

Datenweitergabe an Dritte für KI-basierte Vitalfunktionsoptimierung

Das Med-Tech-Start-up vAItality hat einen KI-basierten Vitalfunktionsoptimierer zur Überwachung des Schlafverhaltens entwickelt, der einen direkten (Daten-)Austausch zur Vorsorge und Therapie zwischen Patientinnen/Patienten und Ärztinnen/Ärzten ermöglicht. Die damit erhobenen PatientInnen-Gesundheitsdaten dienen vor allem der Behandlung für einen gesunden Schlaf und sind nicht nur für die weitere Optimierung des KI-basierten Modells von Belang, sondern bieten auch anderen Akteuren aus dem Gesundheitswesen – wie universitäre Forschungsprojekte oder Pharmaunternehmen – enormes Potenzial für innovative gesundheitsfördernde Geschäftsmodelle und Produkte.

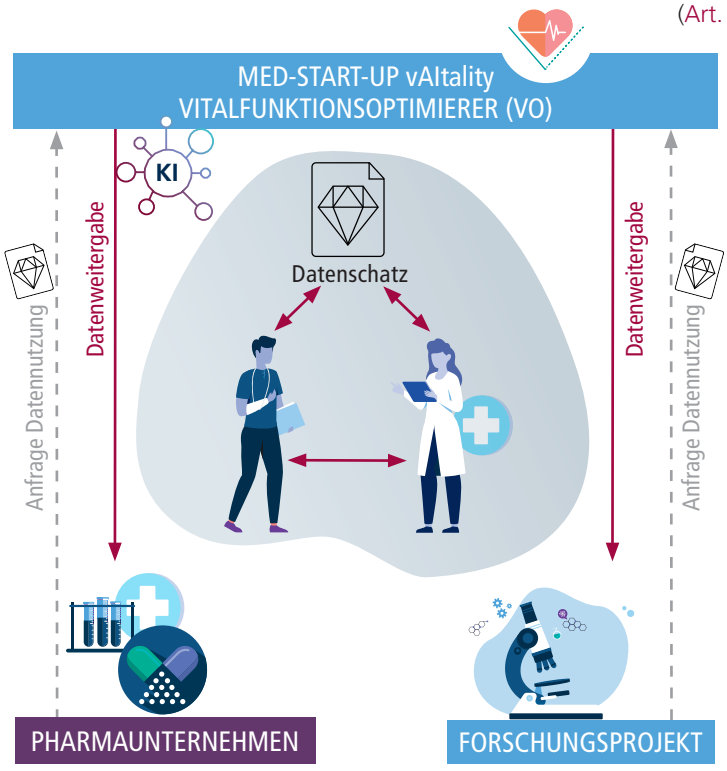
Datenschatz
Gesundheitsdaten

- vergangene, gegenwärtige oder zukünftige Gesundheitsdaten
- körperliche/geistige Gesundheit einer Person
- bei Anmeldung/Erbringung von Gesundheitsdienstleistungen
- Fitness-Apps
- medizinische intelligente Implantate
- etc.

Datenschutz
Datenweitergabe an Dritte

Bei der Weitergabe der persönlichen Daten von PatientInnen an Dritte für weitere innovative Geschäftsmodelle und Produkte, gemeinwohlorientierte Zwecke (z. B. Forschungszwecke) ist zu beachten:

- Einwilligung der betroffenen Personen (Art. 4 Nr. 11 DSGVO)
- Informiertheit der betroffenen Personen (Art. 13 DSGVO)
- Zweckbindung der Datenerhebung (Art. 5 Abs. 1 lit. b) DSGVO)



Datennutzung zur zielgenauen Medikamentierung eines neuen Herzmedikaments

Datennutzung zur Früherkennung vaskulärer Erkrankungen

KI-basierter Vitalfunktionsoptimierer überwacht und stimuliert die Vitalfunktionen (Atmung, Kreislauf, Körpertemperatur) während des Schlafes.
Auf Basis der Vitalfunktionsaufzeichnung können ÄrztInnen die Behandlung abstimmen bzw. sogar nach Notruf durch das Tool informiert werden.

INFO

Technische Ansätze

Differential Privacy (DP):
Verfahren zur Anonymisierung der Daten für den Trainingsdatensatz

☞ Verrauschung führt zu einer signifikanten Reduktion der Genauigkeit der über das ML-Modell erfolgten Klassifikation der PatientInnen.

Federated Learning:
Datenschutzkonformes Training aller KI-Elemente, bei dem die verwendeten Daten auf den Endgeräten der UserInnen (nicht zentral!) gebündelt werden und damit bei diesen verbleiben. (Mehr zu Federated Learning)

☞ Neue Angriffspunkte für Cyberkriminelle

Federated Learning + Differential Privacy (Hybride Modelle):
Absicherung des Austausches des Wights

☞ Fehlkassifikationen/gesundheitsschädliche Fehlstimulierungen aufgrund der signifikanten Reduktion der Datengenauigkeit

Unabhängige Vermittler-Instanzen, eigenverantwortliche Datenorganisation

Personal Information Management Systems (PIMS):
Stärkung der Souveränität von PatientInnen hinsichtlich ihrer Gesundheitsdaten aufgrund privatwirtschaftlichen Interesses

Datentreuhänder mit Datenenklave:
Stärkung der Forschungs-/Gemeinwohlzwecke sowie Datensouveränität von PatientInnen

Anton Merk, 65 Jahre, hat kürzlich seinen Lungenkrebs besiegt; dies mithilfe KI-gestützter Methoden der Krebserkennung, -behandlung und -therapie. Da ihn seit Jahren Schlaf- und Herzrhythmusstörungen plagen, fühlt er sich oft schlapp und schwach. Seine Hausärztin rät ihm, den in Kürze auf den Markt kommenden KI-basierten Vitalfunktionsoptimierer des Med-Tech-Start-up vAltality zu nutzen, der das Schlafverhalten nicht nur überwacht, sondern auch optimieren kann. Anton ist begeistert von dieser Empfehlung und registriert sich als Nutzer des Vitalfunktionsoptimierers ab Produktstart.

vAltality, ein Med-Tech-Start-up, hat sein erstes KI-Produkt entwickelt: den KI-basierten Vitalfunktionsoptimierer. Dieser soll in Kürze auf dem Markt eingeführt werden. Im Vorfeld gilt es für das junge Start-up-Unternehmen noch einiges abzuklären: insbesondere datenschutzrechtliche Vorgaben. Da es sich beim Vitalfunktionsoptimierer um eine KI-basierte Technologie handelt, zu der bisher kaum Gerichtsentscheidungen oder Leitlinien vorliegen, kann der Einsatz mit einem gewissen Risiko verbunden sein. Daher holt sich vAltality vor Einsatz des Vitalfunktionsoptimierers rechtlichen wie technischen Rat ein.

UNIVERSITÄRES FORSCHUNGSPROJEKT

vAltality-Gesundheitsdaten zur Früherkennung vaskulärer Erkrankungen nutzen

Vor Projektstart erhält vAltality eine Anfrage aus der medizinischen Forschung zur entgeltfreien Nutzung der potenziellen PatienInnendaten für ein großes universitäres Forschungsprojekt zur Früherkennung vaskulärer Erkrankungen.

DATEN-WEITERGABE AN DRITTE

PHARMAUNTERNEHMEN

vAltality-Gesundheitsdaten zur Entwicklung eines neuen Herzmedikaments nutzen

Ein Pharmaunternehmen wird aufgrund der erfolgreichen Markteinführung und der positiven Resonanz seitens der PatientInnen und ÄrztInnen auf den Vitalfunktionsoptimierer aufmerksam und sieht in den vAltality-Gesundheitsdaten wirtschaftliches Potenzial. Das Pharmaunternehmen fragt daher bei vAltality an, ob die Gesundheitsdaten für eigene medizinische Zwecke käuflich zu erwerben sind.

Ziele des Unternehmens, auch mit KI-Einsatz

- **Entwicklung eines KI-Vitaloptimierers**
- **Weitergabe der Gesundheitsdaten als Datenschatz für das Gemeinwohl**
- **Generierung neuartiger Geschäftsmodelle mit Dritten**

Datenerhebung für den Vitalfunktionsoptimierer und Datenteilung mit universitärem Forschungsprojekt

SZENARIO 1

Ausgangssituation

Vor Markteinführung des Vitalfunktionsoptimierers hat vAItality u. a. rechtliche Fragen zum Einsatz KI-basierter Anwendungen zu klären. Vor Projektstart erhält vAItality eine **Anfrage aus der medizinischen Forschung zur entgeltfreien Nutzung** der potenziellen PatientInnendaten für ein großes universitäres Forschungsprojekt zur Früherkennung vaskulärer Erkrankungen.

Rechtliche Lage

Bei der Datenerhebung für die **Datennutzung und Datenteilung** gelten folgende Datenschutzauflagen:

- Einwilligung der betroffenen Personen (Art. 4 Nr. 11 DSGVO)
- Informiertheit der betroffenen Personen (Art. 13 DSGVO)
- Zweckbindung der Datenerhebung (Art. 5 Abs. 1 lit. b) DSGVO)

☞ Beide Zwecke – Datennutzung wie Datenteilung (Forschungsprojekt) – sind datenschutzkonform abzudecken.

☞ vAItality erstellt ein Informationsblatt, bei dem die Nutzenden nicht einem einzelnen Zweck, sondern nur beiden gemeinsam (gekoppelt) zustimmen können.

☞ **Frage nach Zweckbindung/„richtigem" Informationsgrad:** Die Formulierung zur Zweckbindung lautet: *„Ihre Gesundheits- und Bewegungsdaten werden für die Optimierung Ihrer Vitalfunktionen mittels eines KI-Systems sowie für ein medizinisches Forschungsprojekt zur Früherkennung von Krankheiten verwendet."*

- Einwilligung zum Forschungszweck nicht ausreichend: kein „broad consent"
- Forschungsprojekt darf die von vAItality erhobenen Daten nicht verwenden
- vAItality darf den Vitalfunktionsoptimierer weiter anbieten
- Informationspolitik von vAItality transparent

Datennutzung: Einwilligung notwendig – Informationsgrad angemessen und verständlich?

nein

Technische Umsetzung: Datenschutz über Datenzugriff-Managementsysteme sichergestellt?

Datentreuhänder mit Datenenklave: Stärkung der Datensouveränität von PatientInnen zu Forschungs-/Gemeinwohlzwecken

☞ PatientInnen können personenbezogene Daten (z. B. Vorerkrankungen) in die Datenenklave hochladen und geben breite Einwilligung für die Datenverwertung zu gemeinwohlorientierten, privatwirtschaftlichen sowie Forschungszwecken, damit Entscheidungs- oder Wahlfreiheit für oder gegen.

ja

Datenzugriff-Managementsysteme zur DSGVO-konformen Datennutzung für KI-System

vAItality registriert sich beim Datentreuhänder Data4Health, der dem Start-up aufgrund der Verbesserung der individuellen PatientInnensituation Gemeinwohlorientierung bescheinigt. vAItality verknüpft in der Datenenklave die dort abgespeicherten personenbezogenen PatientInnendaten des Vitalfunktionsoptimierers mit ihren selbst gesammelten Daten und trainiert die KI-Elemente des Optimierers.

(+) PatientInnenvorteil
> Datensouveränität (+)
> Steigerung der Gesundheit (+)

(+ Unternehmensvorteil
> Umsatzsteigerung (+)

(+) Datenschutz
(+) Gemeinwohl

Auch das Forschungsprojekt registriert sich beim Datentreuhänder und nutzt die dort gespeicherten Daten in der Trusted Execution Environment der Enklave für ihr Forschungsvorhaben.

(+) Forschungsprojekt
> Forschungsergebnisse (+)

nein

Keine Datenteilung mit dem Forschungsprojekt

Das Forschungsprojekt wird aufgrund des rechtlichen Verbots der Datennutzung abgebrochen und muss alle personenbezogenen Daten löschen. Der Forschungsfortschritt ist dahin, weitere potenzielle Fortschritte bei der Erkennung vaskulärer Erkrankungen bleiben aus. Aufgrund begrenzter Fördermittel kann der Forschungsrückstand nicht mehr auf-, auch die nachträgliche Einwilligung nicht mehr eingeholt werden.

(/) Forschungsprojekt
> keine medizinischen Fortschritte (-)
> keine finanzielle Förderung (-)

ja

DSGVO-konforme Datennutzung für KI-System

vAItality versendet das Informationsblatt zusammen mit der Zweckbindung an alle den Vitaloptimierer vertreibenden ÄrztInnen, die es an ihre PatientInnen austeilen. vAItality entstehen dadurch Administrationskosten im fünfstelligen Bereich. Viele PatientInnen – wie Anton – stimmen der Datennutzung für beide Zwecke zu; viele geben aber die Datennutzung für das Forschungsprojekt nicht frei und können den Vitalfunktionsoptimierer daher nicht nutzen.

(-) PatientInnenvorteil
> Wechsel (-)
> Steigerung der Gesundheit (+)

(/) Unternehmensvorteil
> hohe Investitionskosten (-)
> kein vollumfänglicher KI-Einsatz (-)
> Reduzierung des Patient-Innenstamms (-)
> reduzierte Datenqualität (-)

(+) Datenschutz

(/) Gemeinwohl

Datenschützende Datennutzung über DSGVO-Ansatz	Datenschützende Datennutzung über technischen Ansatz

Datenerwerb für Produktentwicklung eines Pharmaunternehmens

SZENARIO 2

Ausgangssituation

Ein Pharmaunternehmen möchte die **Gesundheitsdaten für die Entwicklung eines neuen Herzmedikaments (bereits in der Klinikphase) kaufen**, um mit diesen eine individualisierte Medikamentierung zu erzielen. Dies verspricht u. a. einen Innovationssprung in der Prävention von Herzinfarkten. vAltality erkennt den gesellschaftlichen Mehrwert, schließt einen Vertrag mit dem Pharmaunternehmen über den Datenaustausch und erhält einen siebenstelligen Betrag – PatientInnen profitieren monetär nicht von der Weitergabe ihrer Daten.

Rechtliche Lage

Für die Verwendung der Gesundheitsdaten stehen dem Pharmaunternehmen zwei Möglichkeiten zur Verfügung:

a) Einholung einer aufwändigen **nachträglichen Einwilligung** (DSGVO Art. 6 Abs. 4) aller vAltality-PatientInnen
b) Vollzug einer **individuellen Interessenabwägung** (DSGVO Art. 6 Abs. 1 S. 1 lit. f)

☞ Beim Datenerwerb besteht in der Umsetzung Interpretationsunsicherheit.

Technische Umsetzung: Datenschutz über Datenzugriff-Managementsysteme sichergestellt?

Personal Information Management Systems (PIMS) ermöglichen die Stärkung der Souveränität von PatientInnen hinsichtlich ihrer Gesundheitsdaten, vor allem bei privatwirtschaftlichem Interesse.

☞ Missbrauch der Machtposition seitens der Datentreuhänder möglich; Datenschutz sollte daher per By-Design-Ansatz (z. B. bei PIMS) auf Gemeinwohl ausgerichtet sein.

nein / ja

- Prüfschritt 1: **Plausibilität** Innovationssprung in der Herzinfarktprävention? Eher hoch.
- Prüfschritt 2: **Rechtssicherheit** Gemeinwohlvorrang vor Einzelfallentscheidung?

Höherwertigkeit Gemeinwohlinteresse?

nein/unklar / ja

Keine DSGVO-konforme Aktion notwendig

Entscheidung Kosten/Nutzen bei Datennutzung für KI-System

nein / ja

DSGVO-konforme Aktion notwendig: Nachträgliche Einwilligung

Keine Datennutzung für KI-System

Das Pharmaunternehmen verzichtet wegen der teuren und aufwändigen Einholung der Einwilligungen aller PatientInnen auf die Gesundheitsdaten zur Entwicklung eines neuen Herzmedikaments.

(-) PatientInnenvorteil
(-) Unternehmensvorteil
(+) Datenschutz
(/) Gemeinwohl

DSGVO-konforme Datennutzung für KI-System

Das Pharmaunternehmen holt die Einwilligungen aller PatientInnen ein. Einige PatientInnen verweigern die Einwilligung aber.

(+) PatientInnenvorteil
> Wechsel (-)
> Verbesserung (+)
(-) Unternehmensvorteil
> Einsatz (+)
> Aufwand (-)
> Reduzierter PatientInnenstamm
(+) Datenschutz
(+) Gemeinwohl

DSGVO-konforme Datennutzung für KI-System

Das Pharmaunternehmen entscheidet sich für die individuelle Interessenabwägung, da mit dem Medikament ein Innovationssprung in der Herzinfarktprävention wahrscheinlich ist. Dies verspricht eine Steigerung der gesamtgesellschaftlichen Gesundheit. Die Daten der PatientInnen werden somit ohne nachträgliche Einwilligung genutzt.

(+) PatientInnenvorteil
(+) Unternehmensvorteil
(+) Datenschutz
(+) Gemeinwohl

Datenzugriff-Managementsysteme zur DSGVO-konformen Datennutzung für KI-System

Das Pharmaunternehmen – und auch weitere privatwirtschaftliche Gesundheitsakteure fragen direkt in der PIMS-App zur PatientInnen-Datennutzung an.

(+) PatientInnenvorteil
- Stärkung der Datensouveränität
(+) Unternehmensvorteil
- Datenteilung (+)
- Umsatzsteigerung (+)
(+) Datenschutz
- gleichberechtigte Akteure in der Datenökonomie (+)
(+) Gemeinwohl

☞ **vAltality** empfiehlt seinen PatientInnen, sich auf der PIMS-App sovereign.data zu registrieren und ihre personenbezogenen Daten dort zu hinterlegen: PatientInnen entscheiden selbst, ob sie ihre Daten – entgeltlich oder unentgeltlich – zur Verfügung stellen oder für privatwirtschaftliche Zwecke grundsätzlich ausschließen.

Datenschützende Datennutzung über DSGVO-Ansatz

Datenschützende Datennutzung über Datenzugriff-Managementsysteme

Datennutzung mit technischen Verfahren flexibilisieren

SZENARIO 3

Ausgangssituation

Um den Datenschutz bei der Anwendung des Vitalfunktionsoptimierers sicherzustellen, lässt vAItality technische Aspekte von seinem Entwicklungspartner überprüfen. Dieser empfiehlt, verschiedene **technische Verfahren (auch KI-basierte)** in die Entwicklung des Vitalfunktionsoptimierers einzubeziehen. Dadurch soll eine **datennutzungsorientierte Flexibilität für KI-Systeme** gefördert werden, die dem Gemeinwohl dient, während gleichzeitig die Kontrolle der datengebenden Personen in der Datenökonomie gestärkt wird.

Rechtliche Lage

Einhaltung (datenschutz-)rechtlicher Vorgaben:

☞ Bei der **Datenprozessierung** und **Datenerhebung** ist eine Identifizierbarkeit von Personen bei der Datenverarbeitung zu vermeiden (Art. 4 Nr. 5 DSGVO).

☞ Die **Datenverknüpfung** unterliegt dem Grundsatz der Datenminimierung (Art. 5 Abs.1 lit. c).

☞ In den verschiedenen Daten-Lebenszyklusphasen sind die für den Trainingsdatensatz der KI-Elemente des Vitalfunktionsoptimierers verwendeten PatientInnendaten datenschutzkonform zu gestalten. Ebenso eine datenschutzkonforme Absicherung des Datenaustausches.

Datenschutz technisch sichergestellt?

Differential Privacy (DP) zur Anonymisierung der PatientInnendaten, um den Trainingsdatensatz für KI-Elemente des Vitalfunktionsoptimierers datenschützend auszugestalten.

Neben allen biografischen Daten im vorbereiteten Trainingsdatensatz sind auch einige Vitalfunktionsdaten zu verrauschen, da diese in ihrer Zusammensetzung Identifizierbarkeit erlauben könnten.

Eine mit Intensitätsintervallen trainierende Spitzensportlerin, die den Vitalfunktionsoptimierer nutzt, könnte fälschlicherweise als ältere Frau mit Herzrhythmusstörungen klassifiziert werden, was gesundheitsschädliche Fehlstimulierungen durch den Vitalfunktionsoptimierer provozieren könnte.

☞ **Verrauschung** führt zur signifikanten Reduktion der Genauigkeit der über das ML-Modell erfolgten Klassifikation der PatientInnen.

☞ Datenschutzkonformität beim Training wird so sichergestellt, vAItality kann die individuellen personenbezogenen Daten der PatientInnen nicht einsehen.

Federated Learning zum datenschutzkonformen Training aller KI-Elemente des Vitalfunktionsoptimierers.

vAItality entwickelt zentral ein ML-Modell, bei dem Vitalfunktionswerte von PatientInnen abhängig von ihren biografischen Daten (z. B. Vorerkrankungen, Ernährungs-/Bewegungsgewohnheiten, Alter, Gewicht etc.) klassifiziert werden.

Das auf dem vAItality-Server gespeicherte ML-Modell wird von den Vitalfunktionsoptimierern der PatientInnen heruntergeladen und mit deren Gesundheitsdaten und anderen personenbezogenen Daten lokal trainiert. Nur die Ergebnisse (Weights) des lokal trainierten Modells werden an den zentralen Server zurückgesendet. Dieser aggregiert die Weights aller Vitalfunktionsoptimierer der PatientInnen und aktualisiert damit das ML-Modell.

☞ **Bösartige Extraktionsangriffe** möglich, bei denen ein Server jedem Client zum Training andere Updates der Weights schickt, was die Modellgüte signifikant reduzieren kann.

Federated Learning + Differential Privacy als hybride Modelle zur Absicherung des Austausches der Weights bei Federated Learning.

Die lokalen Trainingsdaten der in den Vitalfunktionsoptimierern individueller PatientInnen trainierten KI-Elemente könnten so weit verändert werden, dass Identifizierbarkeit individueller PatientInnen beim Weight-Austausch nicht mehr möglich ist, sodass Datenschutz sichergestellt wäre.

☞ **DP-Einsatz** führt zur signifikanten Reduktion der Genauigkeit der über das ML-Modell erfolgten Klassifikation der PatientInnen.

nein | ja | ja

Datenschutz rechtlich anerkannt

nein | ja | ja

Keine Datennutzung für KI-System

Die Ungenauigkeit des KI-Modells über die Differential-Privacy-Komponente würde einen Einsatz unrentabel machen. vAItality verzichtet deshalb auf den Einsatz und damit auf die Verbesserung der Nutzererfahrung.

(-) PatientInnenvorteil
(-) Unternehmensvorteil
(-) Datenschutz
(/) Gemeinwohl

Technische Sicherstellung einer DSGVO-konformen Datennutzung für KI-System

vAItality setzt voll auf verteiltes Lernen und geht damit bewusst ein Datenschutzrisiko hinsichtlich möglicher Angriffe auf den Weight-Austausch oder auf die Endgeräte ein.

(+) PatientInnenvorteil
  > Steigerung der Datensouveränität
  > hohe Genauigkeit des KI-Modells
  > Angriff auf Endgeräte (-)
(+) Unternehmensvorteil
  > Umsatzsteigerung (-)
(+) Datenschutz
  > mögliche Angriffe auf Weight-Austausch/ die Endgeräte
(/) Gemeinwohl

Datennutzung für KI-System

vAItality setzt das hybride Modell ein und schützt somit das KI-Modell vor Angriffen; nimmt dafür die leichten Genauigkeitsverluste des KI-Modells in Kauf.

(+) PatientInnenvorteil
  > leichte Reduktion der Genauigkeit des KI-Modells durch Verrauschung
  > Steigerung der Souveränität
  > Schutz vor Angriff auf Endgeräte (+)
(+) Unternehmensvorteil
  > Umsatzsteigerung (-)
  > Rücklagen (-)
(+) Datenschutz
  > Schutz vor Angriffen auf Weight-Austausch
(+) Gemeinwohl

Datenschützende Datennutzung über technischen Ansatz

Literatur

Bitkom (2020): Wie man KI trainiert, ohne den Datenschutz zu verletzen. Online unter: https://bitkom.org/Presse/Presseinformation/Wie-man-KI-trainiert-ohne-den-Datenschutz-zu-verletzen

Bitkom Research, Tata Consultancy Services (2021): Nachhaltig geht nur digital. Wie Deutschland mit KI und Co. die Zukunft gestaltet. Online unter: https://bitkom-research.de/sites/default/files/Bitkom_Research_TCS_Trendstudie_2021_DE.pdf

Blankertz, A. (2020): Designing Data Trusts. Why We Need to Test Consumer Data Trusts Now. Stiftung Neue Verantwortung. Online unter: https://www.stiftung-nv.de/sites/default/files/designing_data_trusts_d.pdf

Blankertz, A. (2021): Stellungnahme von Aline Blankertz, Stiftung Neue Verantwortung für die öffentliche Anhörung des Ausschusses Digitale Agenda am 24. Februar 2021 zum Thema „Datenstrategie der Bundesregierung" (BT-Drs. 19/26450) verbunden mit „Eckpunkte einer Datenstrategie der Bundesregierung" (BT-Drs. 19/16075) und dem Antrag der Fraktion der FDP „Datenpolitik für Selbstbestimmung, Wettbewerb und Innovation" (BT-Drs. 19/26538). Deutscher Bundestag. Ausschuss Digitale Agenda. Ausschussdrucksache 19(23)107. Online unter: https://www.bundestag.de/resource/blob/823798/ed3be0d3950d659afdbd0dbb774a1a2a/Stellungnahme-Blankertz-data.pdf

Blankertz, A. & Specht-Riemenschneider, L. (2021): Neue Modelle ermöglichen. Regulierung für Datentreuhänder. böll.brief Grüne Ordnungspolitik #16. Online unter: https://www.boell.de/sites/default/files/2021-08/bo%23776ll.brief%20G16%20Neue%20Modelle%20ermo%23776glichen.pdf

Buchner, B., Haber, A. C., Hahn, H. K., Kusch, H., Prasser, F., Sax, U. & Schmidt, C. O. (2021): Das Modell der Datentreuhand in der medizinischen Forschung. Datenschutz und Datensicherheit-DuD, 45(12), 806–810. Online unter: https://link.springer.com/content/pdf/10.1007/s11623-021-1534-y.pdf

Dössel, O. & Lenarz, T. (Hrsg.): Gesundheitsdatennutzung – sicher und souverän (acatech IMPULS), München 2023. DOI: https://doi.org/10.48669/aca_2023-10

Enquete-Kommission Künstliche Intelligenz (2020): Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale. Drucksache 19/23700. Deutscher Bundestag. Online unter: https://dserver.bundestag.de/btd/19/237/1923700.pdf

Fereidooni, H. et al. (2021): SAFELearn: Secure Aggregation for private FEderated Learning. 2021 IEEE Security and Privacy Workshops, 56-62. Online unter: https://ieeexplore.ieee.org/document/9474309/authors#authors

Fraunhofer MOEZ (2013): Initiative zu einer Deutschen Daten-Treuhand (DEDATE) als Ultima Ratio der Persönlichen Digitalen Datenwirtschaft (PDD). Online unter: https://www.imw.fraunhofer.de/content/dam/moez/de/documents/Executive-Paper/DEDATE-gesamt.pdf

Gausling, T. (2020): KI und DSGVO im Spannungsverhältnis. S. 11–53. In: Ballestrem, J. G., Bär, U., Gausling, T., Hack, S. & Von Oelffen, S.: Künstliche Intelligenz. Rechtsgrundlagen und Strategien in der Praxis. Springer Gabler, Wiesbaden.

Hoeren, T. & Niehoff, M. (2018): KI und Datenschutz – Begründungserfordernisse automatisierter Entscheidungen. RW Rechtswissenschaft, 9(1), 47–66.

Huang, S., Yang, J., Fong, S. & Zhao, Q. (2020): Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. Cancer Letters, 471, 61–71. Online unter: https://doi.org/10.1016/j.canlet.2019.12.007

Leopoldina, acatech & Union der Deutschen Akademien der Technikwissenschaften (2018): Privatheit in Zeiten der Digitalisierung. Stellungnahme. Online unter: https://www.leopoldina.org/uploads/tx_leopublication/2018_Stellungnahme_BigData.pdf

Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F. & Lin, Z. (2021): When machine learning meets privacy: A survey and outlook. ACM Computing Surveys (CSUR), 54(2), 1–36. Online unter: https://doi.org/10.1145/3436755

Mo, F., Haddadi, H., Katevas, K., Marin, E., Perino, D. & Kourtellis, N. (2021): PPFL: Privacy-preserving Federated Learning with Trusted Execution Environments.

Plattform Lernende Systeme (2019): Lernende Systeme im Gesundheitswesen – Bericht der Arbeitsgruppe Gesundheit, Medizintechnik, Pflege. Online unter: https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG6_Bericht_23062019.pdf

Plattform Lernende Systeme (2021): Potenziale für industrieübergreifendes Flottenlernen. KI-Mobilitätsdatenplattform zur Risikominimierung des automatisierten Fahrens. Online unter: https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG5_Whitepaper_Mobilitaetsplattform.pdf

Plattform Lernende Systeme (2022): KI Kompakt – Verteiltes maschinelles Lernen. Besserer Datenschutz für KI-Anwendungen? Online unter: https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/KI_Kompakt/PLS_KI_Kompakt_ML.pdf

PricewaterhouseCoopers (2018): Auswirkungen der Nutzung von künstlicher Intelligenz in Deutschland. Online unter: https://store.pwc.de/de/publications/auswirkungen-der-nutzung-von-kuenstlicher-intelligenz-in-deutschland

Schallbruch, M., Huth, M., Lundbæk, L.-N., Herdeanu, C. & Attenberger, L. (2021): Künstliche Intelligenz für den öffentlichen Sektor: Masked Federated Learning als datenschutzfreundliche Lösung. Online unter: https://faculty-research.esmt.berlin/sites/faculty/files/2021-06/Xayn_DSI_Positionpaper_DE.pdf

Schwartmann, R. & Weiß, S. (2017): Leitlinien für die rechtssichere Nutzung von Pseudonymisierungslösungen unter Berücksichtigung der Datenschutz-Grundverordnung. Whitepaper zur Pseudonymisierung der Fokusgruppe Datenschutz der Plattform Sicherheit, Schutz und Vertrauen für Gesellschaft und Wirtschaft im Rahmen des Digital-Gipfels 2017. Online unter: https://www.gdd.de/downloads/whitepaper-zur-pseudonymisierung

Specht-Riemenschneider, L. & Kerber, W. (2021): Designing Data Trustees – A Purpose-Based Approach. Konrad-Adenauer-Stiftung (Hrsg.). Online unter https://www.kas.de/documents/252038/16166715/Designing+Data+Trustees.pdf/3523489b-2611-a12a-f187-3e770d1a9d94?version=1.0&t=1647261611824

Stevens, G. & Boden, A. (2022): Warum wir parteiische Datentreuhänder brauchen. Zum Modell der Datentreuhänderschaft als stellvertretende Deutung der Interessen individueller und kollektiver Identitäten.
Online unter: https://www.verbraucherforschung.nrw/sites/default/files/2022-02/zth-06-stevens-boden-warum-wir-parteiische-datentreuhaender-brauchen.pdf

Stiftung Datenschutz (2021): Potenziale von Künstlicher Intelligenz mit Blick auf das Datenschutzrecht. Gutachten.
Online unter: https://stiftungdatenschutz.org/fileadmin/Redaktion/Dokumente/Gutachten-Studien/Stiftung-Datenschutz_Gutachten-Georg-Borges-Potenziale-Kuenstliche-Intelligenz-Datenschutzrecht-2021-12.pdf

Torkzadehmahani, R. et al. (2022): Privacy-preserving artificial intelligence techniques in biomedicine. Methods of Information in Medicine. Online unter: https://doi.org//10.1055/s-0041-1740630

Trésoret, M. (2018): Randnummer 56. In Schlegel, R., Voelzke, T. (Hrsg.): juris PraxisKommentar SGB VIII – Kinder- und Jugendhilfe, 2. Auflage.

TÜV-Verband (2020): Sicherheit und Künstliche Intelligenz. Einstellungen, Hoffnungen, Emotionen.
Online unter: https://www.tuev-verband.de/?tx_epxelo_file[id]=824710&cHash=5b6624273d780ebc87cafc44bc821a35

Vaishya, R., Javaid, M., Haleem Khan, I. & Haleem, A. (2020): Artificial Intelligence (AI) applications for COVID-19 pandemic. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 14 (4), 337–339.

XayNet & CMS (2020): Privacy Aspects of Federated Learning, Legal Review by XayNet & CMS.
Online unter: https://uploads-ssl.webflow.com/5f0c5c0bb18a279f0a62919e/5fcfa8e3389ecc84a9309513_XAIN%20Legal%20Review%202020%20v1.pdf

Über dieses Whitepaper

Die Autorinnen und Autoren sind Mitglieder der Arbeitsgruppe IT-Sicherheit, Privacy, Recht und Ethik der Plattform Lernende Systeme. Als eine von insgesamt sieben Arbeitsgruppen thematisiert sie Fragen zur Sicherheit (Security), Zuverlässigkeit (Safety) und zum Umgang mit Privatheit (Privacy) bei der Entwicklung und Anwendung von Lernenden Systemen. Sie analysiert zudem damit verbundene rechtliche sowie ethische Anforderungen und steht in engem Austausch mit allen weiteren Arbeitsgruppen der Plattform Lernende Systeme.

Autorinnen und Autoren
Marian Gläser, Brighter AI Technologies AG
Dr. Marit Hansen, Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein
Dr. Detlef Houdeau, Infineon Technologies AG
Prof. Dr. Michael Huth, Imperial College London
Prof. Dr. Jörn Müller-Quade, Karlsruher Institut für Technologie (KIT Karlsruhe)
Peter Rost, secunet Security Networks AG
Prof. Dr. Ahmad-Reza Sadeghi, Technische Universität Darmstadt
Prof. Dr. Louisa Specht-Riemenschneider, Universität Bonn
Dr. Dirk Wacker, Giesecke+Devrient GmbH

Redaktion
Jan Biehler, Geschäftsstelle der Plattform Lernende Systeme
Dr. Erduana Wald, Geschäftsstelle der Plattform Lernende Systeme
Christine Wirth, Geschäftsstelle der Plattform Lernende Systeme

Impressum

Herausgeber
Lernende Systeme –
Die Plattform für Künstliche Intelligenz
Geschäftsstelle | c/o acatech
Karolinenplatz 4 | 80333 München
www.plattform-lernende-systeme.de

Gestaltung und Produktion
PRpetuum GmbH, München

Stand
Oktober 2023

Bildnachweis
sdecoret/Adobe Stock/Titel
freepik/S. 7, 13, 29, 30, 34, 35

Empfohlene Zitierweise
Müller-Quade, J., Houdeau, D. et al.: Datenschatz für KI nutzen, Datenschutz mit KI wahren. Technische und rechtliche Ansätze für eine datenschutzkonforme, gemeinwohlorientierte Datennutzung. Whitepaper aus der Plattform Lernende Systeme, München.
DOI: https://doi.org/10.48669/pls_2023-5

Bei Fragen oder Anmerkungen zu dieser Publikation kontaktieren Sie bitte Dr. Thomas Schmidt (Leiter der Geschäftsstelle):
kontakt@plattform-lernende-systeme.de

Über die Plattform Lernende Systeme

Die Plattform Lernende Systeme ist ein Netzwerk von Expertinnen und Experten zum Thema Künstliche Intelligenz (KI). Sie bündelt vorhandenes Fachwissen und fördert als unabhängiger Makler den interdisziplinären Austausch und gesellschaftlichen Dialog. Die knapp 200 Mitglieder aus Wissenschaft, Wirtschaft und Gesellschaft entwickeln in Arbeitsgruppen Positionen zu Chancen und Herausforderungen von KI und benennen Handlungsoptionen für ihre verantwortliche Gestaltung. Damit unterstützen sie den Weg Deutschlands zu einem führenden Anbieter von vertrauenswürdiger KI sowie den Einsatz der Schlüsseltechnologie in Wirtschaft und Gesellschaft. Die Plattform Lernende Systeme wurde 2017 vom Bundesministerium für Bildung und Forschung (BMBF) auf Anregung des Hightech-Forums und acatech – Deutsche Akademie der Technikwissenschaften gegründet und wird von einem Lenkungskreis gesteuert.